



**UNIVERSIDADE DE BRASÍLIA - UnB  
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE  
DEPARTAMENTO DE ECONOMIA  
MESTRADO EM ECONOMIA DO SETOR PÚBLICO DEPARTAMENTO**

**RHAYNNEMAN ROGÉRIO MACÊDO**

**PREVISÃO DE RECEITA DO IMPOSTO SOBRE A PROPRIEDADE DE  
VEÍCULOS AUTOMOTORES (IPVA) DO ESTADO DE GOIÁS**

**Brasília – Brasil  
Outubro de 2023**

**RHAYNNEMAN ROGÉRIO MACÊDO**

**PREVISÃO DE RECEITA DO IMPOSTO SOBRE A PROPRIEDADE DE VEÍCULOS AUTOMOTORES (IPVA) DO ESTADO DE GOIÁS**

**Dissertação apresentada ao Departamento de Economia – FACE da Universidade de Brasília – UnB, como requisito para obtenção do título de Mestre Profissional em Economia do Setor Público – MESP.**

**Orientador: Prof. Dr. Fábio Ávila de Castro**

**Brasília – Brasil  
Outubro de 2023**

**RHAYNNEMAN ROGÉRIO MACÊDO**

**PREVISÃO DE RECEITA DO IMPOSTO SOBRE A PROPRIEDADE DE  
VEÍCULOS AUTOMOTORES (IPVA) DO ESTADO DE GOIÁS**

Trabalho final de Curso apresentado no Mestrado Profissional em  
Economia do Setor Público – MESP do Departamento de Economia –  
FACE da Universidade de Brasília – UnB.

**BANCA EXAMINADORA**

---

**Prof. Dr. Fábio Ávila de Castro**  
**Orientador**

---

**Prof. Dra. Maria Eduarda Tannuri Pianto**

---

**Prof. Dr Paulo Roberto Scalco**

“Dedico este trabalho a companheira Emília motivadora incondicional em todos os momentos da minha trajetória acadêmica.”

## **AGRADECIMENTOS**

Aos meus pais que me amaram antes que eu nascesse, Ademar Macedo de Oliveira e Cléia Rubia Alves Dias, sem eles, eu não seria.

Uma menção especial à Emília Cristine, companheira e apoiadora que se fez presente em toda essa trajetória. Sem você não teria conseguido!

Ao meu estimado orientador, Fábio Ávila de Castro, sou eternamente grato pela confiança e segurança, sua mediação proporcionou avanços que conduziram a conclusão da pesquisa.

Agradeço a todos aqueles que tornaram possível a realização deste mestrado. Em particular, aos professores, pesquisadores, colegas de turma e amigos que deram suporte e incentivo durante todo o processo de aprendizagem. Agradeço, também, à Universidade de Brasília, Faculdade de Economia e a Escola de Governo do Estado de Goiás que me deram a chance de estudar e aprimorar meus conhecimentos.

Aos colegas da minha turma de mestrado, obrigado pela contribuição constante de outros pontos de vista e explicações enriquecedoras. Nossas discussões e suas colaborações foram balizadores para meu desenvolvimento acadêmico e pessoal.

## RESUMO

Este trabalho aborda o tema da previsão das receitas do Imposto sobre a Propriedade de Veículos Automotores (IPVA) no Estado de Goiás. O trabalho se concentra na análise comparativa de diferentes métodos de previsão univariada, enfatizando as técnicas de alisamento exponencial de Holt (aditivo e multiplicativo) e autorregressivos ARIMA e SARIMA. Para o estudo foi considerado o período de arrecadação entre 2003 a 2022, que foi deflacionado pelo Índice Geral de Preços - Disponibilidade Interna (IGPDI) e, em seguida, suavizados pela função logarítmica, antes de ser submetida às estratégias preditivas e processadas com suporte do software R.

Esse estudo contribui para aprimorar a capacidade de previsão das receitas do IPVA no contexto brasileiro, fornecendo uma melhor compreensão sobre as técnicas mais adequadas para esse fim. Além disso, ressalta a importância da consideração de fatores sazonais na análise de séries temporais econômicas.

Os resultados revelaram que o modelo SARIMA se destacou como o mais eficaz na previsão das receitas do IPVA. Além disso, o modelo de Holt-Winters multiplicativo também obteve bom desempenho, aproximando-se do SARIMA em termos de precisão para duas partições de dados.

A dissertação também sugere que a aplicação de modelos mais complexos, como o SARIMA, pode melhorar significativamente a qualidade das previsões de receitas tributárias em níveis estaduais e municipais. Esses resultados têm implicações significativas para a gestão fiscal e o planejamento governamental no Brasil.

Palavras-chave: Previsão, Receita, IPVA, Holt-Winters, Aditivo, Multiplicativo, Séries Temporais, Sazonalidade, ARIMA, SARIMA.

## **ABSTRACT**

This work addresses the topic of forecasting the revenues of the Imposto sobre a Propriedade de Veículos Automotores (IPVA) in the state of Goiás. The work focuses on the comparative analysis of different univariate forecasting methods, emphasizing Holt's exponential smoothing techniques (additive and multiplicative) and autoregressive ARIMA and SARIMA. For the study, the collection period between 2003 to 2022 was considered, which was deflated by the Índice Geral de Preços - Disponibilidade Interna (IGPDI) and then smoothed by the logarithmic function, before being subjected to predictive strategies and processed with support from the R software.

This study contributes to improving the forecasting ability of IPVA revenues in the Brazilian context, providing a better understanding of the most appropriate techniques for this purpose. In addition, it highlights the importance of considering seasonal factors in the analysis of economic time series.

The results revealed that the SARIMA model stood out as the most effective in forecasting IPVA revenues. Additionally, the multiplicative Holt-Winters model also performed well, approaching SARIMA in terms of accuracy for two data partitions.

The dissertation also suggests that the application of more complex models, such as SARIMA, can significantly improve the quality of tax revenue forecasts at state and municipal levels. These results have significant implications for fiscal management and government planning in Brazil.

**Keywords:** Forecasting, Revenue, IPVA, Holt-Winters, Additive, Multiplicative, Time Series, Seasonality, ARIMA, SARIMA.

## **LISTA DE SIGLAS**

ITR – Imposto sobre a Propriedade Territorial Rural

IPTU – Imposto sobre a Propriedade Predial e Territorial Urbana

IPVA – Imposto sobre a Propriedade de Veículos Automotores

ITBI – Imposto sobre a Transmissão de Bens Imóveis e de Direitos a eles Relativos

IR – Imposto sobre a Renda e Proventos de Qualquer Natureza

IGF – Imposto sobre Grandes Fortunas

CBT – Código Brasileiro de Trânsito

SENATRAN – Secretaria Nacional de Trânsito

RENAVAM – Registro Nacional de Veículos Automotores

ACF – Função de Autocorrelação

PACF – Função de Autocorrelação Parcial

ACVF – Função de Autocovariância

IPCA – Índice Nacional de Preços ao Consumidor Amplo

INPC – Índice Nacional de Preços ao Consumidor

IGPM – Índice Geral de Preços do Mercado

IPC – Índice de Preços ao Consumidor

IGPDI – Índice Geral de Preços Disponibilidade Interna

MAE – Mean Absolute Error/Erro médio absoluto

MSE – Mean Squared Error/Erro quadrado médio

RMSE – Root Mean Squared Error/Raiz do erro quadrático médio

MAD – Mean Absolute Deviation/Desvio absoluto médio

MAPE – Mean Absolute Percentage Error/Erro percentual absoluto médio

AIC – Akaike Information Criterion/Critério de avaliação de Akaike

BIC – Bayesian Information Criterion/Critério de avaliação bayesiana

HW – Modelo de Holt-Winters

HW-A – Modelo de Holt-Winters Aditivo

HW-M – Modelo de Holt-Winters Multiplicativo

AR – Modelos Autorregressivos

MA – Modelos de Média Móvel

ARIMA - AutoRegressive Integrated Moving Average/

SARIMA – Seasonal AutoRegressive Integrated Moving Average



## SUMÁRIO

<b>1 - Introdução.....</b>	<b>12</b>
<b>2 – Panorama do IPVA .....</b>	<b>15</b>
2.1 – Histórico do IPVA .....	15
2.2 - O imposto e a Tabela FIPE.....	16
2.3 – IPVA no Estado de Goiás .....	18
<b>3 – Fundamentos Teóricos de Séries Temporais .....</b>	<b>24</b>
3.1 – Conceitos básicos.....	24
3.2 – Análise das séries .....	25
3.3 – Estacionariedade e raiz unitária .....	26
3.4 – Funções de autocovariância e autocorrelação .....	28
3.5 – Modelos de Previsão .....	29
3.5.1 – Método de Suavização Exponencial .....	29
3.5.2 – Modelos Autorregressivos .....	34
3.5.3 – Modelos de Média Móvel .....	34
3.5.4 – Modelo ARMA.....	35
3.5.5 – Modelos ARIMA.....	36
3.5.6 – Modelo SARIMA.....	37
3.6 – Funções de avaliação de modelos .....	37
3.6.1 – Erro médio absoluto.....	39
3.6.2 – Erro quadrático médio .....	39
3.6.3 – Raiz do erro quadrático médio .....	40
3.6.4 – Desvio absoluto médio .....	40
3.6.5 – Erro percentual absoluto médio .....	41
3.6.6 – Função de Verossimilhança .....	41
3.6.7 – Critério de avaliação de Akaike (AIC) .....	41
3.6.8 – Critério de avaliação bayesiana (BIC) .....	42
<b>4 – Base de dados utilizada e análise exploratória dos dados .....</b>	<b>44</b>
4.1 – Descrição da série .....	44
4.2 – Sazonalidade .....	48
4.3 – Deflacionando a série.....	50
4.4 – Estacionariedade.....	52
4.5 – Funções de Autocorrelação e Autocorrelação Parcial .....	54
4.6 – Divisão da base de dados .....	56
<b>5 – Análise dos resultados .....</b>	<b>58</b>

<b>5.1 – Análise da 1ª Partição: 2019</b> .....	<b>58</b>
<b>5.2 – Análise da 2ª Partição: 2020</b> .....	<b>63</b>
<b>5.3 – Análise da 3ª Partição: 2021</b> .....	<b>68</b>
<b>5.4 – Análise da 4ª Partição: 2022</b> .....	<b>72</b>
<b>6 – Conclusões</b> .....	<b>77</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	<b>79</b>
<b>Anexo A – Avaliação dos métodos sem deflação</b> .....	<b>82</b>
<b>Anexo B – Avaliação do SARIMA para a série com e sem deflação</b> .....	<b>90</b>

## LISTA DE TABELAS

Tabela 1 – Arrecadação do IPVA, em 2022, em cada estado brasileiro .....	20
Tabela 2 - Distribuição da frota brasileira por Estados em dez/2022 .....	21
Tabela 3 - Valor das alíquotas do IPVA para cada Estado para 2023 .....	22
Tabela 4 - Resultados da investigação da Estacionariedade .....	54
Tabela 5 - Resultados da investigação na série temporal diferenciada .....	54
Tabela 6 - Divisão da base de dados para treinamento e avaliação .....	57
Tabela 7 - Processamento dos modelos na 1ª partição, previsões: jan-dez de 2019.....	60
Tabela 8 - Outras avaliações dos métodos para 1º partição, jan-dez de 2019.....	60
Tabela 9 - Avaliação dos ajustes e resíduos para 1º partição, jan-dez de 2019 .....	61
Tabela 10 - Processamento dos modelos na 2ª partição, previsões: jan-dez de 2020 .....	65
Tabela 11 - Outras avaliações dos métodos para 2º partição, jan-dez de 2020.....	65
Tabela 12 - Avaliação dos ajustes e resíduos para 2º partição, jan-dez de 2020 .....	66
Tabela 13 - Processamento dos modelos na 3ª partição, previsões: jan-dez de 2021 .....	69
Tabela 14 - Outras avaliações dos métodos para 3º partição, jan-dez de 2021.....	69
Tabela 15 - Avaliação dos ajustes e resíduos dos modelos para 3º partição .....	70
Tabela 16 - Processamento dos modelos na 4ª partição, previsões: jan-dez de 2022 .....	73
Tabela 17 - Outras avaliações dos métodos para 4º partição, jan-dez de 2022.....	74
Tabela 18 - Avaliação dos ajustes e resíduos dos modelos para 4º partição .....	75
Tabela 19 - Resultado, sem deflação, na 1ª partição, previsões: jan-dez de 2019.....	82
Tabela 20- Demais medidas comparativas para 1º partição, jan-dez de 2019.....	83
Tabela 21 - Resultado, sem deflação, na 2ª partição, previsões: jan-dez de 2020.....	85
Tabela 22 - Demais medidas comparativas para 2º partição, jan-dez de 2020.....	85
Tabela 23 - Resultado, sem deflação, na 3ª partição, previsões: jan-dez de 2021.....	86
Tabela 24 - Demais medidas comparativas para 3º partição, jan-dez de 2021.....	87
Tabela 25 - Resultado, sem deflação, na 4ª partição, previsões: jan-dez de 2022.....	88
Tabela 26 - Demais medidas comparativas para 4º partição, jan-dez de 2022.....	88
Tabela 27 - Comparação das diferenças absolutas da previsão SARIMA para série temporal com e sem deflação - 2019.....	90
Tabela 28 - Outras métricas avaliativas para o SARIMA - 2019 .....	91
Tabela 29 - Comparação das diferenças absolutas da previsão SARIMA para série temporal com e sem deflação - 2020.....	91
Tabela 30 - Outras métricas avaliativas para o SARIMA - 2020 .....	91
Tabela 31 - Comparação das diferenças absolutas da previsão SARIMA para série temporal com e sem deflação - 2021.....	92
Tabela 32 - Outras métricas avaliativas para o SARIMA - 2021 .....	92
Tabela 33 - Comparação das diferenças absolutas da previsão SARIMA para série temporal com e sem deflação - 2022.....	92
Tabela 34 - Outras métricas avaliativas para o SARIMA - 2022 .....	92

## LISTA DE FIGURAS E QUADROS

Figura 1 - Participação do IPVA na receita do Estado de Goiás .....	19
Figura 2 - Representação gráfica da série temporal do IPVA em valores correntes.....	44
Figura 3 - Distribuição Normal da série temporal .....	45
Figura 4 - Gráficos dos quartis por ano e mês da série temporal .....	46
Figura 5 - Gráficos da Arrecadação do IPVA e linha de tendência suavizada .....	47
Figura 6 - Gráfico do Ruído da Arrecadação do IPVA.....	47
Figura 7 - Gráfico do Sazonalidade da Arrecadação do IPVA.....	48
Figura 8 - Destaque gráfico da Sazonalidade da Arrecadação do IPVA em 1 ano .....	49
Figura 9 - Gráfico da deflação por IGPD e da Arrecadação do IPVA .....	51
Figura 10 - Demonstração da Série Temporal IPVA em escala Logarítmica .....	52
Figura 11 - ACF da Série Temporal IPVA (60 períodos) .....	55
Figura 12 - PACF da Série Temporal IPVA (12 períodos) .....	56
Figura 13 - Gráfico da Arrecadação IPVA x Série em log com Tendências, para a 1ª Partição .....	59
Figura 14 - Divisão dos dados em treino e validação para 1ª Partição .....	59
Figura 15 - Comparação dos modelos x dados de avaliação da 1ª Partição .....	61
Figura 16 - Treinamento do SARIMA para 1ª Partição.....	62
Figura 17 - Resultado do comando summary() sobre o modelo SARIMA - 1ª Partição.....	62
Figura 18 - Representação das previsões futuras para 1ª Partição .....	63
Figura 19 - Gráfico da Arrecadação IPVA x Série em log com Tendências, para a 2ª Partição .....	64
Figura 20 - Divisão dos dados em treino e validação para 2ª Partição .....	64
Figura 21 - Comparação dos modelos x dados de avaliação da 2ª Partição .....	66
Figura 22 - Treinamento do SARIMA para 2ª Partição.....	66
Figura 23 - Resultado do comando summary() sobre o modelo SARIMA - 2ª Partição.....	67
Figura 24 - Representação das previsões futuras para 2ª Partição .....	67
Figura 25 - Gráfico da Arrecadação IPVA x Série em log com Tendências, para a 3ª Partição .....	68
Figura 26 - Divisão dos dados em treino e validação para 3ª Partição .....	68
Figura 27 - Comparação dos modelos x dados de avaliação da 3ª Partição .....	70
Figura 28 - Treinamento do SARIMA para 3ª Partição.....	71
Figura 29 - Resultado do comando summary() sobre o modelo SARIMA - 3ª Partição.....	71
Figura 30 - Representação das previsões futuras para 3ª Partição .....	72
Figura 31 - Gráfico da Arrecadação IPVA x Série em log com Tendências, para a 4ª Partição .....	72
Figura 32 - Divisão dos dados em treino e validação para 4ª Partição .....	73
Figura 33 - Comparação dos modelos x dados de avaliação da 4ª Partição .....	74
Figura 34 - Treinamento do SARIMA para 4ª Partição.....	75
Figura 35 - Resultado do comando summary() sobre o modelo SARIMA - 4ª Partição.....	75
Figura 36 - Representação das previsões futuras para 4ª Partição .....	76
Figura 37 - Comparação dos modelos x série sem deflação da 1ª Partição .....	84
Figura 38 - Comparação dos modelos x série sem deflação da 2ª Partição .....	86
Figura 39 - Comparação dos modelos x série sem deflação da 3ª Partição .....	87
Figura 40 - Comparação dos modelos x série sem deflação da 4ª Partição .....	89
Figura 41 - Erro percentual mensal entre previsões SARIMA e série temporal com e sem deflação...	93
Figura 42 - Erro percentual anual acumulado entre previsões SARIMA e série temporal com e sem deflação.....	94

## 1 - Introdução

A previsão de séries temporais é um campo de estudo que tem atraído a atenção tanto para a economia quanto para análise de dados, pois contribui para tomada de decisões e para a formulação de estratégias em diversas áreas. A popularização das tecnologias da informação e da maturidade computacional disponibilizou conjuntos de dados temporais cada vez mais significativos, e, dessa forma, possibilita a popularização de complexas técnicas preditivas de tendências e comportamentos. Neste contexto, esta dissertação se propôs a investigar e comparar a precisão de diferentes métodos de previsão aplicados a uma série temporal de importância no cenário econômico estadual e municipal: o Imposto sobre a Propriedade de Veículos Automotores (IPVA).

A capacidade de prever com precisão sua arrecadação é fundamental para o planejamento financeiro e a gestão pública. Contudo, essa tarefa não é trivial, uma vez que a série temporal do IPVA apresenta características de sazonalidade, tendência e variação aleatória que a tornam desafiadora para a modelagem e previsão.

Nesse estudo utilizamos modelos e metodologias bem estabelecidos na literatura econômica para analisar a série temporal do Imposto sobre a Propriedade de Veículos Automotores (IPVA) para o estado de Goiás. Com esse intuito, utilizamos os modelos de alisamento exponencial e os modelos sazonais autorregressivos integrados de médias móveis (SARIMA), utilizando o pacote estatístico R como ferramenta de análise.

Alguns trabalhos acadêmicos anteriores adotaram a mesma abordagem. Sousa et al. (2019) utilizaram modelos SARIMA para estimar a arrecadação do ICMS do estado do Ceará entre janeiro e outubro de 2018, relatando a importância da componente sazonal para as previsões. Fontenele (2017) utilizou a metodologia Box-Jenkins para realizar previsões mensais da arrecadação tributária do ICMS e IPVA para o Estado do Ceará, obtendo dois modelos SARIMA como os de melhor acurácia. Junior (2017) apresenta, dentre outros, um modelo ARIMA de previsão para o IPVA do estado do Ceará. Souza (2016) também utilizou a metodologia Box-Jenkins para

realizar previsões para o IPVA do estado de Sergipe, mostrando que um modelo SARIMA entregou as melhores previsões. Rocha (2003) utiliza os modelos de Holt-Winters e SARIMA para efetuar a previsão do ISS do município do Rio de Janeiro. Gomes (2003) utilizou modelos de alisamento exponencial e SARIMA, além de uma combinação de ambos, para estimar as receitas do Imposto de Renda em 2001, utilizando dados de 1994 a 2000. Finalmente, o modelo ARIMA foi usado por Siqueira (2002) para analisar as séries de dez tributos federais no período de janeiro de 1989 a dezembro de 2000, além de fazer comparações das previsões com a arrecadação de 2001, utilizando a raiz quadrada do erro quadrático médio de previsão. Desse modo, Siqueira conclui que o modelo ARIMA escolhido se mostra superior aos outros métodos utilizados na previsão dos tributos federais.

Assim, o objetivo deste trabalho é utilizar modelos de alisamento exponencial, ARIMA e SARIMA para estimar a previsão de receitas do IPVA de Goiás em um horizonte de previsão de doze meses, correspondente a um ciclo orçamentário, e identificar o modelo com a melhor acurácia frente aos diversos períodos econômicos considerados. Ainda que essa pesquisa represente apenas uma etapa do processo de previsão de receitas, ela acaba por sinalizar aos gestores públicos os modelos mais precisos dentro de cada período temporal escolhido.

Este trabalho se divide em seis capítulos incluindo essa introdução. O capítulo dois trata do panorama nacional e estadual do IPVA, o capítulo três aborda os fundamentos teóricos de séries temporais, o capítulo quatro discute a base de dados utilizada e avalia as características importantes para as estratégias escolhidas, o capítulo cinco apresenta a análise dos resultados e, por fim, o capítulo seis apresenta as conclusões. Além desses, há também um anexo que discute a aplicação da metodologia para a série temporal do IPVA não deflacionada.

Esta dissertação oferecerá percepções importantes para pesquisadores, gestores públicos e profissionais da área financeira, destacando o estudo dos dados e a importância da escolha do método de previsão coerente para séries temporais específicas. Como consequência da análise e da comparação detalhada desses métodos, estamos colaborando para o avanço das técnicas de previsão aplicadas à

economia e aos tributos estaduais, visando aprimorar a capacidade de planejamento e gestão das finanças públicas.

## 2 – Panorama do IPVA

### 2.1 – Histórico do IPVA

O imposto sobre patrimônio no Brasil tem uma longa história, que remonta ao início do século XIX, em 1824, através da Lei Orgânica do Império do Brasil, foi instituído o Imposto Territorial<sup>1</sup>, e esse visava à propriedade de terras rurais. Era cobrado pela União e constituía fonte de receita para o governo imperial.

Já no império, o país recém independente, com a nova realidade e necessitando aumentar a receita, em 1857, a Lei 1.732 inaugurou o Imposto Predial, que tributava a propriedade de imóveis urbanos e, gerido pelos municípios.

Em 1891, agora república, a Constituição Federal do Brasil estabeleceu o Imposto de Consumo<sup>2</sup>, que era um imposto cobrado sobre a comercialização de bens e serviços, que era de responsabilidade compartilhada entre estados e municípios.

No Brasil, em 1910 tivemos as primeiras leis referentes ao trânsito. Com o início da produção do automóvel em série no século XX, o país se adaptou a nova realidade e modernizou seu sistema legislativo. Embora o primeiro ato normativo tivesse sido o decreto N. 8.324, de 27 de outubro de 1910, que aprovava o regulamento para o serviço subvencionado de transportes por automóveis, só em 1941 que surgiu o primeiro Código Nacional de Trânsito e ainda assim foi revogado no mesmo ano. E, assim saiu o segundo Código Nacional de Trânsito, que vigorou até o terceiro Código Nacional de Trânsito de 1966<sup>3</sup>.

Foi durante o regime militar que o IPVA foi criado, Lei Federal nº 6.729, de 28 de novembro de 1979, entretanto, apenas seis anos depois, em 1985, através da

---

<sup>1</sup> COELHO, Alvaro Luiz, Um histórico sobre a tributação no Brasil, **Revista da Receita Federal: estudos tributários e aduaneiros**, v. 3, n. 1–2, p. 29–49, 2017.

<sup>2</sup> CAMPELLO, Luiza de Figueiredo, A instituição do imposto de renda no Brasil, 2017.

<sup>3</sup> CONSELHO NACIONAL DE TRÂNSITO (BRAZIL) (Org.), **100 anos de legislação de trânsito no Brasil: 1910-2010**, Brasília: Contran : Denatran : Ministério das Cidades, 2010.



Emenda Constitucional nº 27, também no dia 28 de novembro, foi incluído na carta magna de 1967. Esse dispositivo foi acolhido pela atual Constituição Federal do Brasil, de 1988, em seu artigo 155 e inciso III. Atualmente, vigora a redação da Emenda Constitucional nº 3, de 17 de março de 1993.

## 2.2 - O imposto e a Tabela FIPE

O IPVA é um imposto estadual que incide sobre a propriedade de veículos automotores, sendo calculado com base no valor do veículo e nas alíquotas estabelecidas por cada estado e cada tipo de veículo. Além disso, é um tributo progressivo, pois seu valor está relacionado à cotação do veículo e, dessa forma, proprietários de automotores mais caros serão tributados à maior monta. A motivação disso é justiça social, pois pessoas que possuem mais condições de contribuir participarão com maior parcela para o financiamento dos serviços públicos.

É interessante ressaltar que o IPVA não é exclusivo de carros, motos ou caminhões. Com base no dispositivo legal, art. 155 da CF88, teoricamente pode incidir sobre todo veículo que use propulsão de motores, isso inclui, até mesmo, bicicletas elétricas. Entretanto, o Superior Tribunal de Justiça, bem como o Superior Tribunal Federal, entendem que o IPVA somente pode incidir sobre os veículos terrestres<sup>4</sup>, o que acabou por impedir os estados de tributarem outros tipos de veículos.

Outro aspecto a ser considerada em relação ao fato gerador do IPVA, a propriedade automotiva, além da exigência da renovação a cada ano e da forma da sua utilização do veículo, diz respeito à complexidade de apurar o valor desse bem, considerando as complexidades da nossa economia, além da natural desvalorização do automóvel. Para tratar essa complexidade, a Fundação Instituto de Pesquisas Econômicas (FIPE), entidade privada, ficou responsável por determinar o valor venal de veículos novos e usados, com base na variação dos preços de mercado. Dessa forma, ela mantém e gerencia o compilado dos resultados das suas pesquisas, a

---

<sup>4</sup> **STF isenta barcos e aviões de pagar IPVA**, Jusbrasil, disponível em:

<<https://www.jusbrasil.com.br/noticias/stf-isenta-barcos-e-avioes-de-pagar-ipva/142213>>. acesso em: 11 maio 2023.

Tabela FIPE. Esses registros são atualizados anualmente e, publicadas, geralmente, no início de cada ano. Ela é a referência para o valor do veículo a ser usada no cálculo do IPVA.

O resultado da pesquisa discrimina vários fatores: marca, modelo, ano de fabricação, versão, opcionais e estado de conservação do veículo. Em seguida, realiza-se uma análise estatística desses dados para determinar o valor médio de mercado de cada veículo. Esse método é conhecido como "Método dos Preços Médios Ponderados"<sup>567</sup>, que consiste em atribuir pesos diferentes aos preços dos veículos coletados de acordo com sua representatividade no mercado. Dessa forma, os valores mais frequentes e representativos têm maior peso na média final. É importante ressaltar que o valor venal da Tabela FIPE é uma estimativa baseada em dados reais de mercado, mas pode variar de acordo com fatores econômicos, sazonais e regionais. Além disso, ele não reflete o preço exato de cada veículo, mas sim uma média de mercado que serve como parâmetro para orientar negociações e cálculos de impostos.

Além disso, existem alguns veículos que estão isentos do pagamento do IPVA, como os veículos oficiais, táxis e os de propriedade de pessoas com deficiência ou doença crônica definidas em lei específica.

Por fim, por se tratar de um imposto sabe-se que sua arrecadação é não vinculada, isto é, não há aplicação em um fim específico determinado em lei. Dessa forma, o recolhimento do tributo abastece os cofres públicos, os quais aplicam esse dinheiro para todas as despesas públicas, como, por exemplo, custeios com saúde, educação, segurança e até mesmo manutenção da malha asfáltica. Há ainda outra obrigação sobre o recurso, à repartição em 50% para o estado e 50% para o

---

<sup>5</sup> **Fipe - Fundação Instituto de Pesquisas Econômicas**, Fipe, disponível em: <<https://www.fipe.org.br/>>. acesso em: 6 maio 2023.

<sup>6</sup> BONA, André, **Tabela Fipe: Você sabe como funciona?**, André Bona, disponível em: <<https://andrebona.com.br/tabela-fipe-voce-sabe-como-funciona/>>. acesso em: 15 maio 2023.

<sup>7</sup> REIS, Tiago, **Tabela Fipe: o que é? Como calcular o preço dos automóveis?**, Suno, disponível em: <<https://www.suno.com.br/artigos/tabela-fipe/>>. acesso em: 15 maio 2023.

município, é o caso do Estado de Goiás. A alíquota costuma variar entre 0,5% e 4% sobre o valor do veículo e tem legislação específica para cada estado.

### **2.3 – IPVA no Estado de Goiás**

No Estado de Goiás, a lei que define o IPVA é a Lei Ordinária Nº 11.651, de 26 de dezembro de 1991, que institui o Código Tributário do Estado de Goiás<sup>8</sup> (CTE-GO), no seu Título IV - DO IMPOSTO SOBRE A PROPRIEDADE DE VEÍCULOS AUTOMOTORES – IPVA, entre os artigos 90 e 111, define os fatos geradores, a base de cálculo, as alíquotas, os casos de isenções e não incidência, do responsável, do pagamento e das penalidades.

Em relação às alíquotas, a Seção IV, art. 93 do CTE-GO e atualizada pela Lei nº 13.772, de 28 de dezembro de 2000, subdivide em aplicação para 4 categorias: transporte coletivo (1,25%), motocicletas (3% - atualizada pela Lei Nº 19.021, de 30 de setembro de 2015), veículos utilitários (3,45%), de passeio (3,75%), esses últimos atualizados pela Lei Nº 14.634, de 29 de dezembro de 2003.

Ao longo do período da série temporal, o IPVA participou de forma tímida da receita do Estado de Goiás, como demonstra a Figura 1. Entretanto, ela ele tem aumentado sua participação, em 2003 era 2,72%, mas em 2022 foi de 4,89%, comparativamente, podemos perceber que cresceu 79,91%. Vários são os fatores econômicos que contribuem para percepção do aumento percentual do IPVA na receita do Estado, como: crescimento da economia brasileira que aumenta a frota, montadoras instaladas no Estado, incentivos de consumo pelo governo, juros baixos, o aumento do preço médio dos veículos (evolução da tecnologia embarcada e novas exigências de segurança), aumento da população, o aumento da frota etc.

---

<sup>8</sup> Código Tributário do Estado de Goiás, disponível em: <<https://appasp.economia.go.gov.br/legislacao/arquivos/Cte/CTE.htm>>. acesso em: 5 abr. 2023.

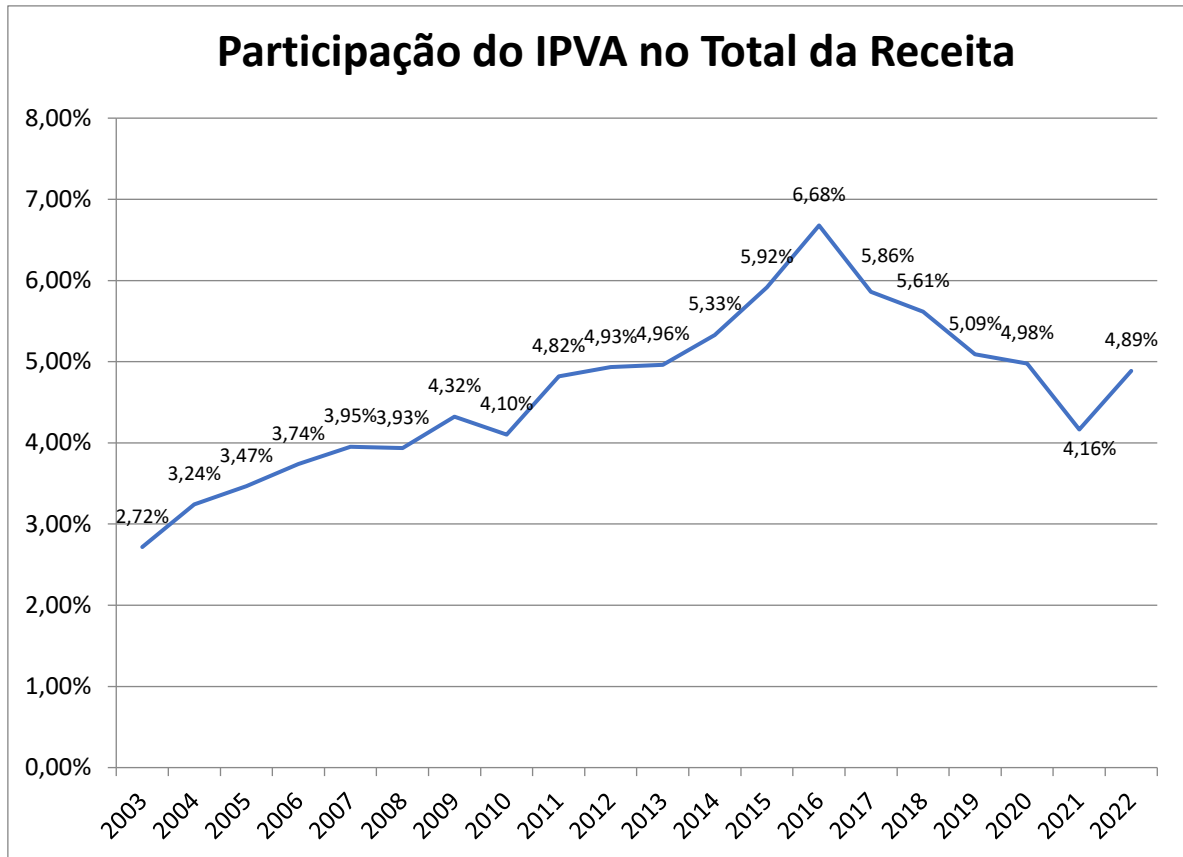


Figura 1 - Participação do IPVA na receita do Estado de Goiás

(fonte: registros administrativos)

Abaixo, apresentamos a Tabela 1 demonstrando o valor total arrecadado pelo IPVA e de todas as arrecadações, em valores correntes, para cada um dos Estados Brasileiros para o ano de 2022. Considerando a ordenação pelo total da arrecadação espontânea para o IPVA e total de impostos, Goiás ocuparia respectivamente as posições de 7º e 8º entre os maiores recolhedores de impostos do país, e em comparação com o país, Goiás contribuiria com 3,76% e 3,64% do valor da tributação total e do IPVA para o país. A Tabela 2, abaixo, considerando veículos habilitados até dezembro de 2022, demonstra a frota total, de automóveis, de motocicletas e outros para cada um dos Estados.

Tabela 1 – Arrecadação do IPVA, em 2022, em cada estado brasileiro

	Arrecadação IPVA		Total Arrecadação	%
<b>Acre</b>	R\$		R\$	5,67%
		116.069.464,73	2.046.763.383,50	
<b>Alagoas</b>	R\$		R\$	7,76%
		509.484.222,85	6.568.897.760,96	
<b>Amazonas</b>	R\$		R\$	3,02%
		477.819.699,79	15.815.290.053,94	
<b>Amapá</b>	R\$		R\$	6,04%
		114.072.694,86	1.887.364.145,21	
<b>Bahia</b>	R\$	1.963.423.780,99	R\$	4,88%
			40.256.913.782,97	
<b>Ceará</b>	R\$	1.480.534.437,21	R\$	7,88%
			18.797.552.782,63	
<b>Distrito Federal</b>	R\$	1.445.468.808,69	R\$	11,94
			12.110.667.312,36	%
<b>Espírito Santo</b>	R\$		R\$	4,97%
		942.576.250,46	18.978.632.642,43	
<b>Goiás</b>	R\$	2.462.605.464,05	R\$	8,30%
			29.676.143.553,38	
<b>Maranhão</b>	R\$		R\$	5,40%
		681.215.771,91	12.617.775.366,05	
<b>Minas Gerais</b>	R\$	7.738.354.642,00	R\$	9,25%
			83.672.919.958,20	
<b>Mato Grosso do Sul</b>	R\$		R\$	5,23%
		932.529.415,68	17.835.855.484,47	
<b>Mato Grosso</b>	R\$	1.238.862.700,90	R\$	5,35%
			23.158.690.029,45	
<b>Pará</b>	R\$		R\$	3,85%
		941.921.353,00	24.455.552.804,74	
<b>Paraíba</b>	R\$		R\$	7,08%
		601.314.499,23	8.498.291.543,96	
<b>Pernambuco</b>	R\$	1.890.446.347,77	R\$	7,70%
			24.553.412.158,48	
<b>Piauí</b>	R\$		R\$	7,19%
		481.646.852,64	6.696.134.615,63	
<b>Paraná</b>	R\$	5.141.380.423,09	R\$	9,86%
			52.128.293.070,85	
<b>Rio de Janeiro</b>	R\$	3.697.426.856,26	R\$	6,52%
			56.704.637.031,03	
<b>Rio Grande do Norte</b>	R\$		R\$	6,47%
		499.306.768,45	7.720.520.442,47	
<b>Rondônia</b>	R\$		R\$	7,68%
		505.156.535,48	6.576.206.918,84	
<b>Roraima</b>	R\$		R\$	4,61%
		98.253.036,34	2.133.413.280,25	
<b>Rio Grande do Sul</b>	R\$	4.657.177.123,96	R\$	9,31%
			50.047.903.912,66	
<b>Santa Catarina</b>	R\$	3.081.924.716,39	R\$	7,56%
			40.755.334.377,88	
<b>Sergipe</b>	R\$		R\$	7,37%
		373.406.168,00	5.069.794.134,00	

<b>São Paulo</b>	R\$ 23.065.575.950,16	R\$ 240.495.579.051,45	9,59%
<b>Tocantins</b>	R\$ 405.164.328,59	R\$ 5.150.626.183,52	7,87%
<b>Brasil</b>	R\$ 65.543.118.313,48	R\$ 814.409.165.781,31	8,05%

Fonte: <https://dados.gov.br/dados/conjuntos-dados/boletim-de-arrecadacao-dos-tributos-estaduais>

Também é característica do cômputo do tributo a particularidade do perfil do veículo, sendo assim, a Tabela 2, abaixo, apresenta a caracterização da frota brasileira. Ainda há a ressalva que agrupei alguns veículos como moto, apesar do Código Brasileiro de Trânsito (CBT)<sup>9</sup> identifica-las em 3 tipos de veículos: as motocicletas, as motonetas e os ciclomotores, e que para a categoria outros a reunião de quaisquer outros veículos que não os já identificados, como por exemplo de transporte coletivo e utilitários. Considerando a participação de Goiás para as frotas, o Estado é responsável por 3,95%, 3,44%, 4,28% e 4,82% do total de frota, de automóveis, de motos e de outros veículos, respectivamente.

*Tabela 2 - Distribuição da frota brasileira por Estados em dez/2022*

<b>Estado</b>	<b>Frota</b>	<b>Automóveis</b>	<b>Motos*</b>	<b>Outros*</b>
<b>Acre</b>	334.377	101.550	175.409	57.418
<b>Alagoas</b>	1.034.187	416.789	434.521	182.877
<b>Amapá</b>	232.691	96.391	91.573	44.727
<b>Amazonas</b>	1.069.794	447.417	420.861	201.516
<b>Bahia</b>	4.887.673	2.095.483	1.795.713	996.477
<b>Ceará</b>	3.625.994	1.295.534	1.787.497	542.963
<b>Distrito Federal</b>	2.021.627	1.384.098	258.535	378.994
<b>Espírito Santo</b>	2.248.960	1.075.778	653.754	519.428
<b>Goiás</b>	4.542.235	2.078.076	1.350.890	1.113.269
<b>Maranhão</b>	2.031.236	518.149	1.209.743	303.344
<b>Mato Grosso</b>	2.568.240	869.703	1.005.916	692.621
<b>Mato Grosso do Sul</b>	1.824.708	825.676	547.875	451.157
<b>Minas Gerais</b>	13.028.883	7.018.707	3.274.795	2.735.381
<b>Pará</b>	2.478.988	717.999	1.316.355	444.634
<b>Paraíba</b>	1.523.167	608.326	694.001	220.840
<b>Paraná</b>	8.575.905	4.898.356	1.677.428	2.000.121
<b>Pernambuco</b>	3.439.164	1.453.963	1.399.494	585.707
<b>Piauí</b>	1.385.426	416.316	757.515	211.595

<sup>9</sup> **L9503Compilado**, disponível em: <[https://www.planalto.gov.br/ccivil\\_03/leis/l9503compilado.htm](https://www.planalto.gov.br/ccivil_03/leis/l9503compilado.htm)>. acesso em: 23 abr. 2023.

<b>Rio de Janeiro</b>	7.475.503	4.857.398	1.402.940	1.215.165
<b>Rio Grande do Norte</b>	1.495.094	629.118	607.376	258.600
<b>Rio Grande do Sul</b>	7.869.630	4.682.229	1.400.821	1.786.580
<b>Rondônia</b>	1.154.287	330.443	591.220	232.624
<b>Roraima</b>	263.345	91.250	114.465	57.630
<b>Santa Catarina</b>	5.974.106	3.294.394	1.291.865	1.387.847
<b>São Paulo</b>	32.293.191	19.632.285	6.540.020	6.120.886
<b>Sergipe</b>	907.388	367.629	391.551	148.208
<b>Tocantins</b>	830.733	256.233	379.991	194.509
<b>Brasil</b>	115.116.53	60.459.290	31.572.124	23.085.118

2

Fonte: <https://www.gov.br/transportes/pt-br/pt-br/assuntos/transito/conteudo-Senatran/frota-de-veiculos-2022>

A seguir trazemos a Tabela 3 que apresenta comparativamente as alíquotas por cada categoria aplicável para cada um dos estados. Apesar da discricionariedade do governo local em definir o valor da alíquota podemos perceber que Goiás possui taxas bastante significativas, quando analisamos comparativamente.

*Tabela 3 - Valor das alíquotas do IPVA para cada Estado para 2023*

<b>Estado</b>	<b>Automóveis</b>	<b>Motos*</b>	<b>Outros*</b>
<b>Acre</b>	2,00%	1,00%	1,00%
<b>Alagoas</b>	3,00%	1,00%	3,25%
<b>Amapá</b>	3,00%	1,50%	3,00%
<b>Amazonas</b>	3,00%	2,00%	3,00%
<b>Bahia</b>	2,50%	2,50%	2,50%
<b>Ceará</b>	2,50%	2,00%	2,50%
<b>Distrito Federal</b>	2,00%	1,00%	2,00%
<b>Espírito Santo</b>	2,00%	1,00%	2,00%
<b>Goiás</b>	3,75%	3,00%	3,45%
<b>Maranhão</b>	2,50%	2,00%	2,50%
<b>Minas Gerais</b>	4,00%	2,00%	3,00%
<b>Mato Grosso</b>	3,00%	2,50%	2,50%
<b>Mato Grosso do Sul</b>	3,00%	2,00%	3,00%
<b>Pará</b>	2,50%	1,00%	2,50%
<b>Paraíba</b>	2,50%	2,50%	2,50%
<b>Paraná</b>	3,50%	3,50%	3,50%
<b>Pernambuco</b>	2,50%	2,00%	2,50%
<b>Piauí</b>	2,50%	2,00%	2,50%
<b>Rio de Janeiro</b>	4,00%	2,00%	3,00%
<b>Rio Grande do Norte</b>	3,00%	1,50%	3,00%
<b>Rio Grande do Sul</b>	3,00%	2,00%	3,00%

<b>Rondônia</b>	3,00%	2,00%	3,00%
<b>Roraima</b>	3,00%	2,00%	2,00%
<b>Santa Catarina</b>	2,00%	1,00%	2,00%
<b>Sergipe</b>	3,00%	2,00%	2,50%
<b>São Paulo</b>	4,00%	2,00%	2,00%
<b>Tocantins</b>	2,00%	2,00%	3,00%

Fonte: <https://www.noticiasautomotivas.com.br/tabela-ipva-2023-de-todos-os-estados/>

Conforme observamos, a arrecadação do IPVA depende de vários fatores, tais como: tamanho da frota, distribuição das categorias e as alíquotas definidas em lei. Além disso, o preço médio dos veículos está sujeito à oferta-demanda, às novas exigências legais que impactam nesse no custo e, ainda as depreciações características do desgaste mecânico e da saída de linha de um modelo ou de uma fabricante. E, por fim, vale ressaltar que a inflação também afeta o comportamento de consumo e o valor dos veículos, e, como mencionado na seção 4.3, os principais indicadores econômicos variaram mais de 300% no período da nossa amostra.



### 3 – Fundamentos Teóricos de Séries Temporais

Essa sessão do trabalho apresenta os recursos conceituais necessários para compreensão sobre séries temporais, modelos de previsão sobre a série e as métricas de avaliação para o modelo.

#### 3.1 – Conceitos básicos

As séries temporais são uma importante área de estudo da matemática e da estatística, amplamente utilizadas em diversas áreas, como economia, finanças, meteorologia, epidemiologia, entre outras, justamente pela possibilidade de previsão de valores futuros a partir de elemento da própria série. Para compreender o conceito de séries temporais, é necessário percorrer um caminho que começa com a definição de conjuntos, de sequências, o estudo de funções e, por fim, o entendimento das séries temporais.

Um conceito fundamental matemático é o de conjuntos e esses, são coleções de elementos distintos e reunidos pela condição de compartilhamento de alguma característica, essa utilizada pela seleção desses integrantes. E, em relação a cardinalidade, isto é, a quantidade de elementos, os conjuntos podem ser classificados entre finitos e infinitos. Para finalidade desse trabalho, só trataremos conjuntos finitos. De forma genérica, o conjunto será denominado por uma letra maiúscula, pode apresentar seus elementos entre chaves ou por uma expressão, por exemplo:  $A = \{2, 4, 6, 8\}$  ou A como o conjunto que agrupa números pares menores que 10.

Outra característica importante dos conjuntos diz respeito à possibilidade ordenação ou não dos seus elementos, isto é, a posição de cada elemento passa a importar e, a simples alteração disposição implica noutro conjunto. Os conjuntos cujos elementos são ordenados serão denominados por sequência. Dessa forma, cada elemento da sequência é identificado pelo seu índice, que é um numeral não nulo que estabelece sua ordem em relação aos outros elementos do conjunto. Em relação ao exemplo anterior, podemos dizer que o algarismo 4 ocupa a segunda posição ou tem índice 2.

As séries temporais tratam de um tipo especial de sequencias, isto é, um conjunto de observações ordenado cujo índice dos elementos refere-se ao instante temporal, que podem ser em intervalos regulares ou não. Dessa forma, podemos dizer que cada observação do conjunto representa uma medida de uma determinada variável em um determinado momento. Assim, podemos dizer que série temporal é um conjunto de observações ordenadas no tempo, que apresentam dependência serial, isto é, dependência entre instantes de tempo, no sentido de que a ordenação dos dados é essencial, uma vez que os dados não podem ser “embaralhados”.

### **3.2 – Análise das séries**

A análise de séries temporais busca identificar comportamentos ao longo do tempo, padrões para as observações, identificação de tendências dos valores e flutuações ao longo do tempo, bem como fazer previsões futuras com base nos dados observados. Graças ao formalismo científico, diversas técnicas estatísticas e modelos matemáticos são aplicadas para analisar e interpretar as séries temporais, permitindo a tomada de decisões e o planejamento adequado em diversas áreas de estudo. Ao estudar séries temporais, é essencial compreender as principais características que a descrevem.

A primeira característica a ser investigada é a tendência e refere-se à direção geral e persistente que uma série temporal segue ao longo do tempo, graficamente falando. Ela destaca a mudança gradual e sistemática na média da série ao longo de períodos de tempo extensos. Em geral, a tendência pode ser ascendente (crescente), descendente (decrecente) ou sem tendência definida.

Os ciclos, por outro lado, são flutuações repetitivas e de duração variável (em geral maior que um ano) em uma série temporal que não são influenciadas por fatores sazonais. Eles podem estar relacionados a fatores econômicos, políticos ou outros fatores exógenos que causam oscilações em um determinado período.

Já a sazonalidade revela a possibilidade de padrões recorrentes e previsíveis ocorrem em intervalos regulares de tempo inferiores a um ano (curto prazo). Esses padrões são influenciados por fatores sazonais, como estações do ano, feriados ou

eventos específicos. A sazonalidade pode afetar o comportamento e a variabilidade da série temporal em períodos específicos.

E, finalmente, o ruído, também conhecido como componente aleatório da série temporal, representa as variações imprevisíveis e não sistemáticas. É resultado de fatores não controláveis e pode mascarar os padrões subjacentes da série. O ruído é o remanescente da série após remover a tendência, a sazonalidade e o ciclo.

### 3.3 – Estacionariedade e raiz unitária

A estacionariedade é uma propriedade estatística essencial que descreve a estabilidade e a consistência das séries ao longo do tempo. Uma série temporal é considerada estacionária quando suas propriedades estatísticas permanecem constantes ao longo do tempo, independentemente do intervalo considerado. Isso significa que a média, a variância e a autocorrelação da série não dependem explicitamente do tempo, assim, a série não apresenta tendências sistêmicas ou padrões sazonais pronunciados, o que a torna mais adequada para análises estatísticas e modelagens econômicas.

Essa condição, de estacionariedade fraca, é fundamental para a aplicação de funções estocásticas em séries temporais econômicas de forma confiável e robusta, possibilitando a obtenção de insights valiosos sobre o comportamento das variáveis econômicas ao longo do tempo e facilitando a interpretação dos resultados obtidos. De forma geral, temos que, dada uma série temporal  $Y_t$ , onde  $Y$  representa um valor aleatório para uma observação e  $t$  o índice ou momento dessa medição, as seguintes propriedades são observadas<sup>10</sup>:

$$\begin{aligned} \text{Média:} & \quad E(Y_t) = \mu \\ \text{Variância:} & \quad \text{var}(Y_t) = E(Y_t - \mu)^2 = \sigma^2 \\ \text{Covariância:} & \quad \gamma_k = E[(Y_t - \mu)(Y_{t+k} - \mu)] \end{aligned}$$

em que  $\gamma_k$  é a covariância na defasagem  $k$ , é a covariância entre os valores de  $Y_t$  e  $Y_{t+k}$ , isto é, entre dois valores de  $Y$  separados por  $k$ .

<sup>10</sup> GUJARATI, Damodar N.; PORTER, Dawn C., **Basic econometrics**, 5th ed. Boston: McGraw-Hill Irwin, 2009.

Por outro lado, a raiz unitária é um conceito essencial no estudo de séries temporais e está intrinsecamente ligada à questão da estacionariedade. Se uma série apresenta uma raiz unitária, isso indica que ela possui uma tendência a se afastar indefinidamente de seu valor médio ao longo do tempo, o que pode resultar em flutuações irregulares e imprevisíveis. Em outras palavras, séries temporais com raiz unitária tendem a ser mais voláteis e imprevisíveis, e isso, dificulta a análise e a modelagem. Assim, a raiz unitária e a estacionariedade estão inversamente relacionadas, sendo que a presença da primeira pode indicar a ausência da segunda.

Nesse trabalho usaremos 3 testes que investigam a estacionariedade através da verificação de tendência à possuir raiz unitária, e são eles: Dickey-Fuller (ADF), Kwiatkowski-Phillips-Schmidt-Shin (KPSS) e Phillips-Perron (PP). Em geral, esses testes examinam o comportamento, cada um da sua forma, das observações em relação à média móvel. Os testes ADF e PP focam na identificação da presença de raiz unitária através da forma como os valores se comportam em relação à média móvel. Noutra perspectiva, o teste KPSS, analisa a variabilidade dos dados para verificar a estacionariedade da série.

Se uma série temporal não for estacionária, os métodos e análises estatísticas preditivas podem ser menos eficazes. A diferenciação é uma abordagem eficaz para remover tendências e padrões de variação não estacionários, resultando em uma série temporal mais homogênea e previsível. Esse método consiste de uma função que calcula a diferença entre dois valores consecutivos da série temporal, dessa forma, considerando que  $y_t$  seja um elemento da série temporal no instante  $t$ , a primeira diferença,  $\Delta y_t$ , é dada pela fórmula:

$$\Delta y_t = y_t - y_{t-1}$$

Ao diferenciar a série, como podemos perceber per fórmula acima, podemos remover a tendência da série temporal e torná-la mais próxima de uma distribuição normal. Entretanto, pode ser necessário realizar a diferenciação uma segunda vez, nesse caso, a fórmula para a segunda diferença,  $\Delta^2 y_t$ , é:

$$\Delta^2 y_t = \Delta y_t - \Delta y_{t-1} = y_t - 2 \cdot y_{t-1} + y_{t-2}$$

Dessa forma, quando não se verifica a estacionariedade, essa técnica desempenha uma etapa importante na preparação dos dados para a aplicação de métodos que dependem dessa condição.

### 3.4 – Funções de autocovariância e autocorrelação

Uma vez caracterizada a série, é conveniente investigar sobre a possibilidade de dependência dos valores observados entre si, internamente na própria série temporal. As funções de autocovariância (COV) e autocorrelação (ACF) são ferramentas estatísticas que nos capacita estudar essa possibilidade de sujeição de valores observados<sup>11</sup>.

A ACF é uma função estocástica que mede a covariância entre os valores de uma série temporal em diferentes pontos no tempo, isto é, entre um elemento da série temporal e outro elemento defasado em relação a ela mesma<sup>12</sup>. A autocorrelação é uma medida da relação linear entre os valores de uma série temporal em diferentes pontos no tempo. Matematicamente, para um atraso  $k$ , a ACF é calculada como a covariância entre  $y_t$  e  $y_{t-k}$ , onde  $y_t$  é o valor de uma observação da série temporal no instante  $t$ :

$$COV(k) = \sum (y_t - \mu)(y_{t-k} - \mu)$$

A ACF, por outro lado, pode ser calculada como:

$$ACF(k) = \frac{COV(k)}{\sigma^2}$$

Além da ACF, também é possível obter a função de autocorrelação parcial (PACF) que examina a dependência direta entre dois elementos da série temporal. É uma medida importante para determinar o número de atrasos para modelos do tipo ARMA que veremos mais adiante. Em relação essa medida, temos:

$$PACF(k) = \frac{COV(k)}{\sqrt{var(y_t|y_{t-1}, \dots, y_{t-k+1}) \cdot var(y_{t-k}|y_{t-1}, \dots, y_{t-k+1})}}$$

onde:

---

<sup>11</sup> *Ibid.*

<sup>12</sup> *Ibid.*; GOMES, Amanda Dos Santos, MODELAGEM E PREVISÃO DA ARRECADAÇÃO DO IMPOSTO DE RENDA NO BRASIL.

$\mu$ : é a média da série temporal até o elemento do instante  $t$

$\sigma^2$ : é o desvio padrão até o elemento do instante  $t$

$var(y_t)$ : é a variância em relação ao elemento do instante  $t$

### 3.5 – Modelos de Previsão

Um dos principais pressupostos para a análise adequada de uma série temporal é a condição de estacionariedade. Essa é uma propriedade estatística crucial que permite a aplicação de técnicas analíticas adequadas e a realização de inferências confiáveis sobre o comportamento da série.

#### 3.5.1 – Método de Suavização Exponencial

A suavização exponencial é o primeiro método preditivo deste trabalho e a sua principal característica é a autodependência, isto é, a sua proposta é utilizar as próprias observações que, depois de computados, predirão valores que serão percebidos em períodos futuros da série temporal.

A metodologia constitui na atribuição de pesos exponenciais e decrescentes aos valores passados da série temporal. A hipótese é que o próximo elemento a ser observado seria pouco diferente do anterior, assim, as observações mais recentes terão um peso maior na previsão do que observações mais antigas. Dessa forma, o valor suavizado previsto será combinação ponderada entre o valor observado mais recente e o valor suavizado anterior.

Uma das virtudes do método de suavização exponencial é a sua simplicidade e facilidade de implementação. Ele se apresenta eficaz em situações em que os dados possuem padrões de tendência e sazonalidade, permitindo que a previsão se ajuste a esses padrões<sup>13</sup>. Por outro lado, se mostra impreciso quando a série possui valores atípicos e flutuações extremas nos dados.

---

<sup>13</sup> NIELSEN, Aileen, **Análise Prática de Séries Temporais: Predição com Estatística e Aprendizado de Máquina**, 1a edição. [s.l.]: Alta Books, 2021.

### 3.5.1.1 – Método de Suavização Exponencial Simples

Esta técnica em particular, foi proposta por Charles C. Holt em 1957, como uma extensão de métodos de médias móveis, como o método da média móvel ponderada. A metodologia se baseia na percepção que as observações recentes são mais significativas na previsão que as antigas. Esse ajuste é alcançado através da atribuição de pesos exponenciais decrescentes aos elementos da série. A fórmula geral é dada por<sup>14</sup>:

$$\hat{y}_t = \alpha \cdot y_t + (1 - \alpha) \cdot \hat{y}_{t-1}$$

onde:

$\hat{y}_t$ : valor previsto

$\hat{y}_{t-1}$ : previsão anterior

$y_t$ : é o valor observado no período t.

$\alpha$ : é o fator de suavização,  $0 < \alpha < 1$ , que controla o peso dado ao valor observado e à previsão anterior.

Como podemos perceber, uma das vantagens dessa estratégia é a sua simplicidade e facilidade de processamento. É um bom processo para séries temporais com padrões de comportamento mais simples, como tendência constante e variações moderadas. No entanto, é ineficaz para contextos de dados complexos como séries sazonais ou mudanças abruptas.

### 3.5.1.2 – Método de Suavização Exponencial Dupla ou de Holt

O método de Suavização Exponencial Duplo ou de Holt, é uma evolução do método de SES, compartilham o mesmo criador, Charles C. Holt em 1957, e acrescenta a complexidade de processar séries com tendências, tanto padrões crescentes quanto decrescentes<sup>15</sup>.

A evolução da metodologia se dá pela estratégia de decomposição da série em dois componentes, daí seu nome, formado por: um nível e outra para tendência. O

---

<sup>14</sup> GUJARATI; PORTER, *Basic econometrics*.

<sup>15</sup> NIELSEN, *Análise Prática de Séries Temporais: Predição com Estatística e Aprendizado de Máquina*.

primeiro componente, o nível, tem a hipótese que o nível da previsão futura considera o nível da observação atual ponderada pela constante  $\alpha$ , juntamente com o nível anterior ajustado pela tendência anterior. Já a composição da tendência. Formalmente teríamos<sup>16</sup>:

$$l_t = \alpha \cdot y_t + (1 - \alpha) \cdot (l_{t-1} + b_{t-1}),$$

onde:

- $l_t$ : nível previsto para o período t
- $\alpha$ : é o fator de suavização do nível
- $y_t$ : é o valor observado no período t
- $l_{t-1}$ : previsão anterior do nível
- $b_{t-1}$ : é a estimativa da tendência anterior

A outra equação, para avaliação da tendência, supõe-se que a diferença entre os níveis atual e anterior ponderada por  $\beta$ , juntamente com a tendência anterior, a fórmula seria:

$$b_t = \beta \cdot (l_t - l_{t-1}) + (1 - \beta) \cdot b_{t-1}$$

onde:

- $b_t$ : é a estimativa da tendência no período t
- $\beta$ : é o fator de suavização da tendência
- $l_t$ : é a previsão do nível para o período t
- $l_{t-1}$ : é a previsão do nível anterior
- $b_{t-1}$ : é a estimativa da tendência anterior

E, finalmente, a próxima previsão,  $\hat{y}_t$ , será a composição desses dois componentes, a saber:

$$\hat{y}_t = l_{t-1} + b_{t-1}$$

A maior complexidade desse modelo é a melhor determinação dos parâmetros  $\alpha$  e  $\beta$  possíveis, essa etapa é crucial para maior acurácia do modelo. Além disso, o modelo não é competente para séries com tendências não lineares ou padrões mais complexos como cenários com mudanças bruscas na tendência ou padrões

---

<sup>16</sup> GUJARATI; PORTER, **Basic econometrics**.



irregulares. Além disso, a estratégia é mais sensível a outliers e variações extremas nos dados.

### 3.5.1.3 – Método de Holt-Winters Aditivo e Multiplicativo

Seguindo a evolução conceitual da técnica, por volta de 1960 e com participação de outros pesquisadores, o modelo de suavização exponencial evoluiu para uma estratégia mais completa com a incorporação do tratamento para séries temporais com padrão de sazonalidade, isto é, séries que possuem variações regulares em ciclos temporais. A nova metodologia é a combinação dos três componentes: nível, tendência e sazonalidade, incorporando a influência de termos passados.

O novo componente, a sazonalidade, é modelada a partir das médias móveis ponderadas para capturar a variação cíclica dos dados ao longo das estações. Outra consequência do novo componente é a necessidade de mais parâmetros, a saber: um  $\alpha$ , que pondera o nível, um  $\beta$ , que ajusta a tendência, e um  $\gamma$ , que acomoda o componente sazonal. Além disso, a componente sazonal pode participar da previsão de duas formas: a primeira, chamada Holt-Winters Aditivo que a sazonalidade é adicionada ao nível e à tendência, e uma segunda, o Holt-Winters Multiplicativo que a sazonalidade multiplica a composição nível e tendência.

O modelo aditivo é ideal para séries temporais que a amplitude das variações sazonais é constante ao longo do tempo, ou seja, as variações sazonais empilham a evolução temporal da série para o nível e a tendência. As fórmulas para o modelo Holt-Winters Aditivo são as seguintes<sup>17</sup>:

$$\begin{aligned} \text{previsão:} & \quad \hat{y}_t = l_{t-1} + b_{t-1} + s_{t-m} \\ \text{nível:} & \quad l_t = \alpha \cdot (y_t - s_{t-m}) + (1 - \alpha) \cdot (l_{t-1} + b_{t-1}) \\ \text{tendência:} & \quad b_t = \beta \cdot (l_t - l_{t-1}) + (1 - \beta) \cdot b_{t-1} \\ \text{sazonalidade:} & \quad s_t = \gamma \cdot (y_t - l_t - b_t) + (1 - \gamma) \cdot s_{t-m} \end{aligned}$$

onde:

$\hat{y}_t$  é a previsão para o período t

---

<sup>17</sup> *Ibid.*

$l_{t-1}$ : é a previsão do nível anterior

$b_{t-1}$ : é a estimativa da tendência anterior

$s_{t-m}$ : é a estimativa do componente sazonal do período  $t - m$ , com  $m$  representando o número de períodos de uma temporada ou ciclo sazonal.

$l_t$ : é a previsão do nível para o instante  $t$

$\alpha$ : é o fator de suavização do nível

$y_t$ : é o valor observado no período  $t$

$b_t$ : é a estimativa da tendência no período  $t$

$\beta$ : é o fator de suavização da tendência

$l_t$ : é a previsão do nível para o período  $t$

$\gamma$ : é o fator de suavização da sazonalidade

Já o modelo multiplicativo, ele representa melhor as séries cuja variação sazonal muda à medida que o nível e tendência da série aumentam. As fórmulas para esse modelo são<sup>18</sup>:

previsão: 
$$\hat{y}_t = (l_{t-1} + b_{t-1}) \cdot s_{t-m}$$

nível: 
$$l_t = \alpha \cdot \frac{y_t}{s_{t-m}} + (1 - \alpha) \cdot (l_{t-1} + b_{t-1})$$

tendência: 
$$b_t = \beta \cdot (l_t - l_{t-1}) + (1 - \beta) \cdot b_{t-1}$$

sazonalidade: 
$$s_t = \gamma \cdot \frac{y_t}{l_t - b_t} + (1 - \gamma) \cdot s_{t-m}$$

onde:

$\hat{y}_t$ : é a previsão para o período  $t$

$l_{t-1}$ : é a previsão do nível anterior

$b_{t-1}$ : é a estimativa da tendência anterior

$s_{t-m}$ : é a estimativa do componente sazonal do período  $t - m$ , com  $m$  representando o número de períodos de uma temporada ou ciclo sazonal.

$l_t$ : é a previsão do nível para o instante  $t$

$\alpha$ : é o fator de suavização do nível

$y_t$ : é o valor observado no período  $t$

$b_t$ : é a estimativa da tendência no período  $t$

$\beta$ : é o fator de suavização da tendência

---

<sup>18</sup> *Ibid.*

$l_t$ : é a previsão do nível para o período  $t$   
 $\gamma$ : é o fator de suavização da sazonalidade

### 3.5.2 – Modelos Autorregressivos

Os modelos autorregressivos (AR) foram elaborados a partir da hipótese que o valor atual na série temporal pode ser determinado a partir de seus valores passados, semelhante à metodologia de alisamento exponencial<sup>19</sup>. Entretanto, aqui o pressuposto é mais simples, o valor previsto é uma combinação linear dos valores anteriores. Além disso, o único parâmetro desse método é denominando  $p$  que constitui quantidade de elementos antigos usados na função adicionados um erro aleatório, temos<sup>20</sup>:

$$\hat{y}_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t$$

onde:

$\hat{y}_t$ : é a previsão para o período  $t$   
 $c$ : é uma constante para o modelo  
 $\phi_i$ : com  $1 \leq i \leq p$ , os coeficientes da autoregressão  
 $y_i$ : com  $1 \leq i \leq p$ , são  $p$  elementos anteriores da série temporal  
 $\varepsilon_t$ : é erro aleatório para o período  $t$ , também conhecido como ruído branco

Esse tipo de estratégia possui várias limitações, entre elas a fragilidade em relação a elementos outliers, além da complexidade de identificar seus parâmetros  $p$  e os coeficientes da autoregressão.

### 3.5.3 – Modelos de Média Móvel

A metodologia de média móvel teve suas raízes nas décadas de 1950 e 1960 e foi motivada pela busca de alternativas aos modelos AR. Apesar desses serem capazes de tratar as autocorrelações da série, são mais susceptíveis a acumular os erros das previsões anteriores após execuções sucessivas. Por isso, surgiu a

---

<sup>19</sup> NIELSEN, *Análise Prática de Séries Temporais: Predição com Estatística e Aprendizado de Máquina*.

<sup>20</sup> GUJARATI; PORTER, *Basic econometrics*.

evolução focada em capturar a dependência temporal através da média ponderada dos erros passados. Dessa forma, os modelos de média móvel pressupõem que valor previsto seria uma função da média ponderada dos valores de erro passados. Assim como para AR, a função é parametrizada e depende da quantidade de elementos passados, nesse caso, de erros. A equação de média móvel é dada por<sup>21</sup>:

$$\hat{y}_t = c + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

onde:

$\hat{y}_t$ : é a previsão para o período t

$c$ : é uma constante para o modelo

$\theta_i$ : com  $1 \leq i \leq q$ , os coeficientes da média móvel

$\varepsilon_i$ : com  $1 \leq i \leq q$ , são  $q$  erros colecionados pela as previsões anteriores

$\varepsilon_t$ : é erro aleatório para o período t, também conhecido como ruído branco

Entre os destaques da estratégia estão sua flexibilidade para capturar dependências temporais complexas, efeitos sazonais, variações não-lineares e, até, quando os resíduos de um modelo autorregressivo exibem autocorrelações significativas. No entanto, ainda há o esforço para determinar o parâmetro  $q$ , que representa a ordem do modelo de média móvel.

### 3.5.4 – Modelo ARMA

O surgimento do modelo ARMA foi uma evolução natural, partiu da combinação dos dois modelos anteriores que formam seu acrônimo: os autorregressivos (AR) e as médias móveis (MA). A abordagem de ARMA surgiu como uma solução para representar séries temporais que exibem autocorrelação significativa e que também são influenciadas por efeitos de médias móveis. Esse modelo foi proposto como uma maneira de capturar tanto a dependência temporal quanto a média ponderada dos erros passados. Dessa forma, ele é formulado como a combinação linear dos termos autorregressivos e de médias móveis, considerando os parâmetros  $p$  e  $q$ . A equação é dada por<sup>22</sup>:

---

<sup>21</sup> *Ibid.*

<sup>22</sup> *Ibid.*

$$\hat{y}_t = c + \sum_{i=1}^p \Phi_i y_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t$$

onde:

$\hat{y}_t$ : é a previsão para o período t

c: é uma constante para o modelo

$\Phi_i$ : com  $1 \leq i \leq p$ , os coeficientes da autoregressão

$y_i$ : com  $1 \leq i \leq p$ , são p elementos anteriores da série temporal

$\theta_j$ : com  $1 \leq j \leq q$ , os coeficientes da média móvel

$\varepsilon_j$ : com  $1 \leq j \leq q$ , são q erros colecionados pela as previsões anteriores

$\varepsilon_t$ : é erro aleatório para o período t, também conhecido como ruído branco

### 3.5.5 – Modelos ARIMA

O modelo ARIMA é, atualmente, uma das abordagens mais poderosas e amplamente utilizadas na análise de séries temporais. Seu desenvolvimento é uma evolução natural do modelo ARMA, potencializado pela ideia de diferenciação da série, transformando-as de não estacionárias para estacionárias. Foi formalizado por George Box e Gwilym Jenkins na década de 1970.

A estratégia consiste em agregar três componentes da série temporal: primeiramente, observamos a dependência linear entre os valores passados da série temporal, modelo AR, em seguida, processa a tendência da série temporal, que representa o I do acrônimo, e, finalmente, o modelo MA computa a variância da série temporal, referente ao modelo MA. Ainda, há se se considerar os seus parâmetros  $p$ ,  $d$  e  $q$ , respetivamente são: a ordem do modelo autorregressivo, o grau de diferenciação e a ordem do modelo de média móvel. O modelo representado pela equação<sup>23</sup>:

$$\hat{y}_t = c + \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} + \varepsilon_t$$

onde:

$\hat{y}_t$ : é a previsão para o período t

c: é uma constante para o modelo

---

<sup>23</sup> *Ibid.*

$\varphi_i$ : com  $1 \leq i \leq p$ , são os parâmetros autorregressivos calculados através dos mínimos quadrados

$y_i$ : com  $1 \leq i \leq p$ , são  $p$  elementos anteriores da série temporal

$\varepsilon_t$ : é erro aleatório para o período  $t$ , também conhecido como ruído branco

### 3.5.6 – Modelo SARIMA

O modelo SARIMA é um método estatístico para modelar e prever séries temporais que possuem sazonalidade, especialização em relação ao ARIMA. Entretanto, esse modelo possui cinco parâmetros, os do modelo anterior acrescido dos parâmetros sazonais característicos:

- $p$ : o número de termos autorregressivos
- $d$ : o número de diferenciações antes de ser modelada
- $q$ : o número de termos médios móveis
- $P$ : o número de diferenciações sazonais
- $Q$ : o número de termos médios móveis sazonais

O modelo pode ser definido pela seguinte fórmula<sup>24</sup>:

$$\hat{y}_t = c + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=1}^q \theta_j B_{t-j} + \varepsilon_t$$

onde:

$\hat{y}_t$ : é a previsão para o período  $t$

$c$ : é uma constante para o modelo

$\phi_i$ : com  $1 \leq i \leq p$ , os coeficientes da autoregressão

$\theta_j$ : com  $1 \leq j \leq q$ , os coeficientes da média móvel

$B_j$ : com  $1 \leq i \leq q$ , é o operador de diferenciação sazonal

$\varepsilon_t$ : é erro aleatório para o período  $t$ , também conhecido como ruído branco

### 3.6 – Funções de avaliação de modelos

Além dos conceitos necessários para analisar e caracterizar uma série temporal e também conhecer sobre os modelos que serão executados, é necessário, pelo

---

<sup>24</sup> *Ibid.*

objetivo desse trabalho, estabelecer os métodos e critérios que serão utilizados para comparar os resultados das previsões processados. Esta etapa é de fundamental importância para a avaliação do desempenho dos modelos estudados, fornecendo uma perspectiva crítica sobre a qualidade das previsões obtidas e sua adequação aos contextos de previsão de receita do IPVA. Conhecer a aplicação apropriada das medidas de acurácia constitui importante papel sobre a validação e na tomada de decisões baseada nas previsões.

A primeira formalização para esse tema trata-se do erro, que representa a diferença entre um valor observado,  $y$ , e o valor preditivo,  $\hat{y}$ . Assim, consideramos uma série temporal com  $n$  observações, o erro previsto para uma observação  $t$ ,  $t < n$ , é dada por:

$$e_t = y_t - \hat{y}_{t|t-1}$$

onde:

$e_t$ : é a medida do erro percebido quando predito no período  $t$

$y_t$ : é o valor da série no instante  $t$  e reservado determinar acurácia do modelo

$\hat{y}_{t|t-1}$ : é o valor predito pelo modelo para o instante  $t$  e usando a série até o elemento anterior, isto é,  $t - 1$ .

A acurácia, como segundo conceito apresentado, refere-se à medida de quão próximas às previsões ou estimativas estão dos valores reais ou observados. Essa medida aponta para capacidade de um modelo ou método em fornecer resultados precisos e próximos das observações separadas para comparação. Além disso, quanto menor for o valor da métrica de acurácia, melhor será o desempenho do modelo, indicando uma maior precisão nas previsões.

No decorrer destas sessões, serão apresentadas várias métricas de avaliação que capturam diferentes aspectos da acurácia do modelo preditivo. A análise detalhada de medidas tradicionais, como o erro médio absoluto (MAE), erro quadrático médio (MSE) e a raiz do erro quadrático médio (RMSE), fornecerá uma base sólida para compreender a magnitude das discrepâncias entre as previsões e os valores reais. Além disso, medidas de precisão relativa, como a desvio absoluto médio (MAD),

erro percentual absoluto médio (MAPE), serão discutidas, considerando a importância de avaliar a acurácia das previsões em termos percentuais.

### 3.6.1 – Erro médio absoluto

Uma vez compreendido o conceito de erro e considerando que as previsões podem estimar valores além do próximo período, isto é, podemos estimar 3, 6 ou até 12 períodos futuros, parece ser natural considerar a média absoluta de cada um dos erros entre valores propostos e para separados para avaliação como critério de investigação de acurácia da técnica. Bem, é essa a ideia do MAE, dado pela fórmula a seguir<sup>25</sup>:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

onde:

$y_i$ : é o valor observado da posição  $i$  na série temporal.

$\hat{y}_i$ : é o valor previsto para  $i$ -ésima posição da série temporal.

$n$ : é o número de elementos da série temporal.

### 3.6.2 – Erro quadrático médio

Outra técnica comum em estatística é investigar as diferenças ao quadrado, dessa forma, teremos uma medida mais sensível às divergências de valores, isso em relação ao MAE. Além disso, por característica do termo quadrático, sempre teremos valores positivos que, somados, ressaltam a qualidade ou não do modelo. Assim, seu cálculo seria<sup>26</sup>:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

onde:

$y_i$ : é o valor observado da posição  $i$  na série temporal.

$\hat{y}_i$ : é o valor previsto para  $i$ -ésima posição da série temporal.

---

<sup>25</sup> *Ibid.*

<sup>26</sup> *Ibid.*



$n$ : é o número de elementos da série temporal.

### 3.6.3 – Raiz do erro quadrático médio

Apesar das vantagens da utilização da MSE, ela eventualmente produz valores muito altos, além de estar em escala diferente das observações da série temporal. Dessa forma, é justificável utilizar a raiz quadrada da MSE<sup>27</sup>, assim:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

onde:

$y_i$ : é o valor observado da posição  $i$  na série temporal.

$\hat{y}_i$ : é o valor previsto para  $i$ -ésima posição da série temporal.

$n$ : é o número de elementos da série temporal.

### 3.6.4 – Desvio absoluto médio

Realizar operações de potenciação e radiciação, eventualmente, pode acrescentar para cada simulação um aumento significativo de operações computacionais tornando a avaliação do modelo um pouco mais custoso, assim, em geral, é possível realizar comparação dos modelos utilizando outro mecanismo matemático: a média das diferenças absolutas entre os valores reais e os valores previstos, mas sem considerar o sinal das diferenças, assim a equação pode ser representada por, será a seguinte operação para avaliação dos modelos:

$$MAD = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)$$

onde:

$y_i$ : é o valor observado da posição  $i$  na série temporal.

$\hat{y}_i$ : é o valor previsto para  $i$ -ésima posição da série temporal.

$n$ : é o número de elementos da série temporal.

---

<sup>27</sup> *Ibid.*

### 3.6.5 – Erro percentual absoluto médio

Até o momento, os processos matemáticos usados calculam um valor de divergência entre valores esperados e observados. Por outro lado, não foi possível avaliar a magnitude do erro em relação aos valores da série temporal. O cálculo do MAPE resolve esse problema apresentando a relação entre erro e valores da própria série de forma percentual<sup>28</sup>. Para realizar o seu cálculo, pode-se utilizar a seguinte fórmula:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\%$$

onde:

$y_i$ : é o valor observado da posição  $i$  na série temporal.

$\hat{y}_i$ : é o valor previsto para  $i$ -ésima posição da série temporal.

$n$ : é o número de elementos da série temporal.

### 3.6.6 – Função de Verossimilhança

A verossimilhança é outra medida que explicita a aderência do modelo estatístico aos dados observados, dessa forma, o maior valor calculado da verossimilhança está diretamente associado à probabilidade dos dados serem gerados pelo modelo, e, que por consequência, menor erro. É um método iterativo que busca identificar os parâmetros que maximizam o valor da verossimilhança<sup>29</sup>. A fórmula para o cálculo da verossimilhança depende de qual modelo específico será avaliado. Além disso, é uma função distribuição acumulada.

### 3.6.7 – Critério de avaliação de Akaike (AIC)

O AIC é a primeira medida estatística apresentada neste trabalho que serve pra avaliar a estratégia do modelo econométrico e, dessa forma, mensurar a sua qualidade. É baseado na observação que um bom modelo deve ajustar bem os dados enquanto minimiza a complexidade. Para isso, a estratégia é combinar as duas

---

<sup>28</sup> HYNDMAN, Ron, *Forecasting: Principles and Practice (3rd ed)*, [s.l.: s.n.], 2018.

<sup>29</sup> *Ibid.*

características, por um lado uma métrica da qualidade do ajuste do modelo aos dados, através da função de verossimilhança, e, ao mesmo tempo, avalia a complexidade do modelo penalizando-o na mesma proporção que utiliza mais parâmetros. Esse critério de avaliação foi desenvolvido por Hirotugu Akaike em 1974 e, foi definido como<sup>30</sup>:

$$AIC = 2k - 2 \log L,$$

onde:

k: é o número de parâmetros

L: é a verossimilhança maximizada

n: é o tamanho da amostra

A interpretação do resultado AIC deve ser utilizado para comparar os valores entre diferentes modelos e, no caso, aquele que possuir o menor resultado implica ser o melhor. No entanto, é importante notar que o AIC não fornece um valor absoluto, mas que a avaliação está na comparação de valores entre diferentes modelos.

### 3.6.8 – Critério de avaliação bayesiana (BIC)

O BIC é outra medida estatística e, assim como o AIC, serve para avaliação e comparação entre modelos preditivos. A saber, foi desenvolvido por Gideon Schwarz em 1978 como uma evolução do modelo AIC buscando uma abordagem mais rigorosa para balancear a complexidade da estratégia. Na sua proposição, modelos com mais parâmetros precisam ser mais penalizados, pois, a hipótese é que um processo que demanda mais parâmetros é conseqüentemente mais complexo e, portanto, propenso a divergir do padrão dos dados e em proporção do tamanho da amostra. Assim, o BIC é calculado como a soma do logaritmo da verossimilhança, do número de parâmetros do modelo e do logaritmo do tamanho da amostra, terá<sup>31</sup>:

$$BIC = k \log n - 2 \log \hat{L},$$

onde:

k: é o número de parâmetros

$\hat{L}$ : é a verossimilhança maximizada

---

<sup>30</sup> GOMES, MODELAGEM E PREVISÃO DA ARRECAÇÃO DO IMPOSTO DE RENDA NO BRASIL.

<sup>31</sup> *Ibid.*

$n$ : é o tamanho da amostra

O BIC é interpretado da mesma forma que o AIC, quanto menor o valor calculado, melhor é o modelo. Todavia, BIC é mais robusto ao excesso do ajuste de treinamento que o AIC, pois é considera no seu cômputo a quantidade de parâmetros.

## 4 – Base de dados utilizada e análise exploratória dos dados

### 4.1 – Descrição da série

Os dados utilizados nesse trabalho são provenientes de registros administrativos da Secretaria de Estado da Economia. São informações dos pagamentos espontâneos realizados pela data do recebimento bancário e agregados mensalmente, e em valores correntes. A série temporal é composta por 240 registros compreendidos entre janeiro de 2003 a dezembro de 2022 e está apresentada na Figura 2, abaixo.

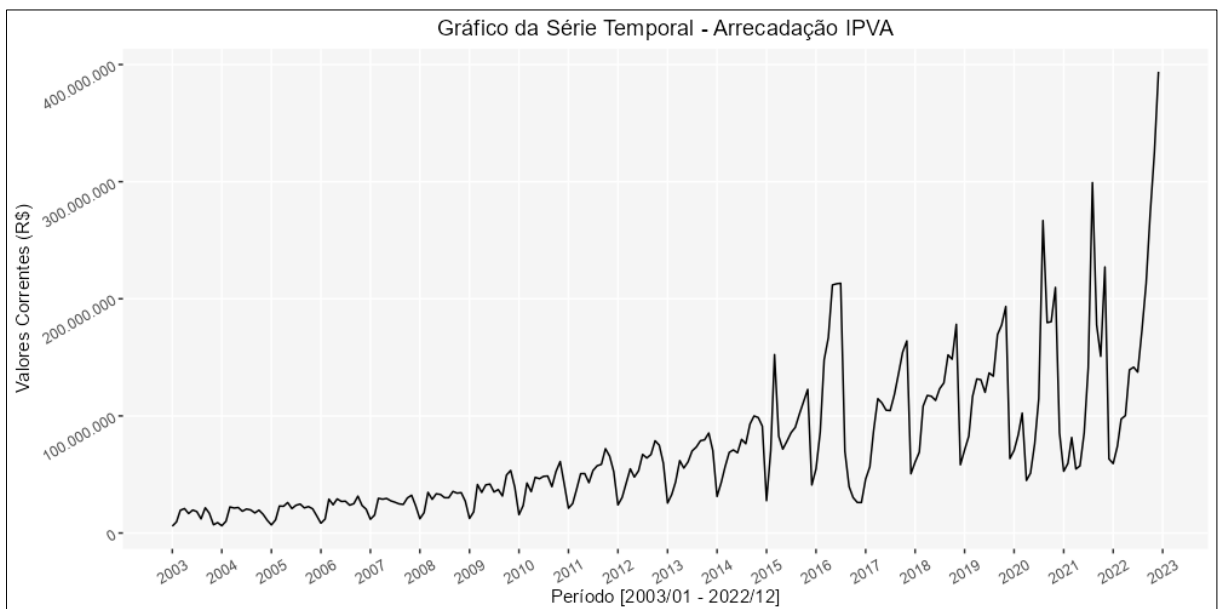


Figura 2 - Representação gráfica da série temporal do IPVA em valores correntes

(fonte: elaboração própria)

Começaremos a análise da série temporal a partir da estatística descritiva dos valores correntes originais, isto é, sem deflação IGPDI. O menor valor observado foi de R\$ 5.881.500,35, percebido em janeiro de 2003, enquanto o maior valor arrecadado foi de R\$ 393.841.549,30 em dezembro de 2022. Esse período, a amplitude de R\$ 387.960.048,95 significou aumento de mais de 6.596,28% de arrecadação ao longo desses 9 anos.

A média mensal de arrecadação é de R\$ 70.312.263,20, mediana de R\$ 52.599.526,55, e desvio padrão de R\$ 61.217.317,23. Observa-se que a mediana da série temporal analisada é inferior à média, sugerindo uma distribuição assimétrica

dos valores, com uma cauda longa negativa, além disso, o desvio padrão significativo reforçando a percepção de assimetria dos valores, conforme demonstra a Figura 3, na qual destacamos sequencialmente, da esquerda para a direita, as linhas verticais: em rosa o 1º quartil, em verde a mediana, em vermelho a médio e em laranja o 3º quartil.

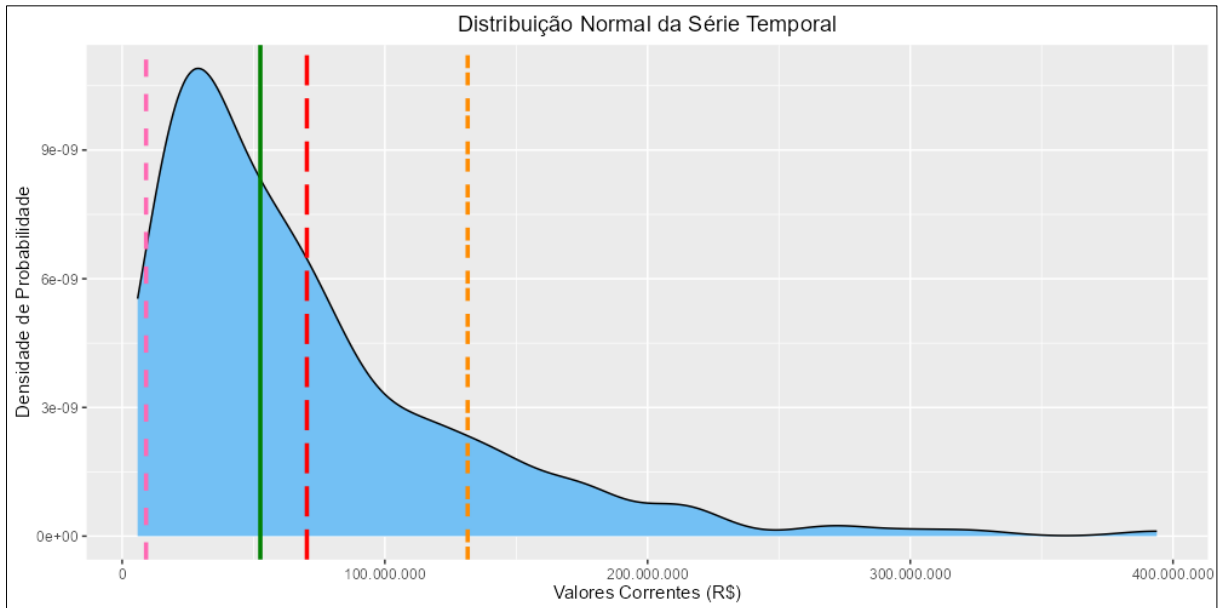


Figura 3 - Distribuição Normal da série temporal

(fonte: elaboração própria)

Além de observarmos o “desbalanceamento” da série, podemos investigar como se comportou os valores em suas amplitudes ao longo dos anos e dos meses, para isso, temos abaixo na Figura 4, dois gráficos do tipo bloxplots que possibilitam interessantes observações. Para a primeira delas, temos que os anos 2016, 2020, 2021 e 2022 a amplitude do valor arrecadado aumentou desproporcionalmente, e a outra, que em relação ao comportamento dos contribuintes, o pagamento de IPVA se acumula no final do ano, preferencialmente em novembro. Investigando os fatores que justificariam o aumento nos anos referidos, temos que: em relação a 2016, além de aspectos macroeconômicos nacionais, a inauguração da fábrica da Nissan em Goiás em 2014. Esse evento é significativo, pois com a novidade da produção local estimulou o consumo de veículos, e conforme demonstra o Instituto Mauro Borges (IMB) entre 2010 e 2016 a atividade do setor industrial de transformação aumentou 42,15% e a atividade de comércio/reparação de veículos 71,80%. Já em relação aos anos finais da série, 2020 à 2021, cabe relembra que vínhamos de uma pandemia em 2019, COVID-19, que parou a cadeia produtiva. Nos anos seguintes, a inflação de demanda

e a perspectiva de retomada de crescimento impulsionou o consumo desse tipo de bens e potencializou a arrecadação. Já o aumento de consumo no final do ano se justifica pela forma oportuna que o Estado de Goiás proporciona aos contribuintes que podem parcelar seus pagamentos, desde que, não ultrapassem para o exercício seguinte.

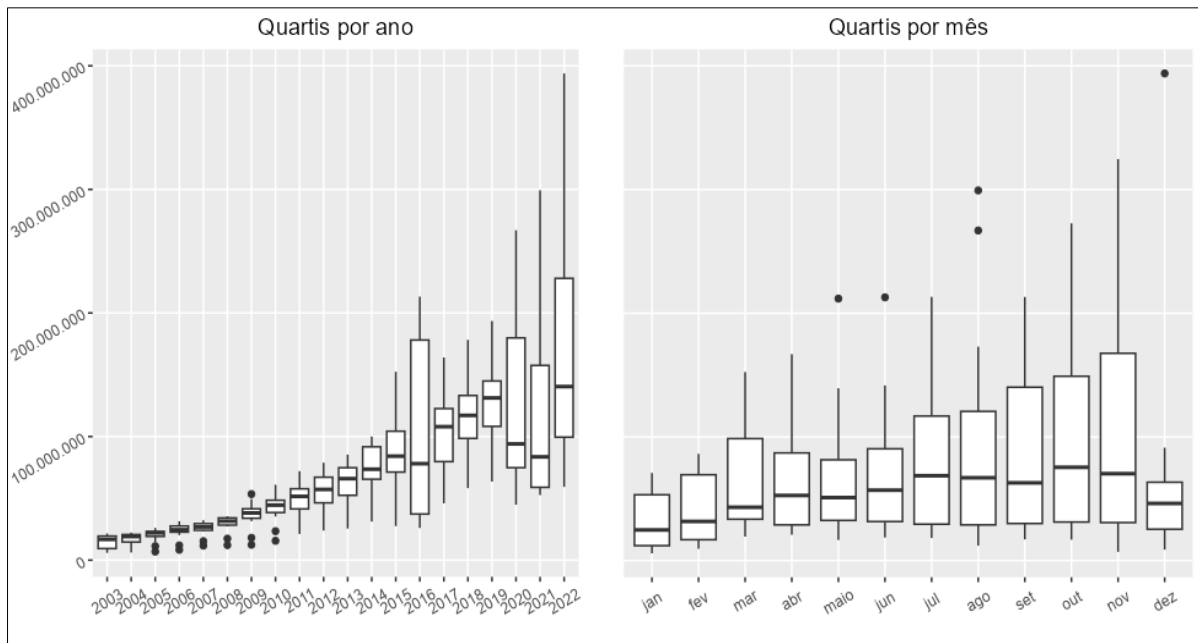


Figura 4 - Gráficos dos quartis por ano e mês da série temporal

(fonte: elaboração própria)

Além da caracterização estatística proporcionada pelos valores arrecadados, há, também, informações importantes em relação das características de série temporal. Ao decompô-la, seus componentes ficam mais nítidos, a saber: a tendência, a sazonalidade e a aleatoriedade ou ruído branco. A investigação sobre eles colabora na interpretação dos resultados pelos modelos preditivos. O software R possui a função “*decompose()*”<sup>32</sup> que desagrega esses componentes possibilitando análise em destaque de cada um. A Figura 5 demonstra a linha de tendência suavizada, em vermelho, e os valores arrecadados do IPVA, tracejada e em azul, para o período, e é importante observar que a tendência se apresenta com característica ascendente.

<sup>32</sup> FERREIRA, Pedro, Análise de Séries Temporais em R.

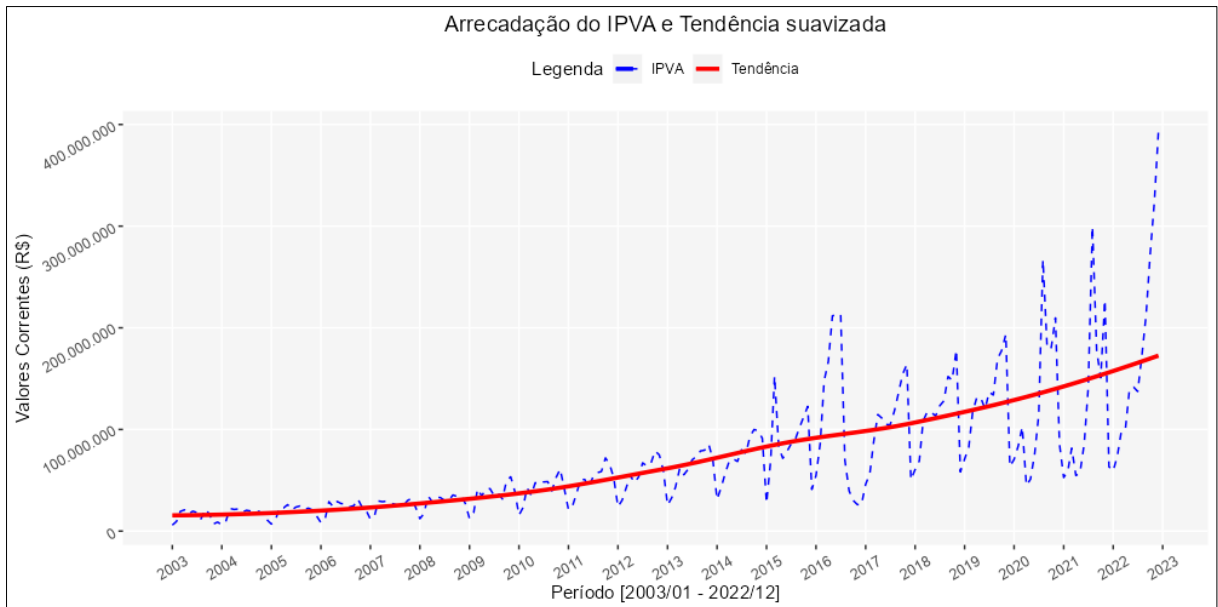


Figura 5 - Gráficos da Arrecadação do IPVA e linha de tendência suavizada

(fonte: elaboração própria)

Além da tendência, há também o componente da aleatoriedade ou ruído da série temporal. Essa representa a parte não previsível e estocástica dos dados, ou seja, as flutuações que não podem ser explicadas por padrões ou tendências claras. Na Figura 6 podemos perceber essa imprevisibilidade, nos primeiros anos, a curva assemelha a “ondas” com bases largas, em seguida, lá pelo ano 2011 já lembra os gráficos de alta frequência, e, ainda, já em 2015, começam as cristas cada vez maiores. Também podemos perceber componentes positivos e negativos, entretanto, o eixo x não serve como eixo de simetria.

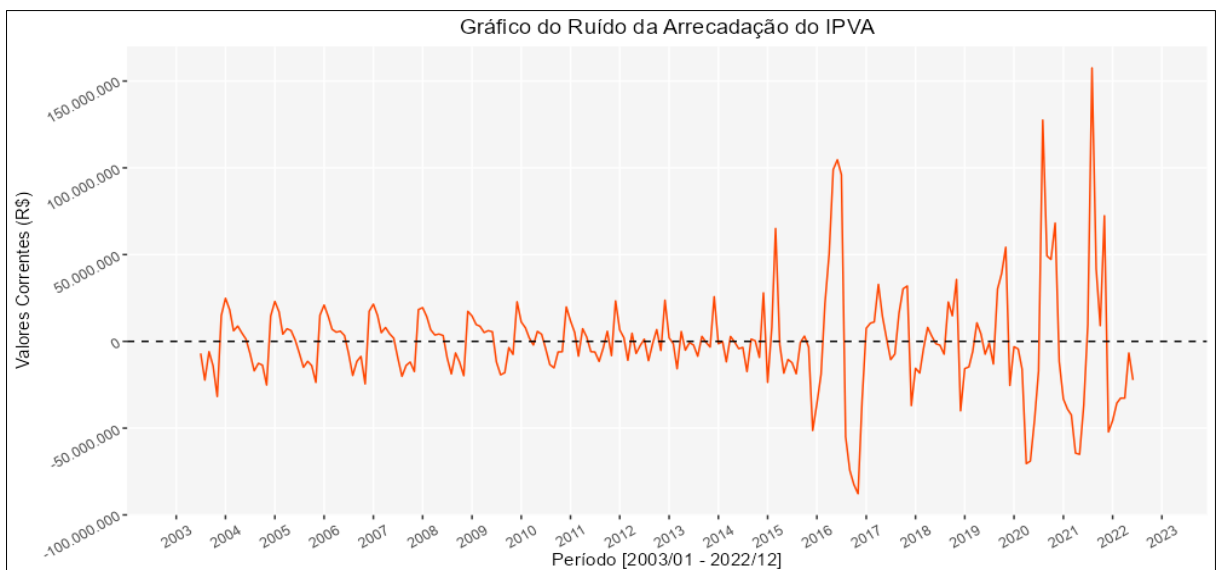


Figura 6 - Gráfico do Ruído da Arrecadação do IPVA

(fonte: elaboração própria)



## 4.2 – Sazonalidade

Além da tendência e do ruído, existe outro componente essencial para a compreensão e análise das séries temporais: a sazonalidade. Essa parte trata dos padrões regulares e repetitivos que ocorrem em um determinado período, geralmente ao longo do ano, trimestre, mês ou qualquer outro intervalo de tempo relevante e curto. Esses padrões podem ser impulsionados por fatores sazonais, como feriados, estações do ano ou comportamentos cíclicos previsíveis em uma série temporal. Uma vez identificada, é possível isolar os efeitos sazonais da série temporal e, dessa forma, visão mais precisa do comportamento dos dados. A Figura 7 abaixo apresenta a componente sazonal extraída pela função “`decompose()`” do software R.

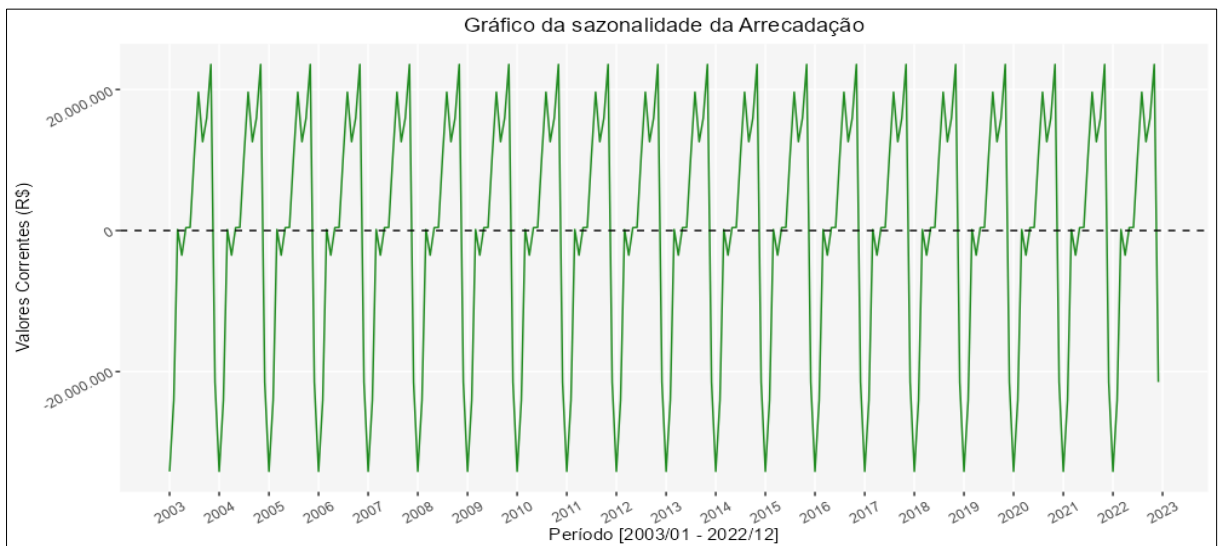


Figura 7 - Gráfico do Sazonalidade da Arrecadação do IPVA

(fonte: elaboração própria)

Ainda vale a pena destacar para conhecimento sobre a arrecadação do IPVA seu comportamento sazonal. Conforme já foi apresentado na Figura 4 e agora detalhado na Figura 8 que há maior recolhimento próximo do final do ano, outubro e novembro, e bem menores no início, janeiro e fevereiro, e ainda no último mês. Essa percepção possibilita organização da administração pública uma vez que é característico esse recolhimento. Esse comportamento da série se justifica por dois motivos, o imposto pode ser pago de forma parcelada, o que acumula parcela em meses seguintes e, o imposto de um exercício precisa ser recolhido dentro do período.

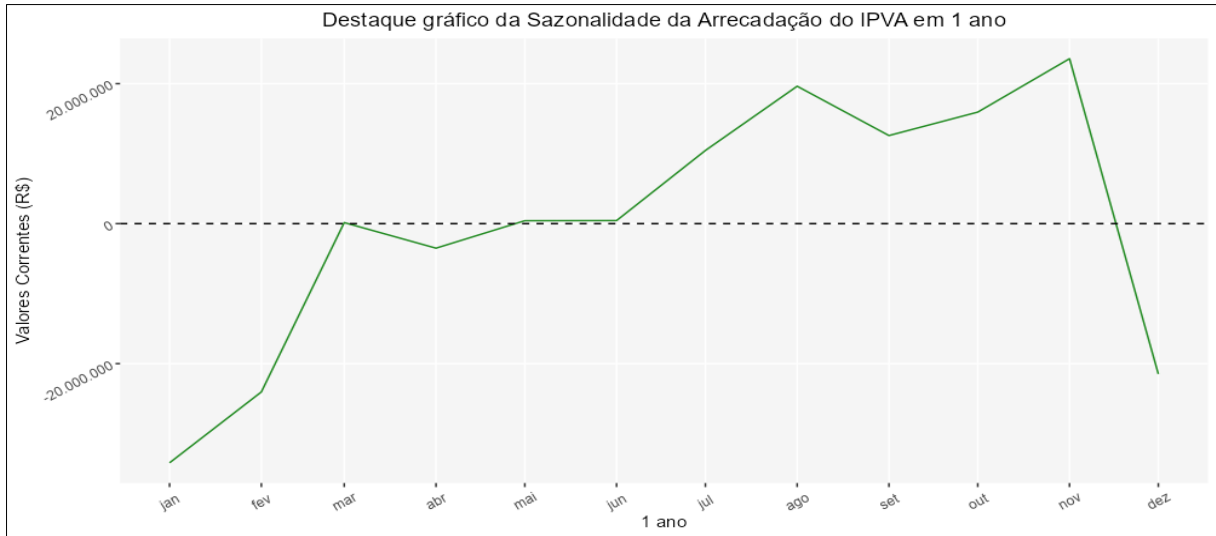


Figura 8 - Destaque gráfico da Sazonalidade da Arrecadação do IPVA em 1 ano

(fonte: elaboração própria)

Além da interpretação gráfica existem várias alternativas matemáticas, também disponíveis no software R, que investigam a sazonalidade da série. Para esse estudo, fizemos uso do Teste de Kruskal-Wallis implementado como “*kruskal.test()*”<sup>33</sup> em R. Esse teste não paramétrico é usado para avaliar se as medianas de diferentes períodos sazonais são iguais. Se houver diferenças significativas nas medianas, isso pode sugerir sazonalidade, assim a Hipótese Nula (H0): as amostras dos grupos vêm de populações idênticas (não há diferença nas medianas). Ele é fundamentado na distribuição das categorias dos elementos como uma transformação que torna os elementos independentes do tipo de distribuição. É calculado usando a seguinte fórmula<sup>34</sup>:

$$H = \left[ \frac{12}{n(n+1)} \cdot \left( \sum_{i=1}^k \frac{R_i^2}{n_i} \right) \right] - 3(n+1)$$

onde:

$n$ : é o tamanho da série temporal

$n_i$ : é o número de observações na amostra  $i$ .

$R_i$ : é a soma dos postos da amostra  $i$ .

$k$ : é número de categorias

<sup>33</sup> *Ibid.*

<sup>34</sup> **Teste de Kruskal-Wallis | PDF | Distribuição Qui Chi | Probabilidade e estatística**, Scribd, disponível em: <<https://pt.scribd.com/document/371557797/Teste-de-Kruskal-wallis>>. acesso em: 11 set. 2023.

Para a arrecadação do IPVA, os resultados obtidos para o valor estatístico é de 26,41, o valor crítico de 7,81 com p-valor de 0,0056 e, considerando o nível de significância de 0,05, concluímos forte evidência da hipótese H0 e, por conseguinte, a série é sazonal.

### 4.3 – Deflacionando a série

Considerando que a amplitude da série temporal da amostra, 9 anos, percebe-se que apenas a substituição da frota por veículos de preço mais altos é insuficiente pra justificar essa variação. Dessa forma, a arrecadação também foi impactada pelos reajustes inflacionários que a economia brasileira passou. A ausência de um índice específico para veículos estimulou a avaliação dos seguintes referenciais econômicos: Índice Nacional de Preços ao Consumidor Amplo (IPCA), Índice Nacional de Preços ao Consumidor (INPC), Índice Geral de Preços ao Mercado (IGPM), Índice de Preços ao Consumidor (IPC) e Índice Geral de Preços - Disponibilidade Interna (IGPDI); e a **Erro! Fonte de referência não encontrada.** apresenta as inflações calculadas para eles e usando a função “*deflate()*”<sup>35</sup> deles. Entre esses, o que possui maior aderência à Tabela Fipe é o IGPDI.

Ele é calculado pela Fundação Getúlio Vargas, entidade reconhecida, e mede a variação dos preços de bens e serviços comercializados no mercado interno, assim como os veículos, o que justifica a escolha. Por outro lado, os demais foram descartados, pois, tanto o INPC, o IPC e o IPCA são restritos, respectivamente, à 5, 10 e 40 salários mínimos, enquanto o IGPM incorpora os preços de produtos e serviços que extrapolam o consumo, como bens de capital e serviços financeiros.

*Inflação estimada 01/2003 e 12/2022  
(Diferentes índices)*

Índice	Inflação
IPCA	217,39%
INPC	219,34%
IGPM	328,62%
<b>IGPDI</b>	<b>322,33%</b>
IPC	204,31%

(fonte: <http://ipeadata.gov.br>)

<sup>35</sup> FERREIRA, Análise de Séries Temporais em R.

A Figura 9 abaixo apresenta uma comparação entre valores originais, em vermelho, e valores deflacionados por IGPD, em azul.

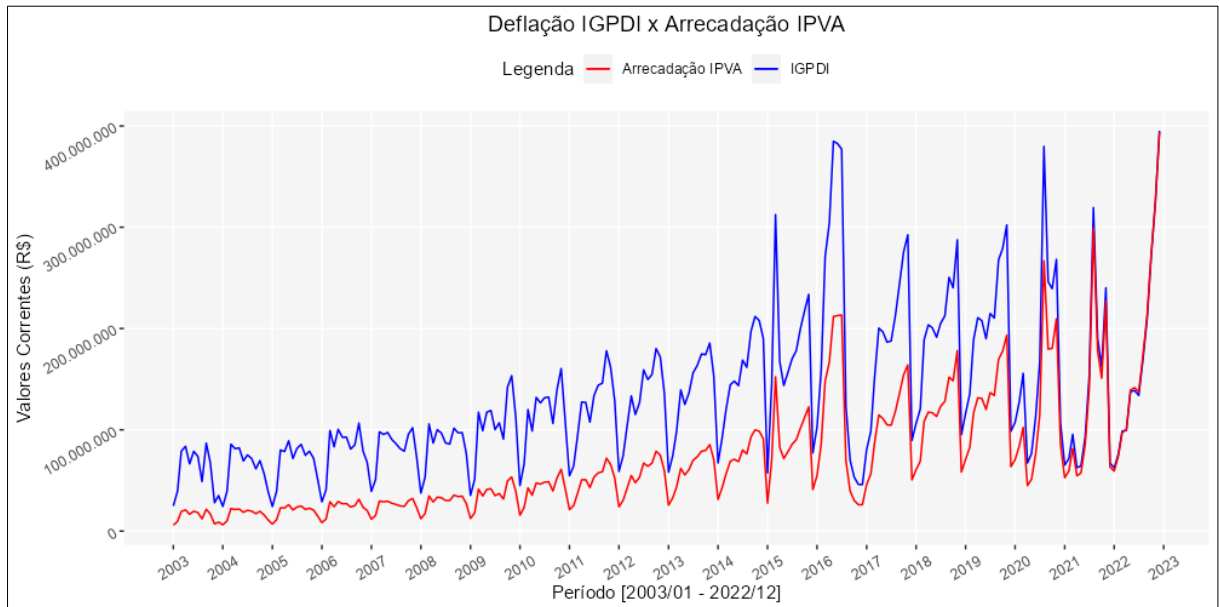


Figura 9 - Gráfico da deflação por IGPD e da Arrecadação do IPVA

(fonte: elaboração própria)

Apesar de adotarmos a série deflacionada pelo IGPD, os resultados da estatística descritiva ainda demonstram que a série ainda é bastante assimétrica, conforme valores descritivos: arrecadação mínima R\$ 17.435.130,11 em 01/2004, valor máximo arrecadado de R\$ 398.377.292,81 em 12/2022, a amplitude de R\$ 380.942.162,70, média mensal de R\$ 104.814.506,54, mediana R\$ 88.947.491,06 e desvio padrão R\$ 64.317.369,59.

Diante disso, adotaremos os valores em escala logarítmica da série deflacionada IGPD, que possuem os seguintes valores descritivos: valor mínimo 16,674 referente à 01/2004, máximo 19,803 referente à 12/2022, com amplitude de 3,129, média: 18,296 e mediana: 18,303 com desvio padrão: 0,595, e, por conseguinte, possui uma dispersão menor entre os valores. Outra forma de perceber isso é a Figura 10 abaixo.

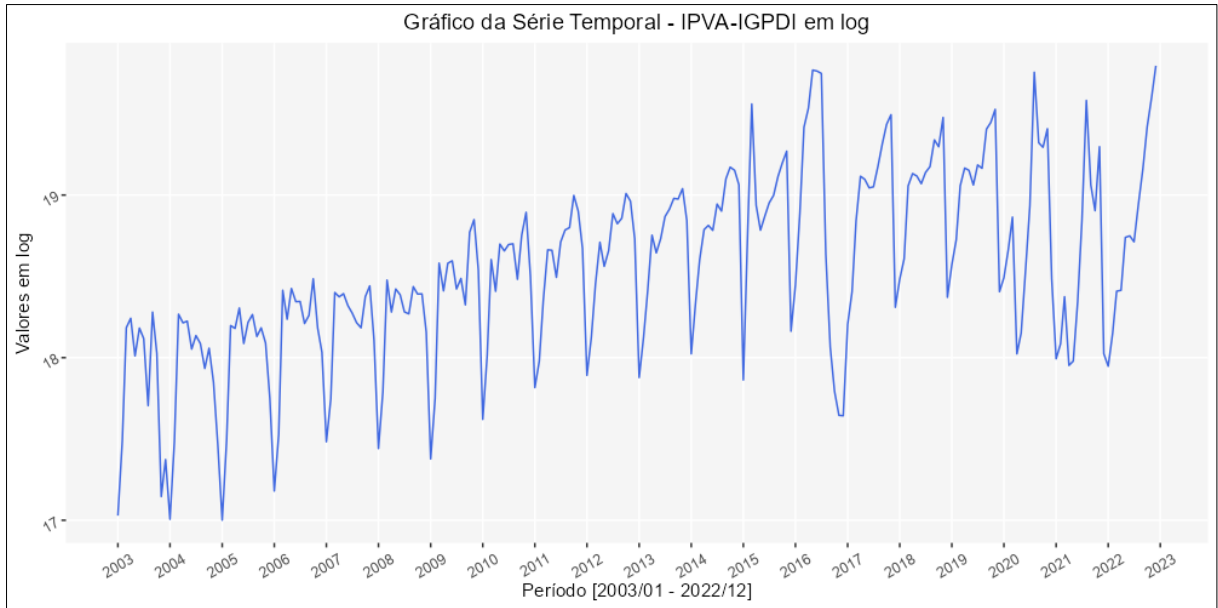


Figura 10 - Demonstração da Série Temporal IPVA em escala Logarítmica

(fonte: elaboração própria)

#### 4.4 – Estacionariedade

Executando os 3 testes para investigar a estacionariedade da série temporal das observações do IPVA, atualizada por IGPDI, com e sem escala logarítmica. Os testes foram realizados no software R apoiados pela biblioteca chamada “*tseries*”. O teste de ADF foca em detectar se uma série temporal possui raiz unitária através da função “*adf.test()*”<sup>36</sup> que mede a significância estatística de uma raiz unitária. Assim, teremos as hipóteses:

- Hipótese Nula (H0): A série temporal possui uma raiz unitária, ou seja, não é estacionária (e em outras palavras, o valor estatístico é maior que o valor crítico).
- Hipótese Alternativa (H1): A série temporal não possui uma raiz unitária, ou seja, é estacionária (e em outras palavras, o valor estatístico é pequeno).

Os resultados do teste de Dickey-Fuller indicam que as séries temporais, deflacionada e logarítmica, são estacionárias. Seus valores estatísticos são:  $-5,834$  e  $-6,775$ , que são significativamente menores que os valores críticos  $0,00394$  (para ambos), indicando que a série é estacionária, enquanto isso, o p-valor do teste é comum e calculado em  $0,01$ , que é inferior ao nível de significância de  $0,05$  reforça

<sup>36</sup> *Ibid.*

essa interpretação. Significando que há uma probabilidade de 1% de se obter uma estatística do teste pelo menos tão extrema quanto à observada, assumindo que a hipótese nula é verdadeira. Rejeitamos, dessa forma, a hipótese nula e concluímos que a série temporal é estacionária.

Já o teste de KPSS que avalia a variabilidade dos dados para verificar a estacionariedade através da função “*kpss.test()*”<sup>37</sup> que mede a significância de uma tendência não estacionária na série temporal. Dito isso, teremos como hipóteses:

- Hipótese Nula (H0): A série temporal é estacionária, ou seja, não raiz unitária (e, nesse caso, valor estatístico alto).
- Hipótese Alternativa (H1): A série temporal é não estacionária, ou seja, possui raiz unitária (e, nesse caso, valor estatístico baixo).

Os resultados do teste KPSS apontam para uma série não estacionária, pois, com o valor estatístico de 3,638, que é maior que o valor crítico 0,739, e mede o quanto a série temporal é estacionária em torno da média, e um p-valor de 0,01, que é menor que o nível de significância padronizada, portanto confirmamos a hipótese nula que a série temporal é estacionária e não possua raiz unitária.

E, o teste de PP que é um teste de regressão em primeira diferença para avaliar se uma série temporal é estacionária, utiliza a função “*pp.test()*”<sup>38</sup> que soma dos resíduos da regressão sem tendência ajustadas por um fator que leva em consideração a autocorrelação desses resíduos, dessa foram, temos as hipóteses:

- Hipótese Nula (H0): A série temporal não é estacionária, ou seja, possui tendência ou raiz unitária (e valor estatístico maior que valor crítico).
- Hipótese Alternativa (H1): A série temporal é estacionária, ou seja, não possui tendência nem raiz unitária (e valor estatístico menor que valor crítico).

O resultado do teste PP indicou que ela é estacionária, pois, com o valor estatístico muito pequeno  $-93,71$  em relação ao valor crítico de  $-3,46$ , e um p-valor

---

<sup>37</sup> *Ibid.*

<sup>38</sup> *Ibid.*

de 0,01, menor que o nível de significância padronizada, rejeitamos a hipótese nula que a série temporal é não estacionária e possui raiz unitária.

Abaixo apresentamos uma Tabela 4 comparativa dos resultados dos testes de estacionariedade:

*Tabela 4 - Resultados da investigação da Estacionariedade*

Medida	ADF		KPSS		PP	
	deflacionada	log	deflacionada	log	deflacionada	log
Valor estatístico	-5,834	-6,775	3,638	4,316	-79,10	-94,27
Valor crítico	0,00394		0,739		-2,75	-3,44
p-valor	0,01		0,01		0,01	
Avaliação	estacionária		não estacionária		estacionária	

*(fonte: elaboração própria)*

Como podemos observar dos resultados anteriores, nem todos os testes classificaram como série estacionária. Com essa dúvida, é necessário diferenciar a série temporal e avaliar novamente. Após uma diferenciação nos valores da série deflacionada, obtivemos a Tabela 5 que demonstra que temos, agora, uma série estacionária.

*Tabela 5 - Resultados da investigação na série temporal diferenciada*

Medida	ADF	KPSS	PP
Valor estatístico	-8,01	0,114	-233,28
Valor crítico	0,00394	0,739	-16,46
p-valor	0,01	0,01	0,01
Avaliação	estacionária	estacionária	estacionária

*(fonte: elaboração própria)*

Seguiremos os estudos, a partir daqui, apenas com a série deflacionada diferenciada, uma vez que não apresentou diferença significativa nos testes entre a série deflacionada e a logarítmica.

#### 4.5 – Funções de Autocorrelação e Autocorrelação Parcial

As funções de autocorrelação (ACF) e de autocorrelação parcial (PACF) contribuem para análise das séries temporais. A ACF que mensura a razão denominada coeficiente de correlação, que varia de -1 a 1, entre os valores de uma série temporal com seus valores passados, por outro lado, enquanto a PACF que também calcula a correlação entre os próprios valores da série temporal, o faz após controlar os efeitos sobre os valores do passado. A ACF demonstra o comportamento de longo alcance em uma série temporal, noutro lado, a PACF possibilita analisar o padrão de autocorrelação no intervalo de tempo menor.

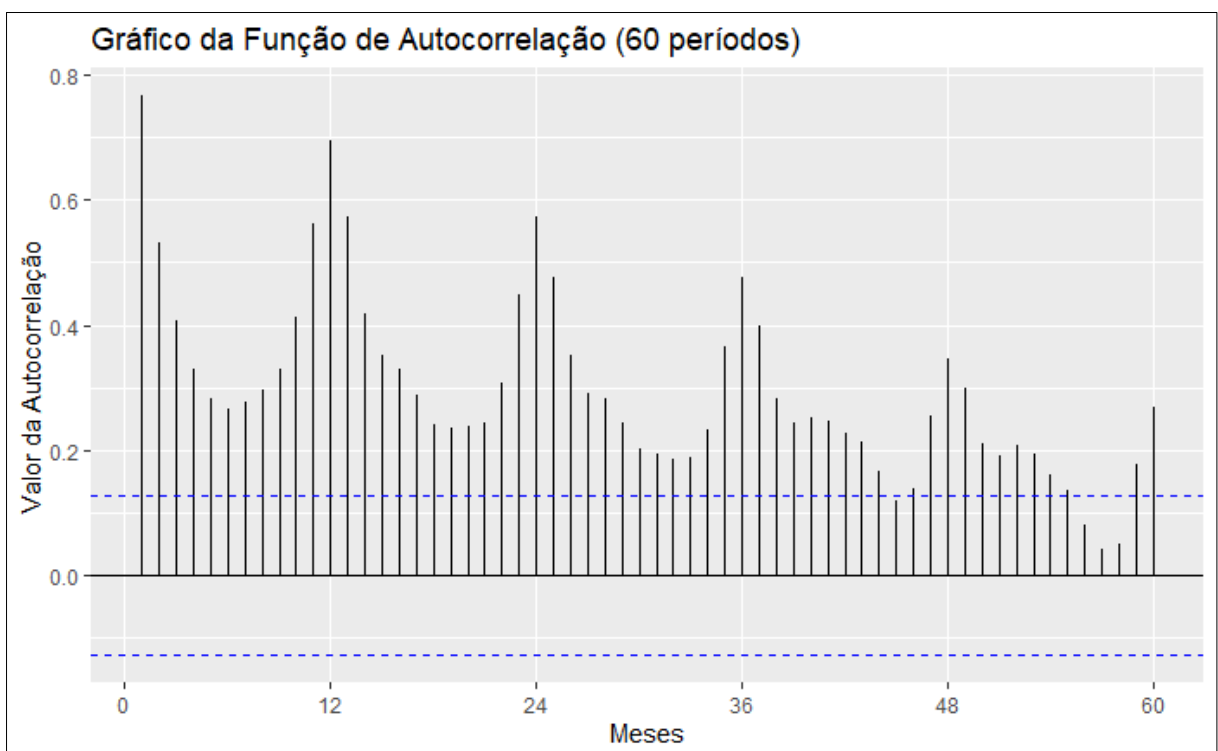


Figura 11 - ACF da Série Temporal IPVA (60 períodos)

(fonte: elaboração própria)

O gráfico da ACF, Figura 11, acima demonstra a correlação significativa entre o valor do imposto após 12 meses, além disso, percebemos que esses valores seguem com similaridade significativa até quase 5 anos. Outro aspecto interessante é que a relação se mostrou sempre positiva, com maior ou menor indução. Os significativos valores do gráfico ACF nos primeiros atrasos indicam a ordem do modelo de média móvel<sup>39</sup>, no caso, igual a 1, além disso, a queda gradual indica os ápices da série temporal indicam a necessidade de diferenciação.

<sup>39</sup> GUJARATI; PORTER, **Basic econometrics**.



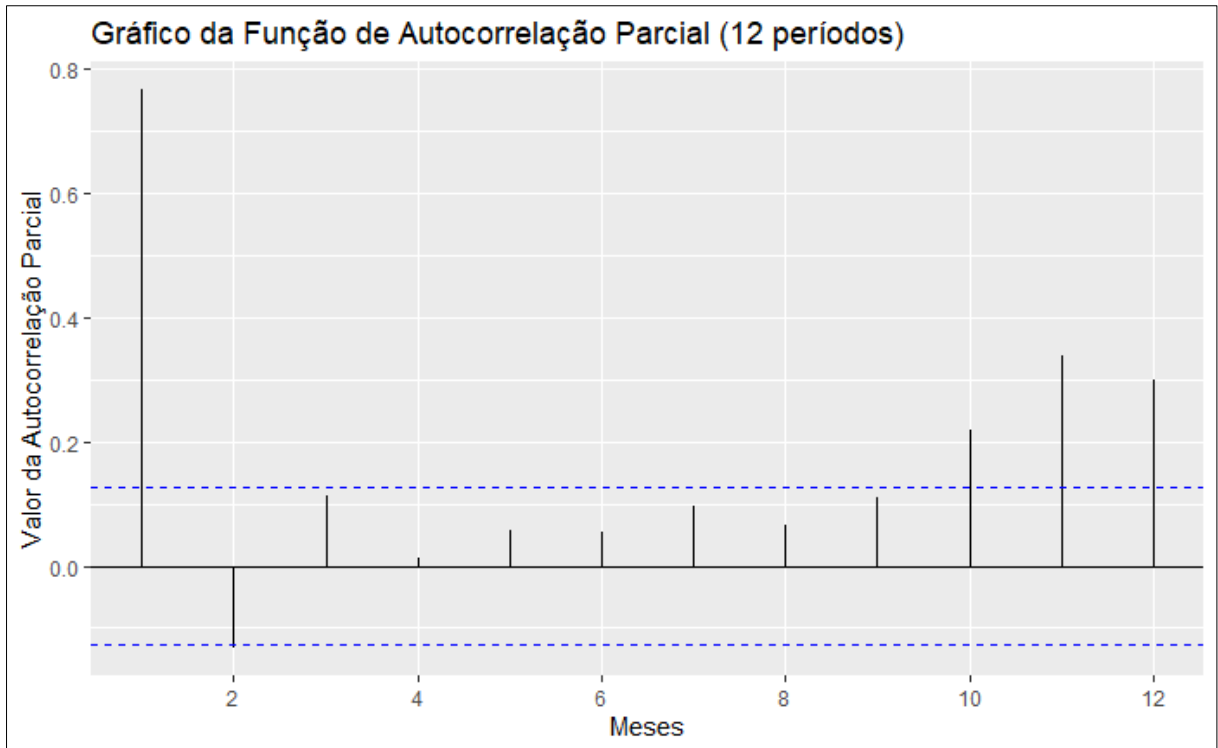


Figura 12 - PACF da Série Temporal IPVA (12 períodos)

(fonte: elaboração própria)

Ao analisar o gráfico da PACF, Figura 12, podemos perceber que apenas o valor do elemento antecessor ainda sofre algum efeito valor atual após controlar a correlação. O valor significativo do gráfico PACF nos primeiros atrasos sugerem a ordem do modelo autorregressivo<sup>40</sup> igual a 1. Além disso, os gráficos de autocorrelação parcial e total possuem comportamentos diferentes, pois em PACF houve a subtração da tendência do elemento sobre os elementos subsequentes.

#### 4.6 – Divisão da base de dados

Ainda sobre a base de dados e o escopo desta pesquisa, é importante esclarecer sobre o processo de avaliação e validação dos modelos preditivos avaliados. Assim, a estratégia adotada foi dividir os dados em duas partes distintas e sequenciais. A primeira parte, em média, 95% dos registros, será usada para aprimoramento do modelo, enquanto a segunda parte, os 5% seguintes da série temporal, em média, serão usados para confrontar valores preditos pela técnica em

<sup>40</sup> *Ibid.*

questão e, por conseguinte, contribuir para determinação da acurácia do procedimento.

Essa técnica de divisão de dados para treinamento e validação será repetida para 4 intervalos contíguos da série temporal, assim, conforme a Tabela 6, teremos a seguinte organização:

*Tabela 6 - Divisão da base de dados para treinamento e avaliação*

	<b>Treinamento</b>	<b>Avaliação</b>
<b>1º Partição</b>	Jan/2003 até Dez/2018 192 períodos	Jan-Dez/2019 12 períodos
<b>2ª Partição</b>	Jan/2003 até Dez/2019 204 períodos	Jan-Dez/2020 12 períodos
<b>3ª Partição</b>	Jan/2003 até Dez/2020 216 períodos	Jan-Dez/2021 12 períodos
<b>4ª Partição</b>	Jan/2003 até Dez/2021 228 períodos	Jan-Dez/2022 12 períodos

*(fonte: elaboração própria)*

## 5 – Análise dos resultados

Neste capítulo serão demonstrados os resultados da aplicação de vários modelos para previsão do IPVA, lembrando que a série temporal foi obtida em registros administrativos. As avaliações serão comparativas entre os seguintes modelos: HW (Aditivo e Multiplicativo), ARIMA e SARIMA. O processo investigativo seguiu a metodologia de parametrização do modelo através do intervalo da série temporal definido pela coluna treinamento da Tabela 6 e, em seguida será realizado as 12 previsões de períodos futuros que serão confrontados com os 12 valores contidos na coluna avaliação. Ainda cabe lembrar que os valores da série temporal, tanto para teste quanto para validação, foram deflacionados pelo IGPD e, em seguida, transformados pela função logarítmica. Apesar disso, os resultados apresentados nesse trabalho estarão apresentados em valores correntes em R\$.

Como discutido na seção 5.3, já contamos com uma série estacionária, característica essa que possibilitou a escolha dos modelos escolhidos para o confronto. Além disso, temos disponíveis as seguintes medidas: MAE, MSE, RMSE, MAD, MAPE, AIC e BIC como balizadores da determinação da melhor estratégia para previsão, entretanto a comparação dos modelos usará primordialmente MAPE.

### 5.1 – Análise da 1ª Partição: 2019

Em relação à primeira partição, podemos perceber pelo gráfico da esquerda na Figura 13 que a série se apresentou bastante “comportada” até o início de 2015, isto é, respeitando uma faixa com tendência de alta. Entretanto, nos anos seguintes ela passou por 3 anos bem “agitados”, isto é, extrapolando essa faixa imaginária com picos e depressões bem significativos. E, no final do período, voltando a ter um comportamento mais “tranquilo” no final da partição. Ainda na Figura 13, do lado direito, demonstramos a série em escala log, e o interessante é que fica ainda mais nítido um perfil limitado a certa faixa de valores na maior parte da série. Apesar de utilizarmos para processamento a série em log, apresentamos graficamente na Figura 14 o recorte dos dados tanto para treino quanto para validação usados nos modelos.

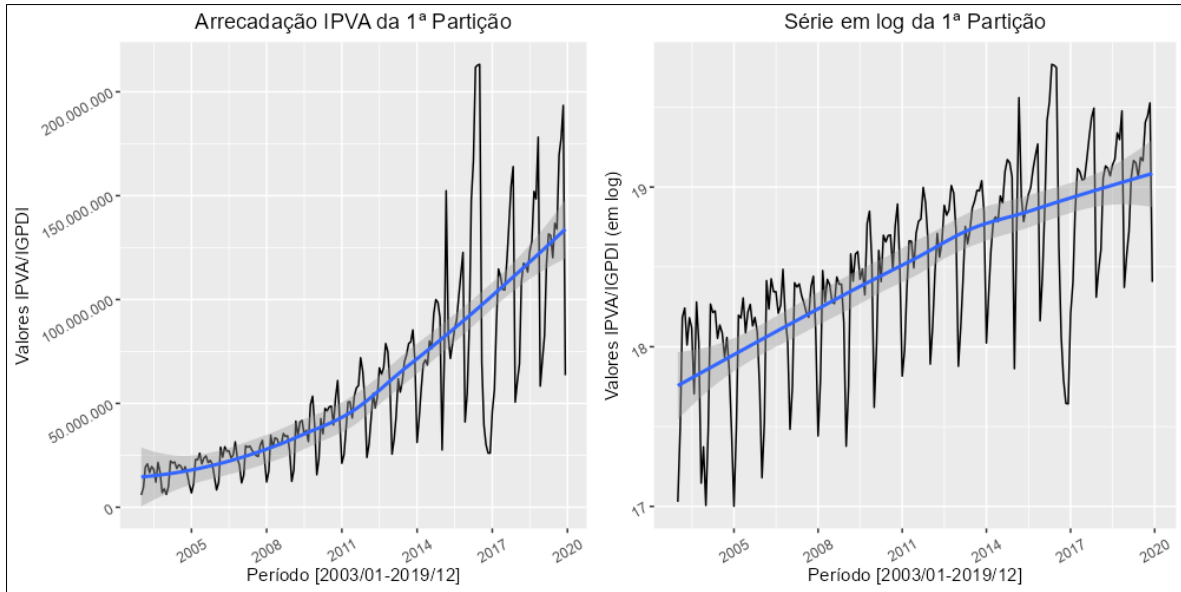


Figura 13 - Gráfico da Arrecadação IPVA x Série em log com Tendências, para a 1ª Partição

(fonte: elaboração própria)

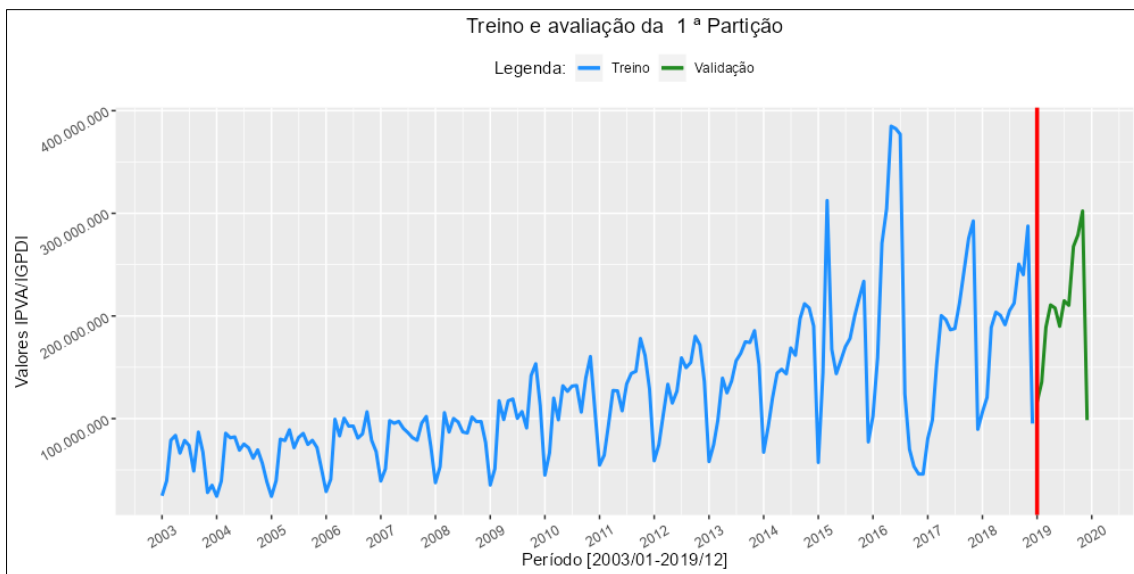


Figura 14 - Divisão dos dados em treino e validação para 1ª Partição

(fonte: elaboração própria)

Abaixo, na Tabela 7, são apresentados os resultados da aplicação dos modelos HW-A, HW-M, SARIMA e SARIMA para aplicação na 1ª partição dos dados, janeiro de 2003 até dezembro de 2018 e previsão de 12 períodos de 2019. E em destaque, nas duas últimas linhas da tabela estão destacados a medida avaliativa MAPE, tanto para valores correntes e em logarítmico. E, os menores valores obtidos foram: 14,16% e 1,80%, respectivamente para série em valor corrente e logarítmica, demonstrando que SARIMA foi o melhor modelo para previsão. Entretanto, comparando as diferenças, mês a mês, observamos que se saiu melhor em 10 meses,

além do total arrecadado de 2019, sendo menos preciso apenas para o ARIMA no mês de fevereiro e para o HW-A no mês de junho.

**Tabela 7 - Processamento dos modelos na 1ª partição, previsões: jan-dez de 2019**

Resumo: Previsões de Receita IPVA - IGPMI - 2019					
Mês/Ano	Arrecadação R\$	HW-A	HW-M	ARIMA	SARIMA
jan/19	116.181.501,70	68.986.663,62	66.493.797,83	88.356.475,89	97.945.247,64
fev/19	135.557.135,88	99.357.165,13	98.267.086,07	127.003.012,51	126.678.263,50
mar/19	188.957.358,78	177.520.049,08	175.541.281,03	144.429.265,85	194.291.757,43
abr/19	210.727.775,32	181.669.973,96	179.799.377,35	151.039.365,37	211.819.221,89
mai/19	207.772.351,15	186.843.399,34	181.696.419,51	153.284.564,90	209.835.335,45
jun/19	189.890.269,60	181.637.267,03	172.416.086,58	153.902.989,19	202.882.960,66
jul/19	214.811.077,23	191.826.680,95	181.489.718,92	153.932.531,56	212.060.525,99
ago/19	210.325.071,36	172.687.667,53	164.307.460,90	153.751.763,92	192.386.301,11
set/19	267.720.179,58	173.118.375,34	164.706.683,25	153.496.910,02	202.724.184,75
out/19	278.913.779,93	185.374.536,32	177.672.148,19	153.216.747,35	199.630.557,12
nov/19	302.313.990,89	171.048.955,74	169.763.086,32	152.928.706,16	217.408.147,90
dez/19	98.501.481,59	115.788.776,77	115.369.073,17	152.639.005,71	87.814.426,48
<b>Total</b>	<b>2.421.671.973,01</b>	<b>1.905.859.510,82</b>	<b>1.847.522.219,12</b>	<b>1.737.981.338,42</b>	<b>2.155.476.929,91</b>
MAPE		31,13182299	35,0497359	44,45444692	14,15810097
MAPE (p/ log)		2,20803649	2,251911311	2,088377878	1,804176601

(fonte: elaboração própria)

Ainda sobre a Tabela 7 é interessante observar que, em relação ao total arrecadado, o SARIMA não seria o mais eficiente, ficando sob esse critério o mérito para o HW-M. E, investigando outras medidas comparativas, percebidas na Tabela 8, temos que os melhores resultados R\$ 25.763.173,48 para MAE, R\$ 39.556.918,17 para RMSE, 30,607 para AIC e 52,957 para BIC pertencem a coluna do método SARIMA e confirmam como método mais preciso. Ainda graficamente, a Figura 15 demonstra as regiões próximas e distantes dos valores previstos por cada modelo em relação aos valores arrecadados e separados para avaliação. É interessante observar que, mesmo mais preciso, o SARIMA entre agosto-novembro se apresentou mais afastado do valor de referência, entretanto, na maior parte do ano foi o modelo que melhor representou a arrecadação.

**Tabela 8 - Outras avaliações dos métodos para 1ª partição, jan-dez de 2019**

Medida	HW-A	HW-M	ARIMA	SARIMA
MAE	45.865.587,71	50.657.078,09	65.997.140,24	25.763.173,69
RMSE	59.296.149,18	63.196.198,28	77.275.015,47	39.556.918,29

AIC	507,0206195	506,5158591	131,2188984	30,60714568
BIC	562,3980409	561,8932804	150,7638706	52,95784363

(fonte: elaboração própria)

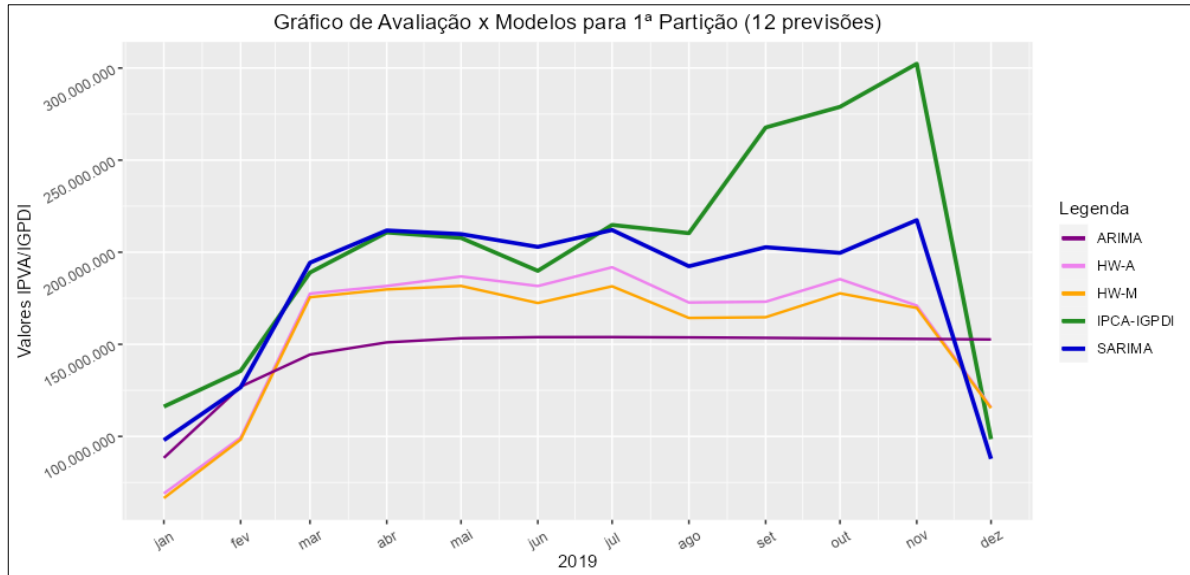


Figura 15 - Comparação dos modelos x dados de avaliação da 1ª Partição

(fonte: elaboração própria)

Investigando a etapa de treinamento entre os modelos, temos na Tabela 9 os resultados do MAPE do ajuste e dos resíduos. Em relação ao ajuste, o melhor resultado, 0,87% é do modelo SARIMA e, em relação aos resíduos, o modelo MW-A resultou no melhor resultado, a saber, 43.370,95% e, logo em seguida, o modelo SARIMA com 53.704,47%. Apesar de não possuir o melhor resultado para essas duas análises, persistimos como SARIMA como melhor resultado uma vez que está bastante próximo do melhor resultado na situação que não foi o melhor modelo. Ainda nessa análise, apresentamos a Figura 16 que demonstra a etapa de treinamento do modelo SARIMA.

Tabela 9 - Avaliação dos ajustes e resíduos para 1º partição, jan-dez de 2019

Medida	HW-A	HW-M	ARIMA	SARIMA
MAPE (ajuste)	0,927542615	0,918397709	1,341439942	0,870927915
MAPE (resíduo)	43.370,95	2.735.693,43	164.339,16	53.704,49

(fonte: elaboração própria)

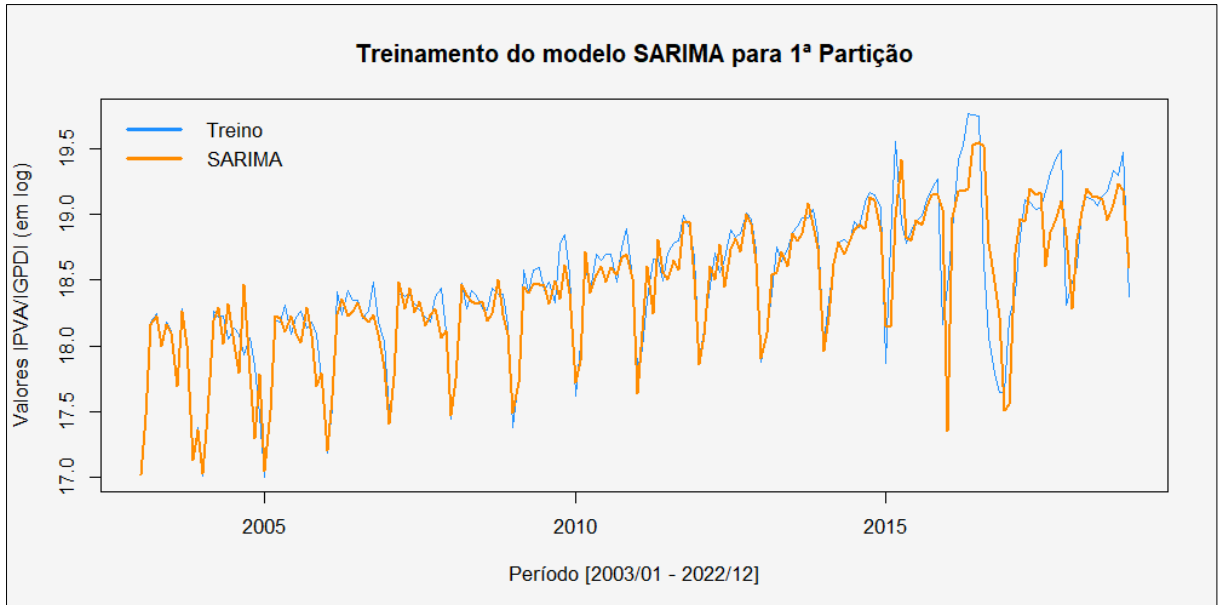


Figura 16 - Treinamento do SARIMA para 1ª Partição

(fonte: elaboração própria)

A parametrização utilizada para o modelo SARIMA( $p,d,q$ )( $P, D, Q, s$ ) foi:

- Dois termos autorregressivos (AR:  $p=2$ )
- Nenhuma diferença ( $d = 0$ )
- Dois termos de média móvel (MA:  $q=2$ )
- Um termo autorregressivo sazonal ( $P = 1$ )
- Uma diferença sazonal ( $D = 1$ )
- Um termo de média móvel sazonal ( $Q = 1$ )
- Um período sazonal de 12

Coefficients:

	ar1	ar2	ma1	ma2	sar1	sma1
	0.1559	0.3773	0.5772	-0.0526	0.0116	-0.5219
s.e.	0.2908	0.1886	0.2875	0.0965	0.1579	0.1517

sigma<sup>2</sup> = 0.06503: log likelihood = -8.55  
 AIC=31.1 AICc=31.75 BIC=53.45

Error measures:

	ME	RMSE	MAE	MPE	MAPE
MASE					
ACF1					
Training set	0.04771914	0.2427695	0.1614718	0.2511137	0.8839747
	0.7457539	-0.04061465			

Figura 17 - Resultado do comando summary() sobre o modelo SARIMA - 1ª Partição

(fonte: elaboração própria)

Além disso, a interpretação a modelagem SARIMA(2,0,2)(1,1,1)[12], conforme a Figura 17, demonstrou os coeficientes dos termos autorregressivos de primeira e segunda ordem de 0,1559 e 0,3773. Isso significa que observações adjacentes na série temporal são correlacionadas positivamente, com uma correlação de 0,1559 entre observações adjacentes e uma correlação de 0,3773 entre observações adjacentes com um atraso de 1 período. Além disso, 0,1559 é mais significativo que 0,3773 indicando que os valores mais recentes propostos da série temporal são os mais relevantes na previsão de valores futuros. O valor estimado de  $\sigma^2$  é 0,06503, sugerindo a variância residual do modelo é relativamente baixa e o valor do logaritmo de verossimilhança é  $-8,55$  comprovando bom ajuste aos dados. E, finalizando a análise dos resultados, representamos, na Figura 18, os valores previstos com intervalo de confiança de 95%, na cor azul e em vermelho, os valores arrecadados reais em escala logarítmica para série toda.

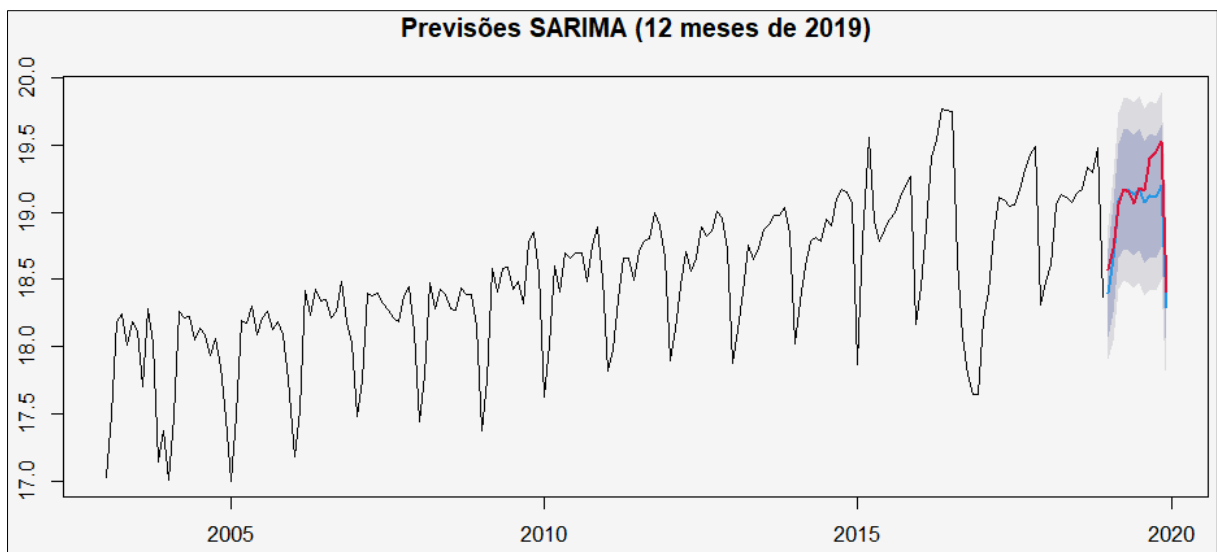


Figura 18 - Representação das previsões futuras para 1ª Partição

(fonte: elaboração própria)

## 5.2 – Análise da 2ª Partição: 2020

O acréscimo de novo ano na série tempo da arrecadação do IPVA na 2ª Partição acrescentou uma complexidade para os modelos. Graficamente, pela Figura 19 podemos perceber que o novo ano traz uma variação significativa dos valores, tanto em relação ao top como em relação ao fundo. Também trazemos para esse novo estudo o gráfico da divisão dos dados em treino e validação apresentado na Figura 20.



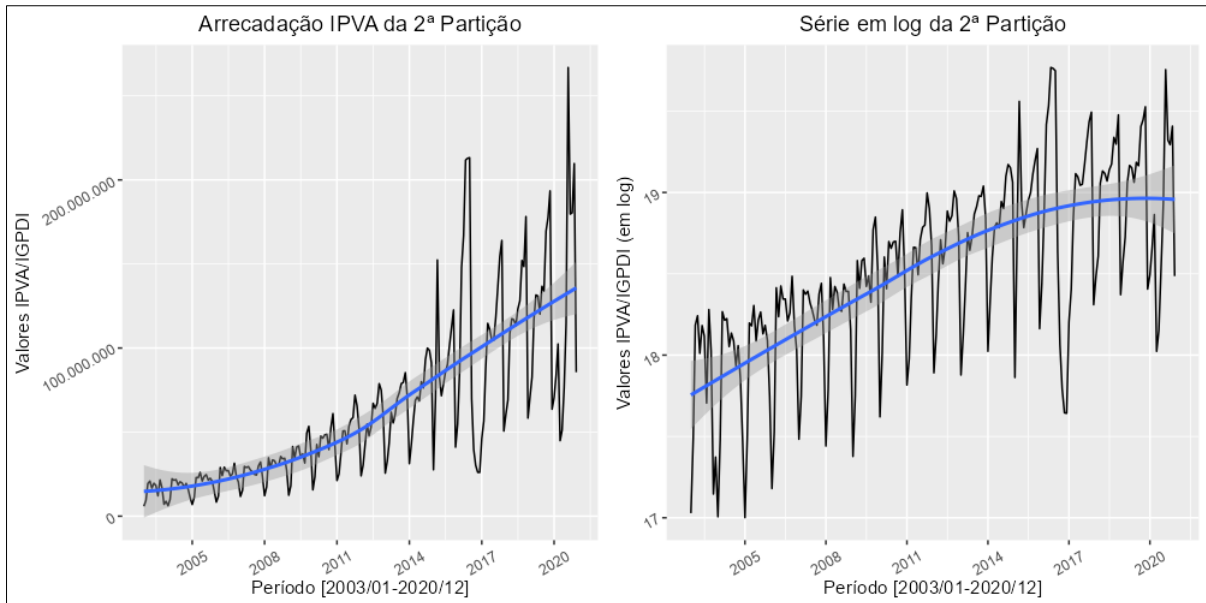


Figura 19 - Gráfico da Arrecadação IPVA x Série em log com Tendências, para a 2ª Partição

(fonte: elaboração própria)

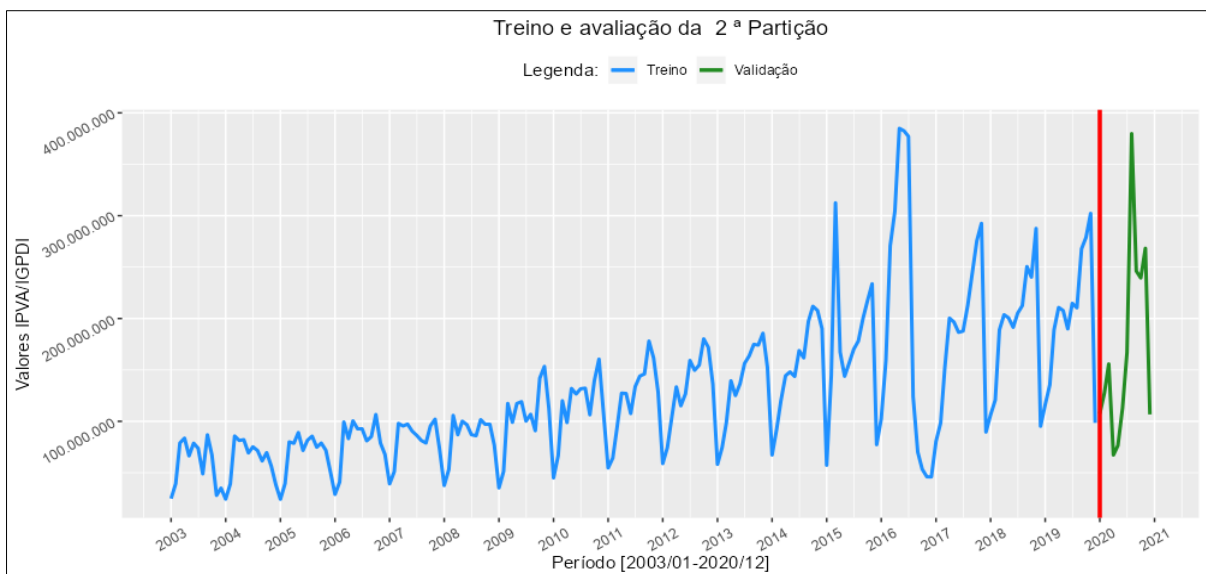


Figura 20 - Divisão dos dados em treino e validação para 2ª Partição

(fonte: elaboração própria)

Abaixo, na Tabela 10, também são apresentados os resultados processados dos modelos para a 2ª partição: as previsões e a medida MAPE. A investigação da tabela em busca dos menores valores para medida comparativa obteve para o MAPE os valores de 27,96% e 2,30% para valores correntes e em escala logarítmica respectivamente, demonstrando que o método SARIMA é novamente o mais preciso. Já a Tabela 11 destaca outras medidas complementares para análise comparativa dos modelos. Os menores valores percebidos em cada linha, R\$ 56.087.867,31 para

o MAE, R\$ 82.914.667,45 para o RMSE, 21,19 para o AIC e 43,99 para o BIC, reforçam os melhores resultados para o SARIMA. Esses resultados, diferente dos percebidos na 1ª Partição demonstraram superioridade em todos os critérios em favor do SARIMA. Entretanto, comparando as diferenças, mês a mês, observamos que se saiu melhor que os demais em 6 diferentes meses de 2020, e menos preciso que o ARIMA entre março-julho, que o HW-M em dezembro e que o HW-A em relação ao total arrecadado. Além disso, a hegemonia do SARIMA em relação a 2020 foi menor que em 2019.

**Tabela 10 - Processamento dos modelos na 2ª partição, previsões: jan-dez de 2020**

Resumo: Previsões de Receita IPVA - IGPD - 2020					
Mês/Ano	Arrecadação Real	HW-A	HW-M	ARIMA	SARIMA
jan/20	107.372.559,10	74.307.140,08	73.494.146,50	96.688.220,64	111.760.585,00
fev/20	127.712.587,57	103.273.540,68	106.628.584,17	136.646.361,79	134.144.064,84
mar/20	155.844.042,49	182.200.068,84	191.736.174,36	155.116.531,00	195.083.843,65
abr/20	67.189.702,83	185.577.565,29	193.318.190,05	162.384.829,18	214.431.972,99
mai/20	76.421.015,30	192.939.918,95	196.596.716,74	164.981.057,97	211.318.863,81
jun/20	112.847.558,59	189.068.220,30	185.810.806,56	165.775.286,66	197.789.577,63
jul/20	167.099.096,42	200.279.167,73	196.253.406,00	165.894.448,92	214.585.775,35
ago/20	379.875.883,87	183.155.727,15	176.214.687,64	165.763.898,06	203.237.780,41
set/20	246.166.630,32	189.951.389,93	184.828.104,31	165.541.893,57	236.315.414,85
out/20	239.379.027,33	208.952.287,97	202.067.856,83	165.287.006,19	239.015.429,08
nov/20	268.352.523,06	195.364.381,62	186.485.643,12	165.020.867,16	260.645.967,08
dez/20	106.736.079,51	130.609.664,91	118.680.959,34	164.751.456,15	92.869.193,80
<b>Total</b>	<b>2.054.996.706,40</b>	<b>2.035.679.073,46</b>	<b>2.012.115.275,62</b>	<b>1.873.851.857,28</b>	<b>2.311.198.468,49</b>
MAPE		39,24125701	40,52301316	40,29505553	27,96114182
MAPE (p/ log)		2,658419974	2,76247823	2,702191274	2,297444064

(fonte: elaboração própria)

**Tabela 11 - Outras avaliações dos métodos para 2ª partição, jan-dez de 2020**

Medida	HW-A	HW-M	ARIMA	SARIMA
MAE	67.365.987,89	69.616.578,88	65.700.745,42	56.087.873,32
RMSE	84.444.287,85	87.978.225,64	87.454.473,29	82.914.668,67
AIC	551,6119449	551,1371888	139,0373476	21,19174772
BIC	608,0199848	607,5452287	158,9460675	43,99421532

(fonte: elaboração própria)

Observando a figura Figura 21, podemos perceber que, diferente da 1ª Partição que foi 3 meses, as previsões do SARIMA se afastaram dos valores coletados para os 5 meses: abril-agosto. Nessa partição houve mais meses que o método esteve pior, entretanto, para os demais o modelo praticamente coincidiu com valores observados.

Em relação aos resultados da fase de treinamento, a Tabela 12 demonstra que para o MAPE do ajuste, 0,84%, e MAPE dos resíduos 60.576,02% o SARIMA só obteve resultados inferiores ao do ARIMA e apenas para o MAPE dos resíduos que foi significativamente melhor 25.298,91%. Por fim, trazemos no gráfico da Figura 22 o registro do treinamento do SARIMA para o período de treinamento.

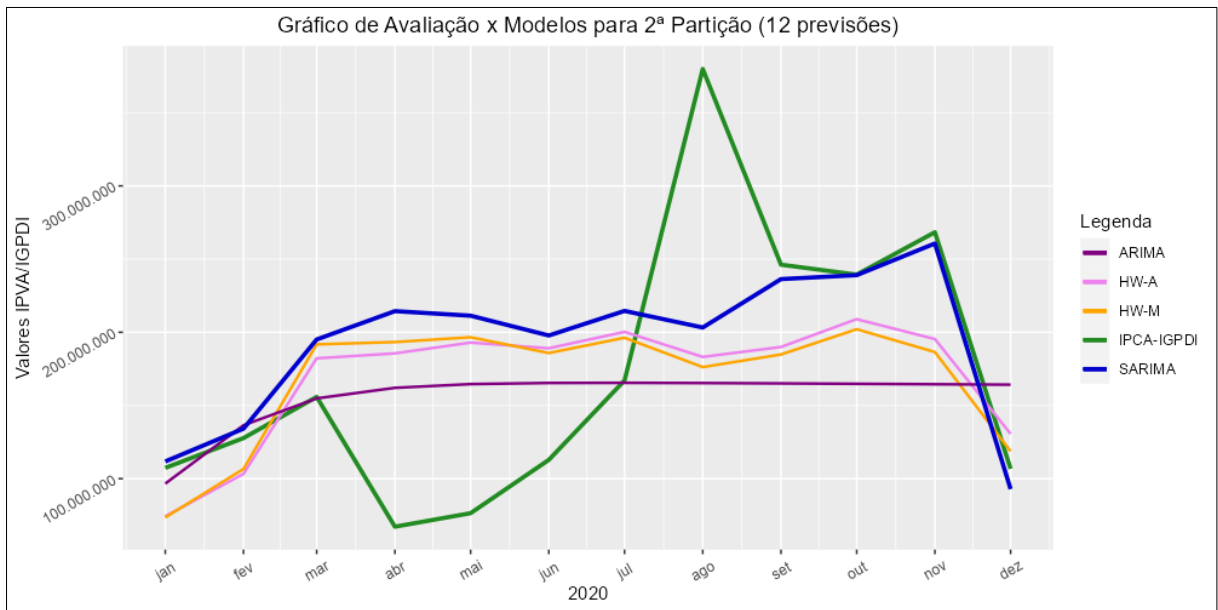


Figura 21 - Comparação dos modelos x dados de avaliação da 2ª Partição

(fonte: elaboração própria)

Tabela 12 - Avaliação dos ajustes e resíduos para 2º partição, jan-dez de 2020

Medida	HW-A	HW-M	ARIMA	SARIMA
MAPE (ajuste)	0,916925124	0,941650706	1,333779186	0,843858751
MAPE (resíduo)	384.926,30	1.673.987,22	25.714,12	60.575,77

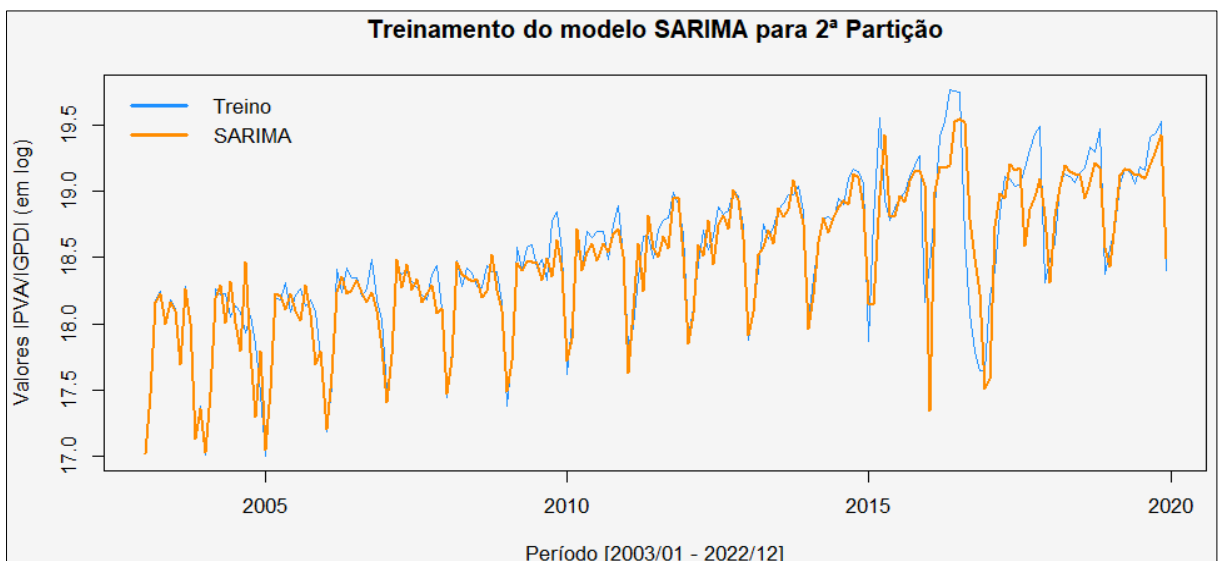


Figura 22 - Treinamento do SARIMA para 2ª Partição

(fonte: elaboração própria)

Não houve alteração nos parâmetros dos métodos e a Figura 23 resgata o resultado de console do software R do comando “summary()” do modelo SARIMA(2,0,2)(1,1,1)[12], e, comparativamente analisado em relação à Partição anterior, pouco mudou, exceto a função de verossimilhança que mostra maior aderência do modelo aos dados treinados, saindo de  $-8,55$  para  $-3,79$ . Além disso, trazemos a representação gráfica dos valores da previsão para 12 períodos à frente pela técnica SARIMA, com destaque para azul os valores propostos e em vermelho os valores observados, na Figura 24.

```

Coefficients:
      ar1      ar2      ma1      ma2      sar1      sma1
 0.1493  0.3737  0.5850 -0.0474 -0.0217 -0.4695
s.e.  0.2798  0.1802  0.2765  0.0932  0.1588  0.1529

sigma^2 = 0.06162:  log likelihood = -3.79
AIC=21.57  AICC=22.18  BIC=44.38

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 0.0454997 0.237033 0.1562923 0.2390307 0.8547895 0.7541287 -0.03806929

```

Figura 23 - Resultado do comando summary() sobre o modelo SARIMA - 2ª Partição

(fonte: elaboração própria)

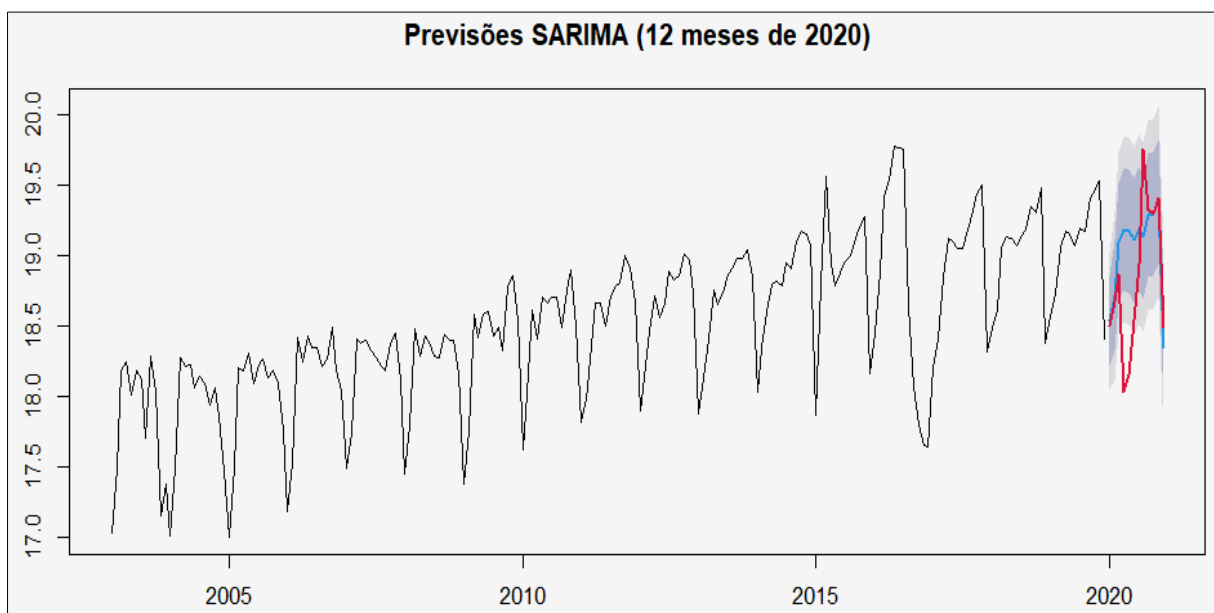


Figura 24 - Representação das previsões futuras para 2ª Partição

(fonte: elaboração própria)

### 5.3 – Análise da 3ª Partição: 2021

Abaixo, na Figura 25 temos o acréscimo de mais um ano a série temporal. A novidade que se apresenta é que ao acrescentar mais um período, a anterior “anormalidade” de escala da variação do ano 2020, em relação a série até aquele momento, se transforma em um novo patamar de valores observados na série temporal. Percebemos um topo ainda mais alto e uma depressão um pouco mais rasa que a do ano anterior. Isso demonstra que o total arrecadado de IPVA aumentou ainda mais. Além disso, mais uma vez demonstramos, na Figura 26, a apresentação gráfica da separação dos dados em treinamento e avaliação para os modelos em estudo.

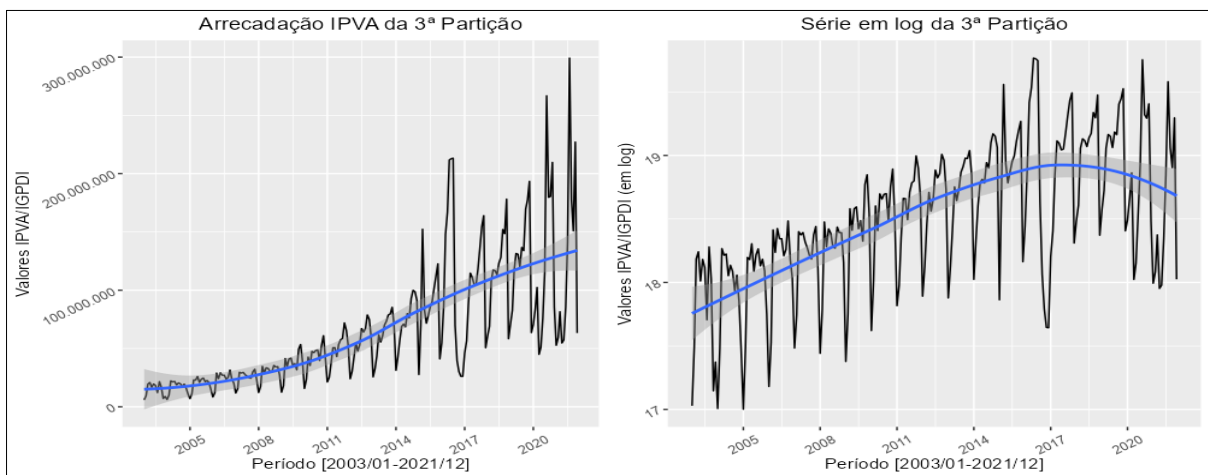


Figura 25 - Gráfico da Arrecadação IPVA x Série em log com Tendências, para a 3ª Partição  
(fonte: elaboração própria)

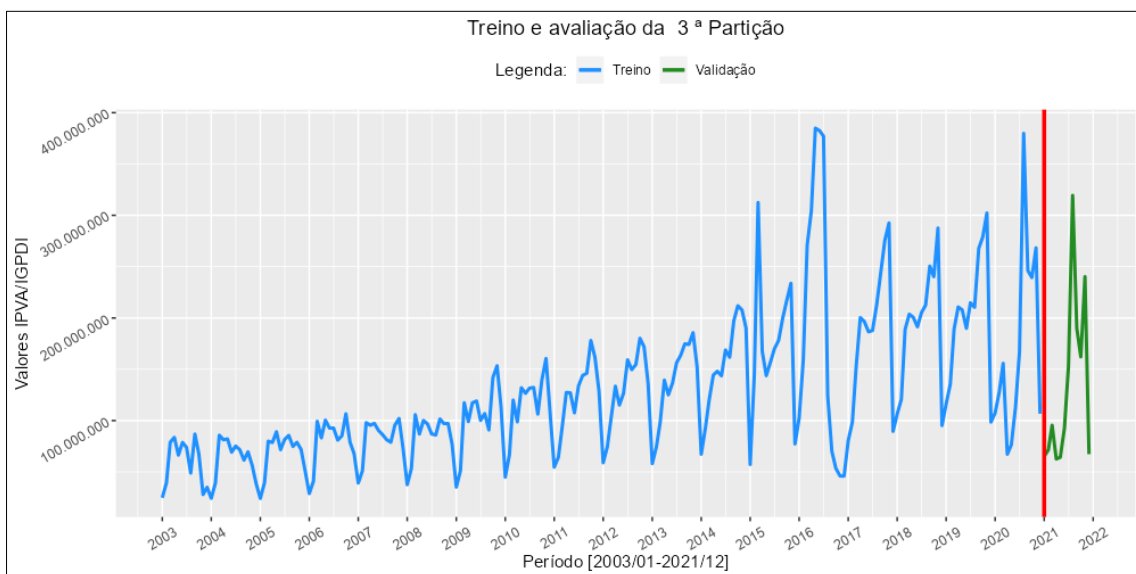


Figura 26 - Divisão dos dados em treino e validação para 3ª Partição  
(fonte: elaboração própria)

Ao final do processamento dos modelos, obtivemos na Tabela 13 os resultados para previsões e também da medida comparativa MAPE. Como comparativo do processamento, temos como melhor solução os valores 33,82% e 1,85% para o MAPE dos valores correntes e em escala logarítmica respectivamente para o modelo SARIMA. E os demais resultados de medidas comparativas: R\$ 53.971.880,04 do MAE, R\$ 56.609.281,86 do RMSE, 40,35 do AIC e 63,58 do BIC reforçam como modelo mais eficiente o SARIMA. Entretanto, comparando as diferenças, mês a mês, observamos que se saiu melhor que os demais em apenas 5 diferentes meses de 2021, e menos preciso que o ARIMA em 3 meses e em relação ao total arrecadado, que o HW-M em novembro e que o HW-A em 3 meses. Além disso, a hegemonia do SARIMA para o ano 2021 enfraqueceu, produzindo menos resultados precisos quando comparado com 2019 e 2020.

**Tabela 13 - Processamento dos modelos na 3ª partição, previsões: jan-dez de 2021**

Resumo: Previsões de Receita IPVA - IGPD - 2021					
Mês/Ano	Arrecadação Real	HW-A	HW-M	ARIMA	SARIMA
jan/21	65.199.162,74	76.946.858,81	114.681.707,84	104.516.475,82	119.933.146,71
fev/21	71.460.637,03	106.953.908,73	121.068.759,43	134.956.929,19	135.108.760,64
mar/21	95.539.210,08	186.027.556,11	171.950.445,30	148.382.874,63	180.081.831,55
abr/21	62.603.525,92	183.065.789,85	166.857.309,29	153.595.101,46	120.740.416,80
mai/21	64.271.821,01	190.851.127,94	200.862.491,31	155.433.447,30	127.916.591,24
jun/21	92.715.879,46	189.444.631,00	220.042.572,80	155.964.766,72	149.549.337,44
jul/21	152.872.645,16	203.416.428,59	244.620.557,82	156.001.089,55	190.031.119,65
ago/21	319.442.185,76	198.536.356,12	247.913.199,12	155.852.004,35	280.004.637,95
set/21	190.018.437,78	198.550.942,94	206.626.884,57	155.634.235,52	242.747.689,64
out/21	162.100.399,78	218.427.900,82	215.647.145,46	155.391.606,26	240.594.933,24
nov/21	240.312.633,61	206.088.104,79	235.828.671,24	155.140.540,58	266.077.932,55
dez/21	67.301.616,04	133.036.245,70	112.456.006,68	154.887.186,68	99.839.207,61
<b>Total</b>	<b>1.583.838.154,39</b>	<b>2.091.345.851,39</b>	<b>2.258.555.750,84</b>	<b>1.785.756.258,05</b>	<b>2.152.625.605,00</b>
MAPE		38,51080297	38,18248693	43,62948448	33,82165806
MAPE (p/ log)		2,248595169	2,161959483	2,033865293	1,847713713

(fonte: elaboração própria)

**Tabela 14 - Outras avaliações dos métodos para 3ª partição, jan-dez de 2021**

Medida	HW-A	HW-M	ARIMA	SARIMA
MAE	68.147.367,83	68.895.291,21	65.135.720,34	53.971.878,85
RMSE	79.260.485,86	79.174.022,01	77.397.646,01	56.609.280,72
AIC	624,606293	628,0192504	155,3642304	40,35202624
BIC	681,986026	685,3989833	175,6159008	63,5788662

(fonte: elaboração própria)

Ao final, a representação gráfica comparativa dos modelos para 12 previsões futuras do ano 2021, destacado pela Figura 27 mostra, assim como para a 2ª Partição, que o modelo SARIMA não foi bem nos primeiros meses do ano, até junho. Entretanto, para os 6 meses seguintes o modelo acabou apresentando resposta das previsões mais adequada. A Tabela 15 confirma esse entendimento, pois, ao avaliar o MAPE dos ruídos, o SARIMA pontuou 65.132,55% ficando atrás do ARIMA com 22.570,92 e do HW-A com 35.789,27%. Entretanto, ao avaliar o MAPE do ajuste, o método SARIMA 0,87% possui o melhor resultado. Trazemos também o gráfico que apresenta a atividade de treinamento do modelo SARIMA para esse período na Figura 27.

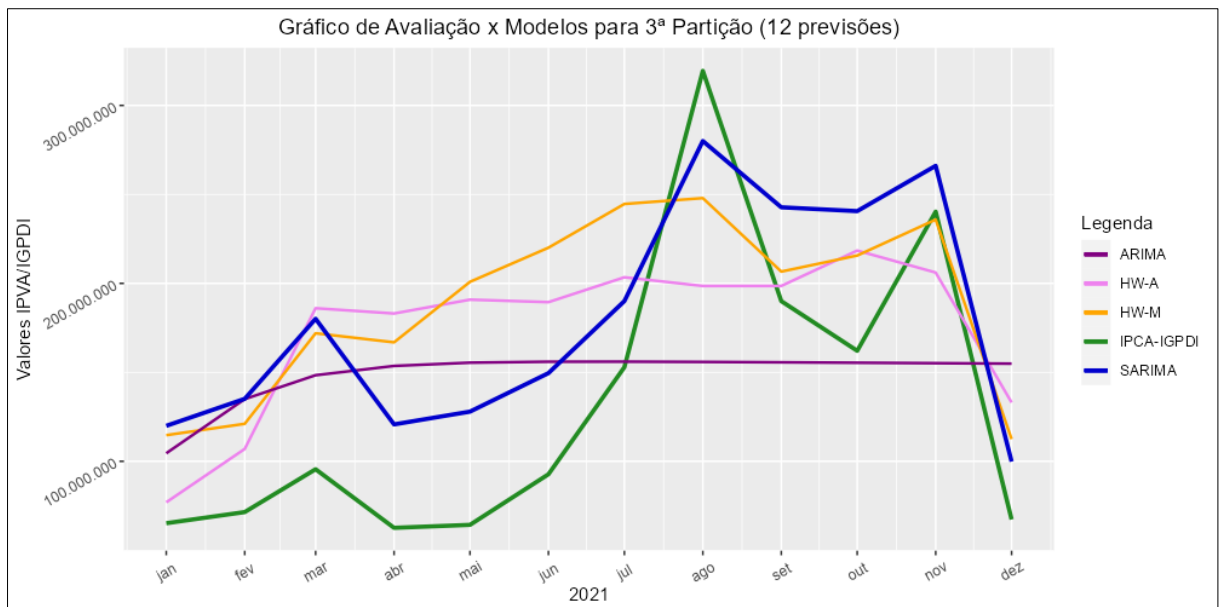


Figura 27 - Comparação dos modelos x dados de avaliação da 3ª Partição

(fonte: elaboração própria)

**Tabela 15 - Avaliação dos ajustes e resíduos dos modelos para 3ª partição**

Medida	HW-A	HW-M	ARIMA	SARIMA
MAPE (ajuste)	0,984509387	0,938330553	1,355231704	0,873884249
MAPE (resíduo)	35.789,27	930.796,70	22.570,92	65.132,36

(fonte: elaboração própria)

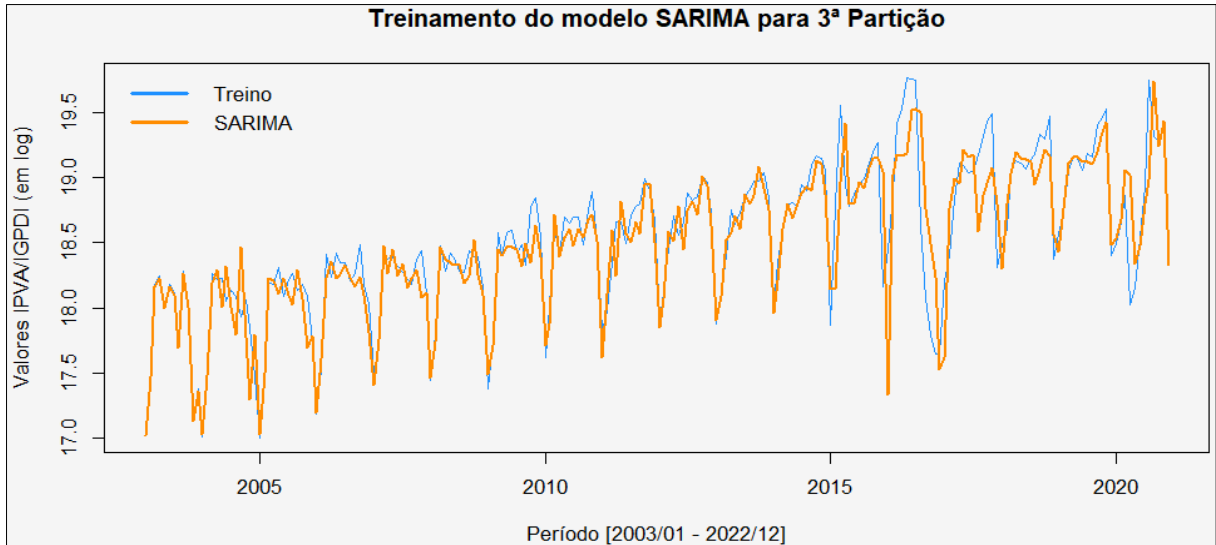


Figura 28 - Treinamento do SARIMA para 3ª Partição

(fonte: elaboração própria)

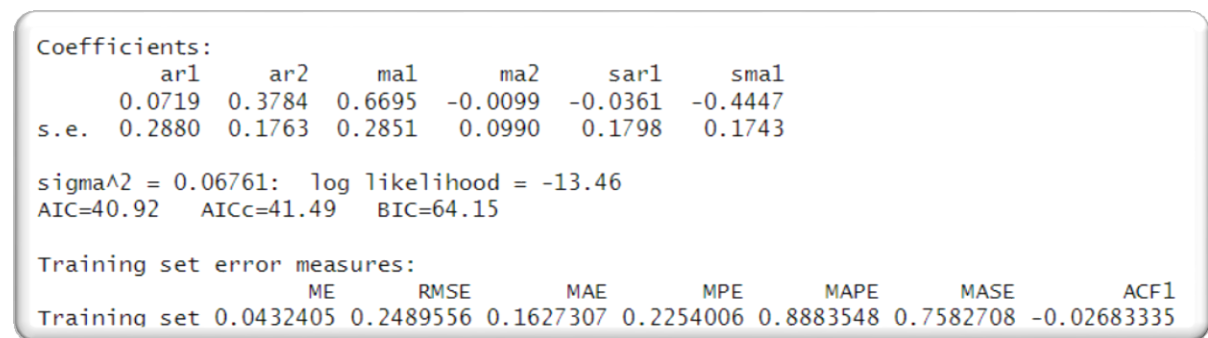


Figura 29 - Resultado do comando `summary()` sobre o modelo SARIMA - 3ª Partição

(fonte: elaboração própria)

A Figura 29, que mostra o resultado do comando “`summary()`”, demonstra que pouco mudou nos resultados de adequação do método SARIMA entre as partições 2 e 3. A Figura 30 que mostra os valores previstos pelo SARIMA, em azul e com confiabilidade de 95%, e os coletados para esse trabalho, em vermelho.



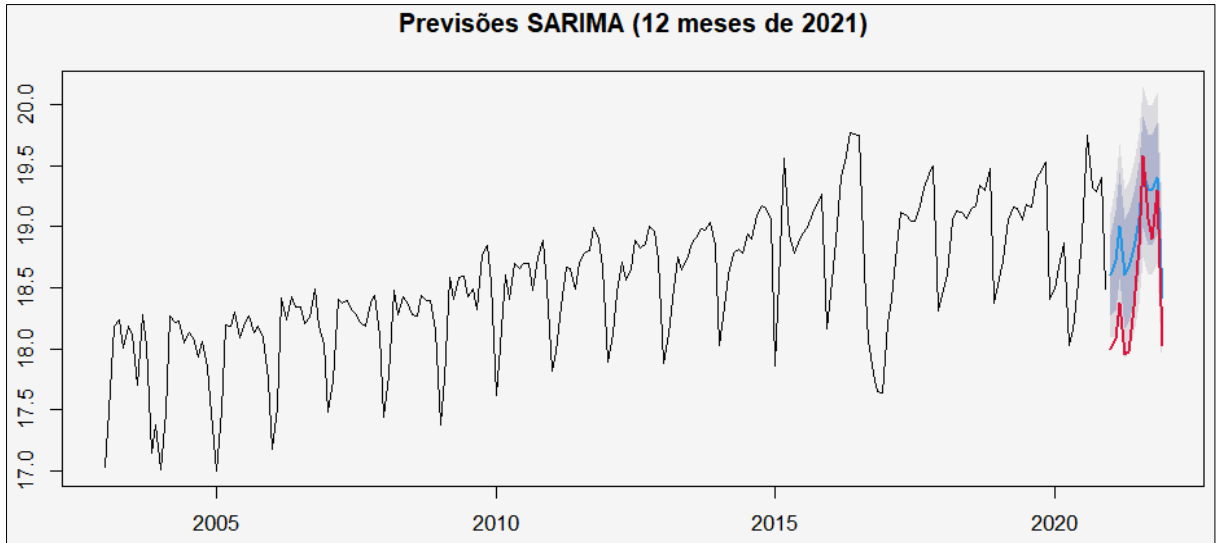


Figura 30 - Representação das previsões futuras para 3ª Partição

(fonte: elaboração própria)

#### 5.4 – Análise da 4ª Partição: 2022

Para essa última partição a ser estudada, o novo ano, mais uma vez, elevou o patamar dos valores da série temporal. A Figura 31 demonstra o novo topo mais alto e, no caso, um fundo bem mais raso. Observamos que o gráfico da direita dessa figura demonstra como a estratégia de suavizar a série com valores logarítmicos é uma boa estratégia. Nos últimos anos incluídos, 2020-2022, a série escalou para um novo nível arrecadatário. Para Figura 32 temos demonstrado a divisão do conjunto de observações para o treinamento e avaliação.

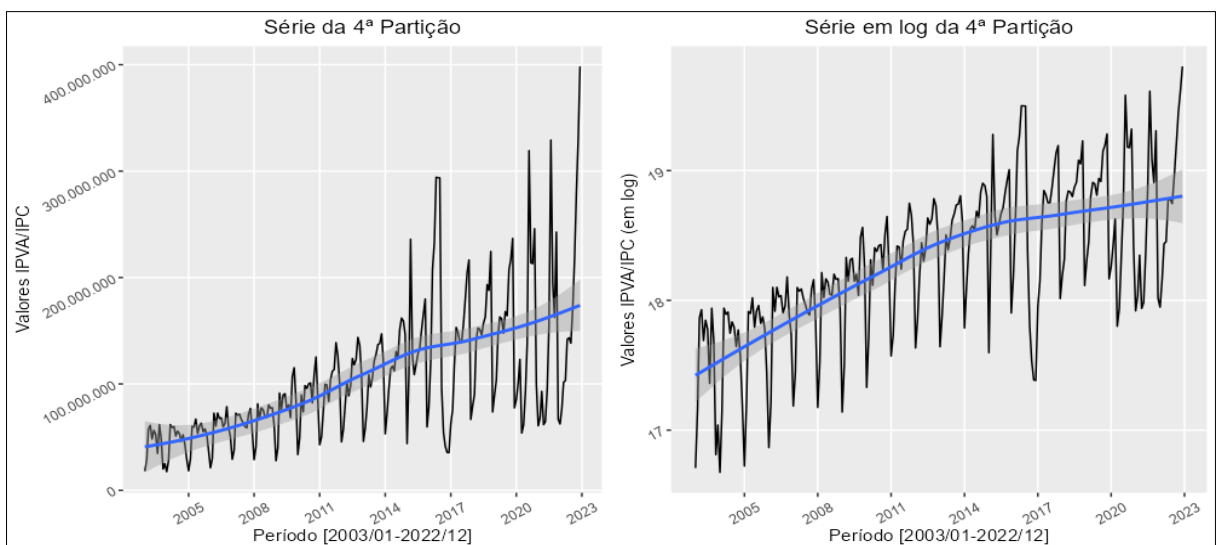


Figura 31 - Gráfico da Arrecadação IPVA x Série em log com Tendências, para a 4ª Partição

(fonte: elaboração própria)

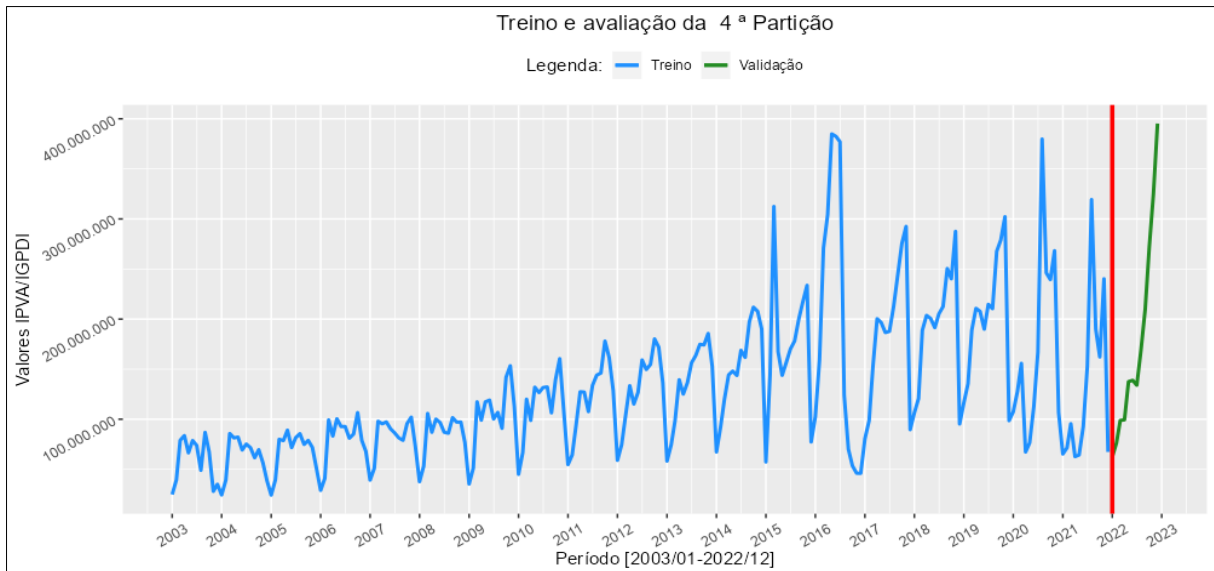


Figura 32 - Divisão dos dados em treino e validação para 4ª Partição

(fonte: elaboração própria)

O estudo dos resultados computados, apresentado pela Tabela 16, nos apresenta uma novidade. Comparativamente, para o MAPE calculado sobre os valores correntes demonstrou melhor resultado para o modelo ARIMA com medida de 55,37% enquanto para o SARIMA o valor foi 55,40%, variação de 0,03% em favor do primeiro modelo. Já comparando o resultado em escala logarítmica, o MAPE do método SARIMA foi 0,66% enquanto do ARIMA 1,94%, o pior resultado entre os modelos. Aliado aos melhores resultados percebido pela Tabela 17, R\$ 62.246.545,90 para o MAE, R\$ 105.092.055,57 para RMSE, 43,59 para o AIC e 67,22 para o BIC, temos confirmação do modelo SARIMA como o mais preciso. Entretanto, comparando as diferenças, mês a mês, observamos que se saiu melhor que os demais em 6 diferentes meses de 2022, e menos preciso que o ARIMA em 4 meses, que o HW-M em abril e que o HW-A em maio e para o total arrecadado. Além disso, a hegemonia do SARIMA em relação a 2020 foi menor que em 2019.

**Tabela 16 - Processamento dos modelos na 4ª partição, previsões: jan-dez de 2022**

Resumo: Previsões de Receita IPVA - IGPDI - 2022					
Mês/Ano	Arrecadação Real	HW-A	HW-M	ARIMA	SARIMA
jan/22	62.278.131,29	76.883.307,33	75.639.147,73	66.154.947,47	59.791.176,54
fev/22	76.268.894,38	95.544.651,10	90.372.849,43	100.157.437,68	77.386.933,60
mar/22	98.791.817,33	131.114.343,05	122.532.202,81	117.040.439,34	109.007.346,50
abr/22	99.368.438,19	107.854.705,78	102.980.992,34	124.016.992,29	77.675.104,78
mai/22	137.560.761,21	134.339.414,16	128.709.952,15	126.637.435,47	82.767.197,60
jun/22	138.853.506,35	168.427.688,16	160.970.212,49	127.518.570,60	111.005.064,64

jul/22	133.884.287,62	202.719.430,48	192.659.306,72	127.729.997,94	164.056.450,50
ago/22	169.135.770,25	237.195.136,31	221.482.805,99	127.687.497,69	292.420.840,14
set/22	209.635.417,88	150.801.901,12	144.457.842,77	127.549.725,43	207.877.495,51
out/22	271.483.231,99	149.850.378,70	145.214.311,78	127.376.799,47	190.559.413,02
nov/22	325.232.385,77	184.218.192,98	177.089.710,46	127.191.436,59	248.102.245,07
dez/22	395.267.657,56	68.912.516,74	69.405.696,78	127.002.208,15	79.734.080,65
<b>Total</b>	<b>2.117.760.299,83</b>	<b>1.707.861.665,91</b>	<b>1.631.515.031,47</b>	<b>1.426.063.488,13</b>	<b>1.700.383.348,56</b>
MAPE	68,71727685	68,01175346	55,36611049	55,40645278	
MAPE (p/ log)	1,368921037	1,526106314	1,939373382	0,662088792	

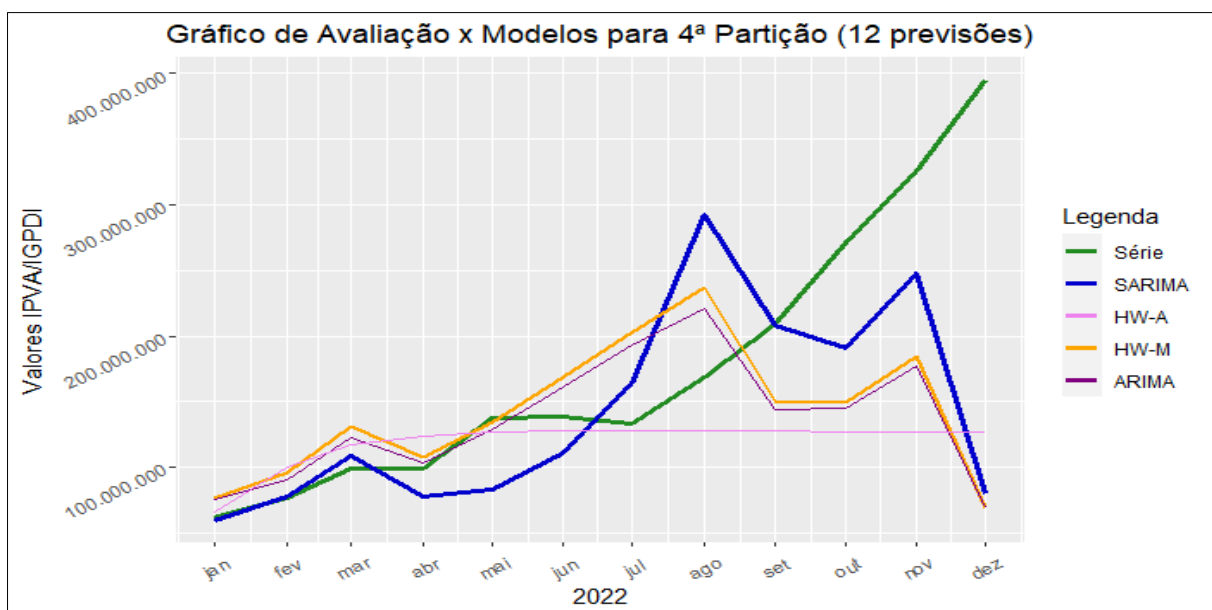
(fonte: elaboração própria)

**Tabela 17 - Outras avaliações dos métodos para 4º partição, jan-dez de 2022**

Medida	HW-A	HW-M	ARIMA	SARIMA
MAE	74.351.288,96	71.863.217,71	69.418.490,24	62.246.546,13
RMSE	114.238.641,33	114.032.349,77	108.865.813,32	105.092.055,85
AIC	684,26264617	683,81732904	176,54710141	43,59464036
BIC	742,5615219	742,1162047	197,1231752	67,22158921

(fonte: elaboração própria)

A vantagem do modelo SARIMA também pode ser percebida graficamente pela Figura 33 que, apesar de os valores previstos se afastarem os observados no final do ano de 2022, percebemos que a série esteve relativamente próxima dos valores de avaliação pelo menos 10 meses.



**Figura 33 - Comparação dos modelos x dados de avaliação da 4ª Partição**

(fonte: elaboração própria)

Revisando a etapa de treinamento, temos na Figura 34 o resultado dessa atividade, enquanto na Tabela 18 o resultado da medida MAPE do ajuste 0,89% e MAPE dos resíduos 47.382,93% como os melhores resultados coletados.

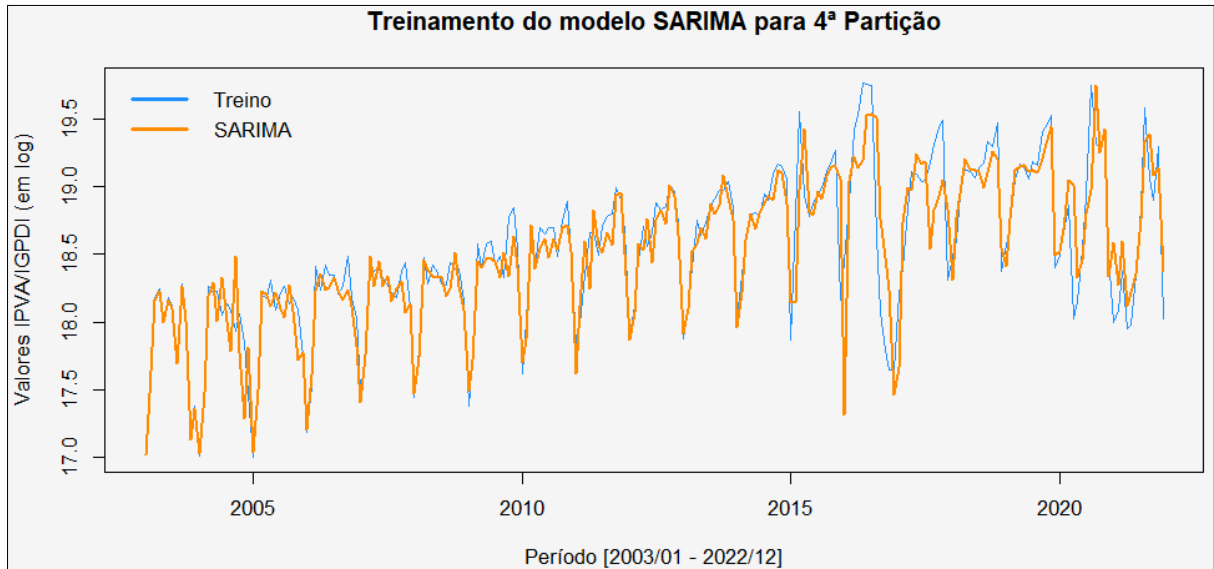


Figura 34 - Treinamento do SARIMA para 4ª Partição

(fonte: elaboração própria)

Tabela 18 - Avaliação dos ajustes e resíduos dos modelos para 4ª partição

Medida	HW-A	HW-M	ARIMA	SARIMA
MAPE (ajuste)	0,988524708	0,985695977	1,395542604	0,888601743
MAPE (resíduo)	52.443,99	1.632.096,58	57.188,61	47.382,99

(fonte: elaboração própria)

```

Coefficients:
      ar1      ar2      ma1      ma2      sar1      sma1
 0.0750  0.3569  0.6558 -0.0024  0.0221 -0.4602
s.e.    0.3242  0.1951  0.3206  0.1001  0.1999  0.1945

sigma^2 = 0.06745: log likelihood = -13.84
AIC=41.67  AICc=42.21  BIC=65.3

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Trainina set 0.03645467 0.2492504 0.1645765 0.1879154 0.8976855 0.7735111 -0.0189639
  
```

Figura 35 - Resultado do comando `summary()` sobre o modelo SARIMA - 4ª Partição

(fonte: elaboração própria)

A apresentação do resultado do comando `summary()`, na Figura 35, para a modelagem SARIMA reforça nossa observação gráfica que a verossimilhança possui o pior resultado quando comparado com outras partições, uma vez que no final do ano a previsão e os valores coletados seguem por direções diferentes. Ainda assim, o

modelo foi o que melhor representou os valores observados. Complementarmente, apresentamos na Figura 36 o gráfico que demonstra em azul os valores da previsão enquanto em vermelho os valores observados.

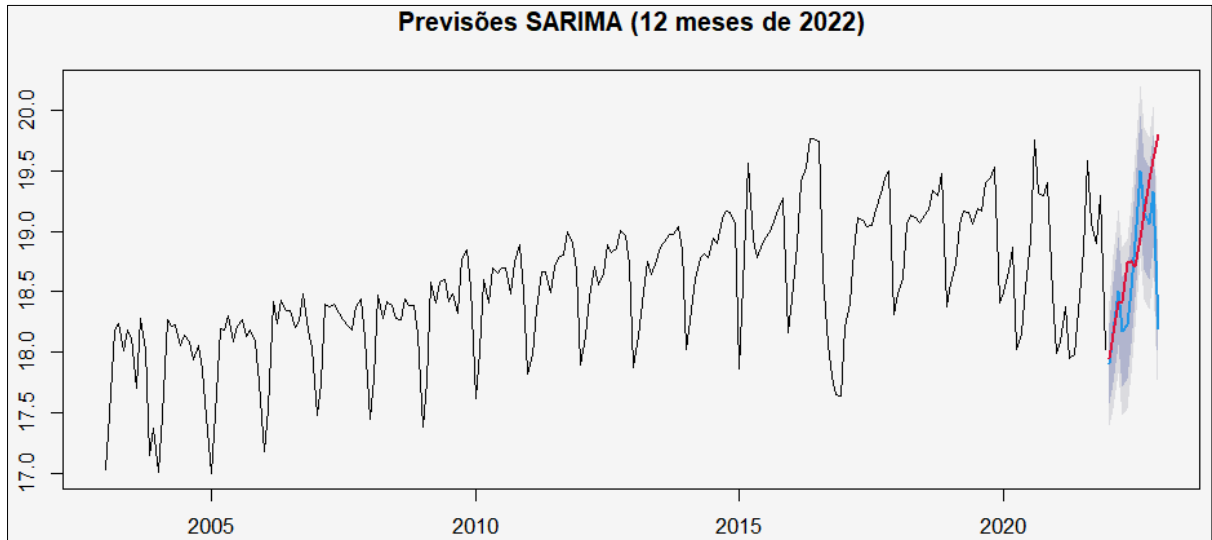


Figura 36 - Representação das previsões futuras para 4ª Partição

(fonte: elaboração própria)

## 6 – Conclusões

Ao término desta pesquisa, os resultados obtidos revelam importantes percepções sobre a eficácia dos modelos avaliados. O que demonstra que a estratégia de particionar a série em 4 grupos de investigação possibilitou consolidar a escolha. Conforme os dados foram apurados, pudemos certificar que os modelos SARIMA e HW-M apresentaram os melhores resultados.

O desempenho da previsão para série temporal do IPVA demonstrou a capacidade de capturar a sazonalidade e as tendências subjacentes aos dados, fornecendo previsões precisas ao longo do período de análise. Além disso, é importante ressaltar que os modelos Holt-Winters multiplicativo e SARIMA ficaram próximos na avaliação, pois ambos são modelos sazonais. No entanto, a escolha pelo modelo SARIMA, além de resultados ligeiramente melhores, se estabelece por se tratar de uma estratégia mais complexa, o que pode explicar sua pequena vantagem em relação ao modelo HW-M.

Por outro lado, foi perceptível que o modelo ARIMA, embora consagrado, não conseguiu rivalizar com o desempenho dos outros dois modelos. Esse resultado sugere que a estrutura de dependência temporal do IPVA pode não ser adequadamente capturada pelo ARIMA, ou que sua aplicação direta não se mostrou tão eficaz na presença de sazonalidade e tendências.

Além da sazonalidade característica do IPVA, a variação da amplitude ao longo tempo se justificada por vários fatores: crescimento populacional e do poder de aquisitivo do período que impulsiona a demanda por veículos novos, inflação brasileira que influencia o preço do bem, custo do veículo estimulado pelas novidades tecnológicas ou de segurança, adoção de veículos elétricos e incentivos governamentais para o consumo. Além do fato óbvio que a posse do bem perdura algum tempo e, por conseguinte, o contribuinte paga o imposto por anos seguidos, criando a clara tendência crescente da série.

Em resumo, esta dissertação proporcionou uma análise abrangente da previsão da série temporal do IPVA, destacando a eficácia do modelo SARIMA e do

método HW-M. Esses resultados são fundamentais para orientar a escolha de abordagens de previsão em contextos similares, onde a sazonalidade desempenha um papel significativo. Além disso, enfatiza a importância de considerar diferentes técnicas de modelagem para obter previsões precisas e confiáveis em séries temporais complexas e dinâmicas.

## REFERÊNCIAS BIBLIOGRÁFICAS

BONA, André. **Tabela Fipe: Você sabe como funciona?** André Bona. Disponível em: <<https://andrebona.com.br/tabela-fipe-voce-sabe-como-funciona/>>. Acesso em: 15 maio 2023.

CAMPELLO, Luiza de Figueiredo. A instituição do imposto de renda no Brasil. 2017.  
COELHO, Alvaro Luiz. Um histórico sobre a tributação no Brasil. **Revista da Receita Federal: estudos tributários e aduaneiros**, v. 3, n. 1–2, p. 29–49, 2017.

CONSELHO NACIONAL DE TRÂNSITO (BRAZIL) (Org.). **100 anos de legislação de trânsito no Brasil: 1910-2010**. Brasília: Contran : Denatran : Ministério das Cidades, 2010.

FERREIRA, Pedro. Análise de Séries Temporais em R.

FONTENELE, Sarah V.P. **Arrecadação Tributária: Modelos de Previsão para o Estado do Ceará**. Dissertação de Mestrado Profissional em Economia do Setor Público. Universidade Federal do Ceará, Fortaleza, 2017.

GOMES, Amanda dos Santos. **Modelagem e Previsão da Arrecadação do Imposto de Renda no Brasil**. Dissertação de Mestrado em Estatística – Universidade Federal de Pernambuco, Recife, 2003.

GUJARATI, Damodar N.; PORTER, Dawn C. **Basic econometrics**. 5th ed. Boston: McGraw-Hill Irwin, 2009.

HYNDMAN, Ron. **Forecasting: Principles and Practice (3rd ed)**. [s.l.: s.n.], 2018. Disponível em: <<https://otexts.com/fpp3/>>. Acesso em: 20 abr. 2023.

JUNIOR, José F. **Proposta de Modelo de Previsão do IPVA do Estado do Ceará**. Fortaleza: Instituto de Pesquisa e Estratégia Econômica do Ceará (IPECE), 2017. (Textos para Discussão n.124)

NIELSEN, Aileen. **Análise Prática de Séries Temporais: Predição com Estatística e Aprendizado de Máquina**. 1a edição. [s.l.]: Alta Books, 2021.

REIS, Tiago. **Tabela Fipe: o que é? Como calcular o preço dos automóveis?** Suno. Disponível em: <<https://www.suno.com.br/artigos/tabela-fipe/>>. Acesso em: 15 maio 2023.



ROCHA, Fabio G. **Contribuição de Modelos de Séries Temporais Para a Previsão da Arrecadação de ISS**. Dissertação de Mestrado em Economia. Universidade Estadual de Campinas, Campinas, 2003.

SIQUEIRA, Marcelo L. **Modelos de Séries Temporais para a Previsão da Arrecadação Tributária Federal**. Dissertação de Mestrado em Economia. Universidade Federal de Pernambuco, Recife, 2002.

SOUSA, Katsuk R.; CORREIA, J.J.A.; SANTOS, R.R.; SANTOS, J. F. Previsão de Receitas Tributárias: Uma análise do ICMS do Ceará por meio de séries temporais. **Refas**, v.5, n.5, p. 1-17, 2019.

SOUZA, Alberth A. A. **Comparação entre modelos de regressão e séries temporais para a arrecadação do IPVA em Sergipe**. Monografia de conclusão do curso de Bacharel em Estatística. Universidade Federal de Sergipe, Aracaju, 2016.

**Código Tributário do Estado de Goiás**. Disponível em: <<https://appasp.economia.go.gov.br/legislacao/arquivos/Cte/CTE.htm>>. Acesso em: 5 abr. 2023.

**Fipe - Fundação Instituto de Pesquisas Econômicas**. Fipe. Disponível em: <<https://www.fipe.org.br/>>. Acesso em: 6 maio 2023.

**L9503Compilado**. Disponível em: <[https://www.planalto.gov.br/ccivil\\_03/leis/l9503compilado.htm](https://www.planalto.gov.br/ccivil_03/leis/l9503compilado.htm)>. Acesso em: 23 abr. 2023.

**STF isenta barcos e aviões de pagar IPVA**. Jusbrasil. Disponível em: <<https://www.jusbrasil.com.br/noticias/stf-isenta-barcos-e-avioes-de-pagar-ipva/142213>>. Acesso em: 11 maio 2023.

**Teste de Kruskal-Wallis | PDF | Distribuição Qui Chi | Probabilidade e estatística**. Scribd. Disponível em: <<https://pt.scribd.com/document/371557797/Teste-de-Kruskal-wallis>>. Acesso em: 11 set. 2023.

Governo federal. Dados Abertos – Ministério da Fazenda. Disponível em: <https://dados.gov.br/dados/conjuntos-dados/boletim-de-arrecadacao-dos-tributos-estaduais>. Acesso em: 10 de abril de 2023

Governo federal. Dados Abertos – Ministério dos Transportes. Disponível em: fonte: <https://www.gov.br/transportes/pt-br/pt-br/assuntos/transito/conteudo-Senatran/frota-de-veiculos-2022>. Acesso em: 15 de abril de 2023

Governo do Estado de Goiás. Instituto Mauro Borges – Síntese de Indicadores Socioeconômicos, Painéis de Indicadores em fonte:

[https://www.imb.go.gov.br/index.php?option=com\\_content&view=article&id=88&Itemid=216](https://www.imb.go.gov.br/index.php?option=com_content&view=article&id=88&Itemid=216).

Acesso em: 05 de maio de 2023

## Anexo A – Avaliação dos métodos sem deflação

O principal objetivo do trabalho consiste na comparação dos métodos preditivos: HW (Aditivo e Multiplicativo), ARIMA e SARIMA sobre uma base de dados relevante para administração pública do Estado de Goiás. Entretanto, para utilização do método “profissionalmente” é necessário considerar a comparação deles sem deflação da série por algum índice. Isso ocorre, pois, ao deflacionarmos a série até o momento presente, teríamos que inferir o mesmo índice para meses seguintes e sobre os valores previstos. Dessa forma, percebe-se a possibilidade de inserir erros que comprometeriam a aplicação da estratégia. Diante disso e complementarmente a essa pesquisa, apresentamos os resultados sob a série sem deflação.

Em relação aos valores, conforme pudemos observar na Figura 9 as séries dos valores arrecadados e a deflacionada, possuem comportamentos semelhantes e tendência parecida, variando, essencialmente na escala.

Abaixo, na Tabela 19, são apresentamos os resultados da aplicação dos modelos HW-A, HW-M, SARIMA e SARIMA para aplicação na 1ª partição dos dados, janeiro de 2003 até dezembro de 2018 e previsão de 12 períodos de 2019. E em destaque, nas duas últimas linhas da tabela está destacado os valores da medida avaliativa MAPE, tanto para valores correntes e em logarítmico. E, os menores valores obtidos foram: 16,31% e 2,05%, respectivamente para série em valor corrente e logarítmica, demonstrando que SARIMA foi o melhor modelo para predição. E em comparação com o resultado obtido na seção 6.1 (14,16% e 1,80%), percebemos que a aplicação do método na série sem deflação obteve acurácia menor. Ainda na Tabela 19, investigando mais detalhadamente, mês a mês, observamos que o SARIMA se saiu melhor em 8 meses do ano, mas que não foi o melhor resultado para o total arrecadado do ano, perdendo, nesse requisito, por pouco, para o HW-M. Esse resultado também difere do observado quando aplicado a técnica na série deflacionada, que, na situação anterior, também se mostrou mais próximo do total.

*Tabela 19 - Resultado, sem deflação, na 1ª partição, previsões: jan-dez de 2019*

Resumo: Previsões de Receita IPVA - 2019					
Mês/Ano	Arrecadação Real	HW-A	HW-M	ARIMA	SARIMA
jan/19	70.839.309,70	44.928.904,64	47.071.063,15	53.124.590,87	63.444.731,19

fev/19	82.709.737,41	63.329.045,39	71.471.528,53	75.347.747,67	79.935.956,86
mar/19	116.734.994,16	114.026.129,36	127.283.857,37	85.405.383,66	120.102.662,30
abr/19	131.581.405,81	119.358.166,13	130.572.916,11	89.241.572,15	130.697.302,61
mai/19	130.907.946,52	118.001.990,58	136.643.922,99	90.537.289,63	127.898.468,98
jun/19	120.119.692,52	115.052.550,02	130.195.911,32	90.873.257,94	122.684.757,07
jul/19	136.741.236,39	120.902.322,59	139.652.345,65	90.856.955,24	127.414.067,10
ago/19	133.874.936,95	112.264.977,82	124.858.460,15	90.713.069,45	116.337.584,24
set/19	169.544.341,04	118.233.878,61	122.031.379,45	90.523.576,68	123.544.591,66
out/19	177.521.908,99	127.474.021,58	136.942.373,10	90.318.235,99	122.277.257,64
nov/19	193.473.014,82	119.116.879,92	127.574.846,26	90.107.828,73	134.009.337,44
dez/19	63.574.243,55	81.182.293,94	84.546.280,13	89.896.255,72	51.449.728,75
<b>Total</b>	<b>1.527.622.767,86</b>	<b>1.253.871.160,58</b>	<b>1.378.844.884,21</b>	<b>1.026.945.763,74</b>	<b>1.319.796.445,84</b>
MAPE	26,27928011	20,12907391	52,49481163	16,31633092	
MAPE (p/ log)	2,398735985	2,461775615	2,306734948	2,058996507	

(fonte: elaboração própria)

Para a Tabela 20, destacamos as menores ocorrências para cada medida comparativa, e podemos observar que, em todas, o SARIMA obteve melhor acurácia, tal qual ocorreu quando comparado com a série deflacionada pelo IGPDÍ.

*Tabela 20- Demais medidas comparativas para 1ª partição, jan-dez de 2019*

Medida	HW-A	HW-M	ARIMA	SARIMA
MAE	25.747.309,00	20.772.191,02	46.110.085,71	18.307.648,95
RMSE	33.028.068,66	28.452.153,79	53.868.393,57	27.892.007,59
AIC	509,88	544,08	138,15	38,89
BIC	565,26	599,45	157,70	61,24

E, a comparação gráfica apresentada pela Figura 37 demonstra o comportamento das previsões de cada método ao longo do ano de 2019.

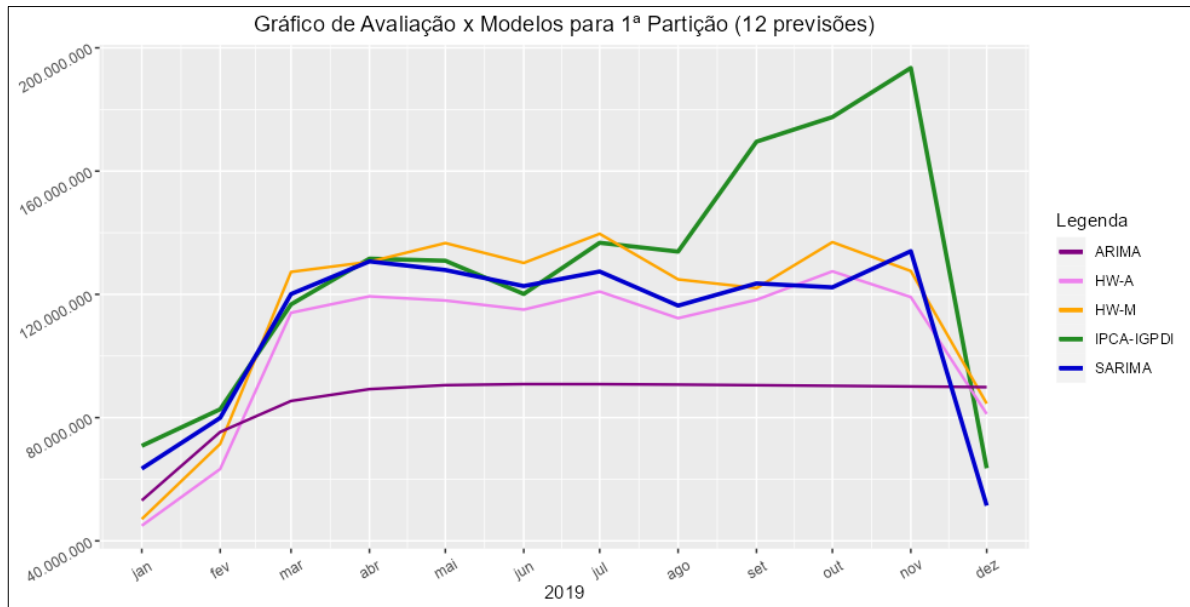


Figura 37 - Comparação dos modelos x série sem deflação da 1ª Partição

A parametrização utilizada para a série sem deflação continuou a mesma, SARIMA(2,0,2)(1,1,1)[12], onde:

- Dois termos autorregressivos (AR:  $p=2$ )
- Nenhuma diferença ( $d = 0$ )
- Dois termos de média móvel (MA:  $q=2$ )
- Um termo autorregressivo sazonal ( $P = 1$ )
- Uma diferença sazonal ( $D = 1$ )
- Um termo de média móvel sazonal ( $Q = 1$ )
- Um período sazonal de 12

Já para a 2ª partição, a Tabela 21 apresenta os resultados da aplicação das estratégias para obtenção dos valores para o ano 2020. E para essa situação, o resultado foi surpreendente, o método mais eficiente foi o ARIMA, que foi mais eficiente para 6 meses consecutivos, entre fevereiro e julho, seguido pela técnica HW-M para 4 meses seguintes, agosto-novembro. O método SARIMA destacou-se apenas em janeiro e em relação ao total arrecadado para o ano. Entretanto, mesmo não possuindo os resultados mais próximos, absolutamente, ele obteve os melhores resultados para o MAPE, 34,86% e 1,51% respectivamente para a série em valores correntes e em logarítmica. Dessa forma, podemos concluir que, assim como para a primeira partição, o SARIMA continua com a melhor acurácia preditiva. Também

apresentamos na Tabela 22 os resultados obtidos para as outras medidas, destacando os melhores resultados.

**Tabela 21 - Resultado, sem deflação, na 2ª partição, previsões: jan-dez de 2020**

Resumo: Previsões de Receita IPVA - 2020					
Mês/Ano	Arrecadação Real	HW-A	HW-M	ARIMA	SARIMA
jan/20	70.506.629,71	50.970.865,20	61.944.383,15	61.433.236,37	73.344.556,99
fev/20	83.941.036,18	74.063.967,16	91.300.192,76	85.489.071,73	85.594.742,62
mar/20	102.443.123,76	130.144.274,60	160.374.383,96	96.623.679,80	123.290.558,84
abr/20	44.893.109,67	134.756.924,33	154.274.751,77	101.021.389,71	135.379.007,21
mai/20	51.086.441,34	140.241.837,40	173.512.169,75	102.580.487,31	132.759.572,34
jun/20	76.244.793,49	136.636.200,91	180.197.061,98	103.030.884,85	123.497.133,36
jul/20	114.707.748,22	145.140.440,59	176.753.565,69	103.060.500,05	134.385.649,00
ago/20	266.869.997,55	131.773.575,76	154.041.815,75	102.932.441,15	128.583.401,72
set/20	179.635.888,01	131.292.038,16	163.200.695,55	102.745.906,80	151.132.106,93
out/20	180.447.380,17	140.952.636,30	167.520.263,58	102.538.096,78	153.732.982,93
nov/20	209.733.831,78	132.597.718,37	175.633.144,48	102.322.931,39	169.057.998,86
dez/20	85.617.403,62	83.404.340,92	111.923.709,73	102.105.613,76	58.417.936,98
<b>Total</b>	<b>1.466.127.383,50</b>	<b>1.431.974.819,69</b>	<b>1.770.676.138,14</b>	<b>1.165.884.239,70</b>	<b>1.469.175.647,78</b>
MAPE		41,38021761	35,51721422	49,7388128	34,86481214
MAPE (p/ log)		2,786308129	2,268449582	2,728852041	1,516799522

**Tabela 22 - Demais medidas comparativas para 2º partição, jan-dez de 2020**

Medida	HW-A	HW-M	ARIMA	SARIMA
MAE	52.436.790,54	56.187.967,00	50.427.705,83	43.817.367,64
RMSE	64.515.437,73	70.777.799,88	69.185.844,60	58.428.907,67
AIC	565,08	582,63	146,31	29,54
BIC	621,49	639,04	166,21	52,34

A Figura 38 destaca o comportamento dos valores previstos para de todos os métodos, ela confirma a percepção dos valores que, o SARIMA, não é a estratégia que mais próxima na maior parte do ano, entretanto, ela figura, globalmente, como o mais aderente ao comportamento dos valores observados no ano.

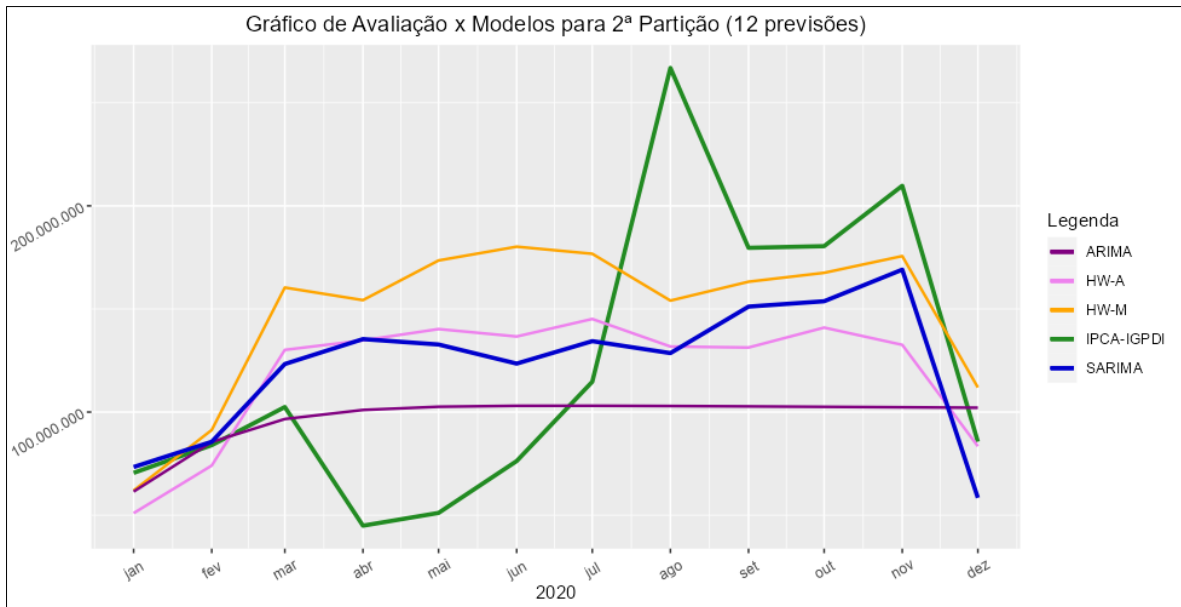


Figura 38 - Comparação dos modelos x série sem deflação da 2ª Partição

A 3ª partição produziu os resultados apresentados na Tabela 23. A investigação mensal apresenta o SARIMA como o processo mais eficiente, ele previu os melhores valores para 5 meses além do total do ano. E o segundo melhor modelo foi o HW-M para 4 meses. E essa situação se reflete também no MAPE, em relação ao valor corrente, o SARIMA possui o melhor resultado 25,06%, mas para a série logarítmica o HW-M foi bem melhor, com 1,99% frente aos 2,65% do SARIMA. Também apresentamos na Tabela 24 os resultados obtidos para as outras medidas, e, nesse caso, o SARIMA não teve sua vantagem ameaçada.

**Tabela 23 - Resultado, sem deflação, na 3ª partição, previsões: jan-dez de 2021**

Resumo: Previsões de Receita IPVA - 2021					
Mês/Ano	Arrecadação Real	HW-A	HW-M	ARIMA	SARIMA
jan/20	52.695.989,07	99.639.999,28	55.929.278,62	77.973.277,50	93.067.240,86
fev/20	59.436.673,48	110.199.629,53	82.411.772,88	94.624.444,95	98.085.460,00
mar/20	81.613.513,32	154.231.920,61	140.038.905,79	101.837.089,97	125.701.217,82
abr/20	54.639.378,81	143.568.138,86	136.149.467,27	104.634.841,23	79.459.210,86
mai/20	57.340.071,90	173.942.388,47	143.313.484,17	105.602.957,96	83.334.308,12
jun/20	85.526.769,76	194.796.034,88	141.263.854,80	105.846.688,57	98.299.550,07
jul/20	141.176.199,39	220.660.614,50	150.423.395,13	105.808.860,31	127.405.462,54
ago/20	299.266.290,13	229.727.418,59	141.227.654,73	105.662.647,03	198.387.377,81
set/20	177.773.052,09	187.742.922,59	143.672.781,33	105.475.146,48	172.343.273,86
out/20	150.826.959,30	193.736.814,50	154.264.201,32	105.272.262,56	173.829.764,06
nov/20	227.169.464,68	214.122.362,67	151.128.843,94	105.063.980,52	196.646.830,94
dez/20	63.251.319,77	97.595.933,54	91.837.904,57	104.854.145,74	73.509.397,18
<b>Total</b>	<b>1.450.715.681,70</b>	<b>2.019.964.178,04</b>	<b>1.531.661.544,54</b>	<b>1.222.656.342,82</b>	<b>1.520.069.094,13</b>

MAPE	38,36393641	38,34584944	57,20941193	25,06628818
MAPE (p/ log)	2,182515884	1,999190382	2,780393001	2,625026108

**Tabela 24 - Demais medidas comparativas para 3ª partição, jan-dez de 2021**

Medida	HW-A	HW-M	ARIMA	SARIMA
MAE	61.201.703,62	51.442.076,39	59.149.899,87	30.879.794,56
RMSE	69.487.473,71	67.038.243,73	76.487.388,06	39.246.889,09
AIC	628,50	659,97	162,85	48,95
BIC	685,88	717,35	183,10	72,18

A Figura 39 destaca o comportamento dos valores previstos para de todos os métodos, nela podemos perceber a dobradinha SARIMA e HW-M como métodos de melhor acurácia para arrecadação desse exercício.

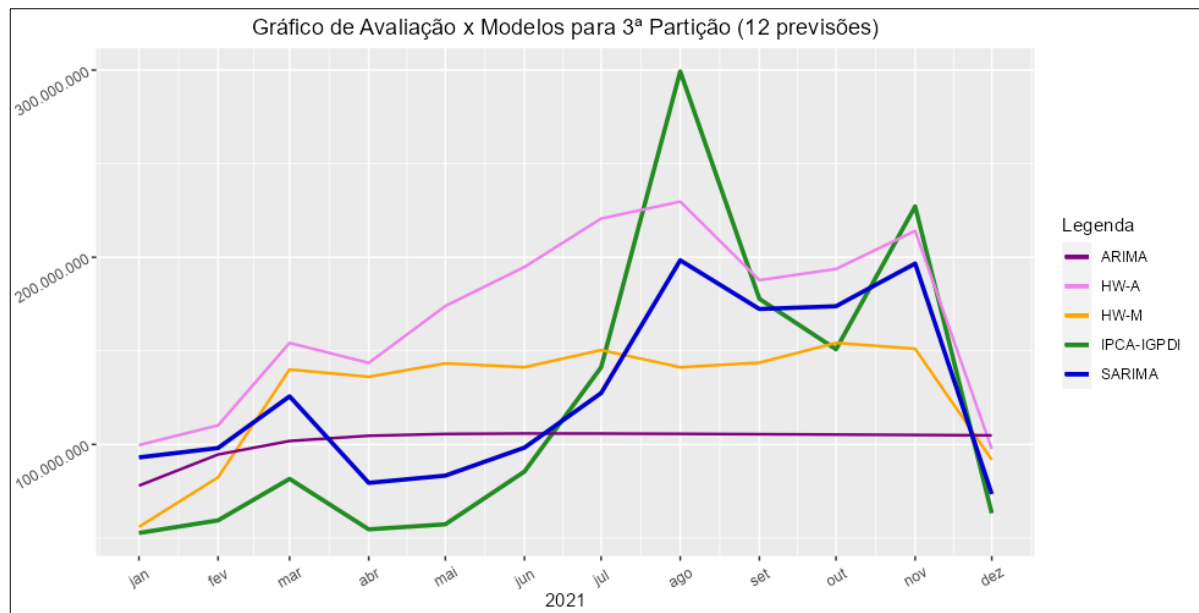


Figura 39 - Comparação dos modelos x série sem deflação da 3ª Partição

Para a última partição, a 4ª, os resultados foram os mais complexos de serem analisados. A observação mensal aponta como técnica mais efetivas os métodos HW-M e ARIMA, ambos, produzindo melhores resultados para 4 meses, seguido pelo SARIMA que foi o melhor para 3 meses. E, curiosamente, o HW-A foi o que resultou melhor resultado de arrecadação anual. Já a observação do MAPE, em relação à valores correntes, tanto HW-M quanto HW-A foram melhores, com valores respectivos de 60,31% e 65,88%. O SARIMA apresentou o 3º melhor resultado com 82,58%. Nos chama a atenção porcentagens tão altas, indicando que nenhum método foi capaz de prever adequadamente os valores arrecadados para esse ano. Já o MAPE para



escala logarítmica, o SARIMA foi o melhor, com resultado 0,76% enquanto os demais praticamente passam 1%. Adjunto, apresentamos a Tabela 26 que apresenta as demais medidas, destacados melhores resultados em vermelho. Nele também podemos observar que não houve método efetivo para prever os valores, destacamos os métodos HW-M e SARIMA como os melhores.

*Tabela 25 - Resultado, sem deflação, na 4ª partição, previsões: jan-dez de 2022*

Resumo: Previsões de Receita IPVA - 2021					
Mês/Ano	Arrecadação Real	HW-A	HW-M	ARIMA	SARIMA
jan/20	59.263.372,78	73.151.859,52	53.124.890,79	57.033.670,70	51.043.530,72
fev/20	74.037.691,45	93.074.186,85	77.410.909,57	81.098.270,49	62.831.761,80
mar/20	97.342.233,98	129.137.112,53	132.908.558,15	92.529.706,47	84.911.251,23
abr/20	100.229.851,80	106.090.114,20	121.907.227,66	97.140.312,21	55.602.006,99
mai/20	139.326.779,79	132.143.160,96	129.351.524,02	98.821.030,07	59.714.905,50
jun/20	141.606.757,29	167.534.030,27	126.655.170,35	99.341.810,04	84.752.356,39
jul/20	137.392.154,49	202.269.485,95	144.300.741,30	99.417.323,67	130.142.005,55
ago/20	172.915.911,34	240.986.207,14	142.406.153,05	99.324.146,08	257.110.243,21
set/20	213.134.958,74	152.886.711,23	137.389.612,51	99.167.639,75	174.072.277,62
out/20	272.658.061,80	152.952.006,72	147.606.742,79	98.987.705,39	159.058.843,25
nov/20	324.627.334,53	191.829.074,13	147.592.106,38	98.799.418,33	214.532.820,06
dez/20	393.841.549,30	71.226.720,01	86.027.688,28	98.608.468,02	68.099.440,45
<b>Total</b>	<b>2.126.376.657,29</b>	<b>1.713.280.669,51</b>	<b>1.446.681.324,83</b>	<b>1.120.269.501,22</b>	<b>1.401.871.442,78</b>
MAPE		65,88800849	60,31591027	86,26356805	82,58942005
MAPE (p/ log)		1,232285977	0,976393466	2,252617913	0,768123794

*Tabela 26 - Demais medidas comparativas para 4ª partição, jan-dez de 2022*

Medida	HW-A	HW-M	ARIMA	SARIMA
MAE	72.667.169,53	67.895.528,53	85.019.026,18	74.407.823,19
RMSE	112.116.789,32	112.008.777,97	126.386.697,86	112.439.033,57
AIC	685,73	730,30	183,51	49,18
BIC	744,03	788,59	204,09	72,81

A Figura 40 destaca o comportamento dos valores previstos para de todos os métodos, nela podemos perceber que, a partir de setembro, nenhum método acompanha mais a arrecadação. Esse fenômeno foi responsável pela indefinição da melhor estratégia.

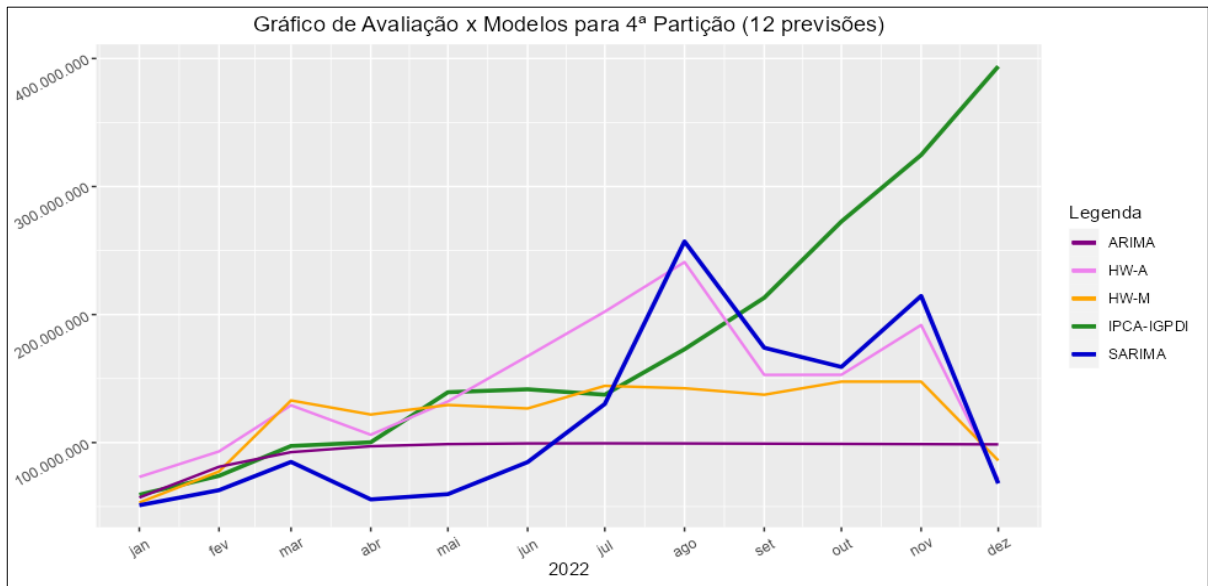


Figura 40 - Comparação dos modelos x série sem deflação da 4ª Partição

## Anexo B – Avaliação do SARIMA para a série com e sem deflação

Outra comparação interessante para esse estudo trata da investigação do melhor método, o SARIMA, quando aplicado na mesma série com e sem deflação. Essa avaliação irá comparar os resultados obtidos das 4 partições buscando destacar a melhor aplicação do método à realidade do Estado de Goiás.

As Tabela 27, Tabela 29, Tabela 31 e Tabela 33 apresentam para cada uma das 4 partições, anos 2019 até 2022, o valor da Arrecadação do IPVA Real, o valor da previsão pelo modelo SARIMA e a diferença absoluta entre esses, além disso, trás para o mesmo mês o valor da arrecadação atualizada pelo IGPDPI, o valor obtido pela modelo SARIMA e a diferença absoluta. Além disso, nessas tabelas estão destacados em verde os menores resultados das diferenças para cada uma das linhas. Além disso, para complementar a análise, as Tabela 28, Tabela 30, Tabela 32 e Tabela 34 apresentam demais medidas avaliativas dos resultados dos modelos, tanto para o resultado obtido, quanto avaliação do esforço do cálculo e, ainda, métricas da etapa de ajuste e do resíduo. E, como padrão dessa seção, destacamos os melhores resultados em verde.

*Tabela 27 - Comparação das diferenças absolutas da previsão SARIMA para série temporal com e sem deflação - 2019*

Comparação previsões 2019 do SARIMA com e sem deflação						
Mês/Ano	Arrecadação Real	SARIMA	Diferença	Arrecadação IGPDPI	SARIMA IGPDPI	Diferença IGPDPI
jan/19	70.839.309,70	63.444.731,19	7.394.578,51	116.181.501,70	97.945.247,25	18.236.254,45
fev/19	82.709.737,41	79.935.956,86	2.773.780,55	135.557.135,88	126.678.262,93	8.878.872,95
mar/19	116.734.994,16	120.102.662,30	-3.367.668,14	188.957.358,78	194.291.756,60	-5.334.397,82
abr/19	131.581.405,81	130.697.302,61	884.103,20	210.727.775,32	211.819.221,10	-1.091.445,78
mai/19	130.907.946,52	127.898.468,98	3.009.477,54	207.772.351,15	209.835.334,56	-2.062.983,41
jun/19	120.119.692,52	122.684.757,07	-2.565.064,55	189.890.269,60	202.882.960,02	-12.992.690,42
jul/19	136.741.236,39	127.414.067,10	9.327.169,29	214.811.077,23	212.060.525,45	2.750.551,78
ago/19	133.874.936,95	116.337.584,24	17.537.352,71	210.325.071,36	192.386.301,01	17.938.770,35
set/19	169.544.341,04	123.544.591,66	45.999.749,38	267.720.179,58	202.724.184,91	64.995.994,67
out/19	177.521.908,99	122.277.257,64	55.244.651,35	278.913.779,93	199.630.557,28	79.283.222,65
nov/19	193.473.014,82	134.009.337,44	59.463.677,38	302.313.990,89	217.408.148,30	84.905.842,59
dez/19	63.574.243,55	51.449.728,75	12.124.514,80	98.501.481,59	87.814.426,67	10.687.054,91
<b>Total</b>	<b>1.527.622.767,86</b>	<b>1.319.796.445,84</b>	<b>207.826.322,02</b>	<b>2.421.671.973,01</b>	<b>2.155.476.926,09</b>	<b>266.195.046,92</b>
MAPE		16,31633092			14,15810089	
MAPE (p/ log)		2,058996507			1,804176498	

Tabela 28 - Outras métricas avaliativas para o SARIMA - 2019

Medida	SARIMA	SARIMA IGPD
MAE	18.307.648,95	25.763.173,48
RMSE	27.892.007,59	39.556.918,17
AIC	38,89	30,61
BIC	61,24	52,96
MAPE (ajuste)	0,94	0,87
MAPE (resíduo)	101.584,48	53.704,47

Tabela 29 - Comparação das diferenças absolutas da previsão SARIMA para série temporal com e sem deflação - 2020

Comparação previsões 2020 do SARIMA com e sem deflação						
Mês/Ano	Arrecadação Real	SARIMA	Diferença	Arrecadação IGPD	SARIMA IGPD	Diferença IGPD
jan/20	70.506.629,71	73.344.556,99	-2.837.927,28	107.372.559,10	111.760.598,15	-4.388.039,04
fev/20	83.941.036,18	85.594.742,62	-1.653.706,44	127.712.587,57	134.144.068,41	-6.431.480,84
mar/20	102.443.123,76	123.290.558,84	-20.847.435,08	155.844.042,49	195.083.846,92	-39.239.804,43
abr/20	44.893.109,67	135.379.007,21	-90.485.897,54	67.189.702,83	214.431.976,07	-147.242.273,24
mai/20	51.086.441,34	132.759.572,34	-81.673.131,00	76.421.015,30	211.318.866,64	-134.897.851,34
jun/20	76.244.793,49	123.497.133,36	-47.252.339,87	112.847.558,59	197.789.575,88	-84.942.017,28
jul/20	114.707.748,22	134.385.649,00	-19.677.900,78	167.099.096,42	214.585.777,30	-47.486.680,88
ago/20	266.869.997,55	128.583.401,72	138.286.595,83	379.875.883,87	203.237.789,72	176.638.094,15
set/20	179.635.888,01	151.132.106,93	28.503.781,08	246.166.630,32	236.315.439,87	9.851.190,45
out/20	180.447.380,17	153.732.982,93	26.714.397,24	239.379.027,33	239.015.456,87	363.570,46
nov/20	209.733.831,78	169.057.998,86	40.675.832,92	268.352.523,06	260.646.000,23	7.706.522,83
dez/20	85.617.403,62	58.417.936,98	27.199.466,64	106.736.079,51	<b>92.869.196,80</b>	<b>13.866.882,72</b>
<b>Total</b>	<b>1.466.127.383,50</b>	<b>1.469.175.647,78</b>	<b>-3.048.264,28</b>	<b>2.054.996.706,40</b>	<b>2.311.198.592,84</b>	<b>-256.201.886,44</b>
MAPE		34,86			27,96	
MAPE (p/ log)		1,52			2,30	

Tabela 30 - Outras métricas avaliativas para o SARIMA - 2020

Medida	SARIMA	SARIMA IGPD
MAE	43.817.367,64	56.087.867,31
RMSE	58.428.907,67	82.914.667,45
AIC	29,54	21,19
BIC	52,34	43,99
MAPE (ajuste)	0,90	0,84
MAPE (resíduo)	198.280,17	60.576,02

**Tabela 31 - Comparação das diferenças absolutas da previsão SARIMA para série temporal com e sem deflação - 2021**

Comparação previsões 2021 do SARIMA com e sem deflação						
Mês/Ano	Arrecadação Real	SARIMA	Diferença	Arrecadação IGPI	SARIMA IGPI	Diferença IGPI
jan/21	52.695.989,07	93.067.240,86	-40.371.251,79	65.199.162,74	119.933.147,51	-54.733.984,76
fev/21	59.436.673,48	98.085.460,00	-38.648.786,52	71.460.637,03	135.108.760,98	-63.648.123,95
mar/21	81.613.513,32	125.701.217,82	-44.087.704,50	95.539.210,08	180.081.832,01	-84.542.621,94
abr/21	54.639.378,81	79.459.210,86	-24.819.832,05	62.603.525,92	120.740.419,33	-58.136.893,41
mai/21	57.340.071,90	83.334.308,12	-25.994.236,22	64.271.821,01	127.916.593,23	-63.644.772,22
jun/21	85.526.769,76	98.299.550,07	-12.772.780,31	92.715.879,46	149.549.338,90	-56.833.459,44
jul/21	141.176.199,39	127.405.462,54	13.770.736,85	152.872.645,16	190.031.120,36	-37.158.475,20
ago/21	299.266.290,13	198.387.377,81	100.878.912,32	319.442.185,76	280.004.636,04	39.437.549,72
set/21	177.773.052,09	172.343.273,86	5.429.778,23	190.018.437,78	242.747.690,70	-52.729.252,92
out/21	150.826.959,30	173.829.764,06	-23.002.804,76	162.100.399,78	240.594.934,67	-78.494.534,89
nov/21	227.169.464,68	196.646.830,94	30.522.633,74	240.312.633,61	266.077.934,10	-25.765.300,49
dez/21	63.251.319,77	73.509.397,18	-10.258.077,41	67.301.616,04	99.839.207,55	-32.537.591,50
<b>Total</b>	<b>1.450.715.681,70</b>	<b>1.520.069.094,13</b>	<b>-69.353.412,43</b>	<b>1.583.838.154,39</b>	<b>2.152.625.615,37</b>	<b>-568.787.460,98</b>
MAPE		25,07			33,82	
MAPE (p/ log)		2,63			1,85	

**Tabela 32 - Outras métricas avaliativas para o SARIMA - 2021**

Medida	SARIMA	SARIMA IGPI
MAE	30.879.794,56	53.971.880,04
RMSE	39.246.889,09	56.609.281,86
AIC	48,95	40,35
BIC	72,18	63,58
MAPE (ajuste)	0,94	0,87
MAPE (resíduo)	58.801,17	65.132,55

**Tabela 33 - Comparação das diferenças absolutas da previsão SARIMA para série temporal com e sem deflação - 2022**

Comparação previsões 2022 do SARIMA com e sem deflação						
Mês/Ano	Arrecadação Real	SARIMA	Diferença	Arrecadação IGPI	SARIMA IGPI	Diferença IGPI
jan/22	59.263.372,78	51.043.530,72	8.219.842,06	62.278.131,29	59.791.175,98	2.486.955,31
fev/22	74.037.691,45	62.831.761,80	11.205.929,65	76.268.894,38	77.386.933,08	-1.118.038,70
mar/22	97.342.233,98	84.911.251,23	12.430.982,75	98.791.817,33	109.007.346,52	-10.215.529,19
abr/22	100.229.851,80	55.602.006,99	44.627.844,81	99.368.438,19	77.675.106,98	21.693.331,21
mai/22	139.326.779,79	59.714.905,50	79.611.874,29	137.560.761,21	82.767.199,51	54.793.561,70
jun/22	141.606.757,29	84.752.356,39	56.854.400,90	138.853.506,35	111.005.066,19	27.848.440,16
jul/22	137.392.154,49	130.142.005,55	7.250.148,94	133.884.287,62	164.056.451,20	-30.172.163,58
ago/22	172.915.911,34	257.110.243,21	-84.194.331,87	169.135.770,25	292.420.833,58	-123.285.063,33
set/22	213.134.958,74	174.072.277,62	39.062.681,12	209.635.417,88	207.877.492,94	1.757.924,95
out/22	272.658.061,80	159.058.843,25	113.599.218,55	271.483.231,99	190.559.410,76	80.923.821,23
nov/22	324.627.334,53	214.532.820,06	110.094.514,47	325.232.385,77	248.102.241,82	77.130.143,95
dez/22	393.841.549,30	68.099.440,45	325.742.108,85	395.267.657,56	79.734.080,05	315.533.577,51
<b>Total</b>	<b>2.126.376.657,29</b>	<b>1.401.871.442,78</b>	<b>724.505.214,51</b>	<b>2.117.760.299,83</b>	<b>1.700.383.338,61</b>	<b>417.376.961,22</b>
MAPE		82,59			55,41	
MAPE (p/ log)		0,77			0,66	

**Tabela 34 - Outras métricas avaliativas para o SARIMA - 2022**

Medida	SARIMA	SARIMA IGPI
--------	--------	-------------

MAE	74.407.823,19	62.246.545,90
RMSE	112.439.033,6	105.092.055,57
AIC	49,18	43,59
BIC	72,81	67,22
MAPE (ajuste)	0,95	0,89
MAPE (resíduo)	72.388,93	47.382,93

Dos resultados obtidos, contabilizando as células destacadas, podemos perceber que em 60,42% dos melhores resultados ocorreram da aplicação do método SARIMA com a série temporal sem deflação. A Figura 41 demonstra mensalmente e comparativamente qual o menor erro de previsão, isto é, a linha que permanece mais próxima de zero representa o melhor previsão. Além disso, podemos ver como a previsão ora está mais próxima com deflação IGPD I ora sem deflação. Por outro lado, observando o total arrecadado ao final do exercício, para os anos 2019 e 2022, o SARIMA sobre a série sem deflação foi melhor, enquanto para os exercícios 2020 e 2021, o resultado foi melhor sobre a série temporal com deflação IGPD I. E mais, como demonstra a Figura 42, o erro da previsão sobre a série sem deflação foi significativamente menor quando acumulado os erros.

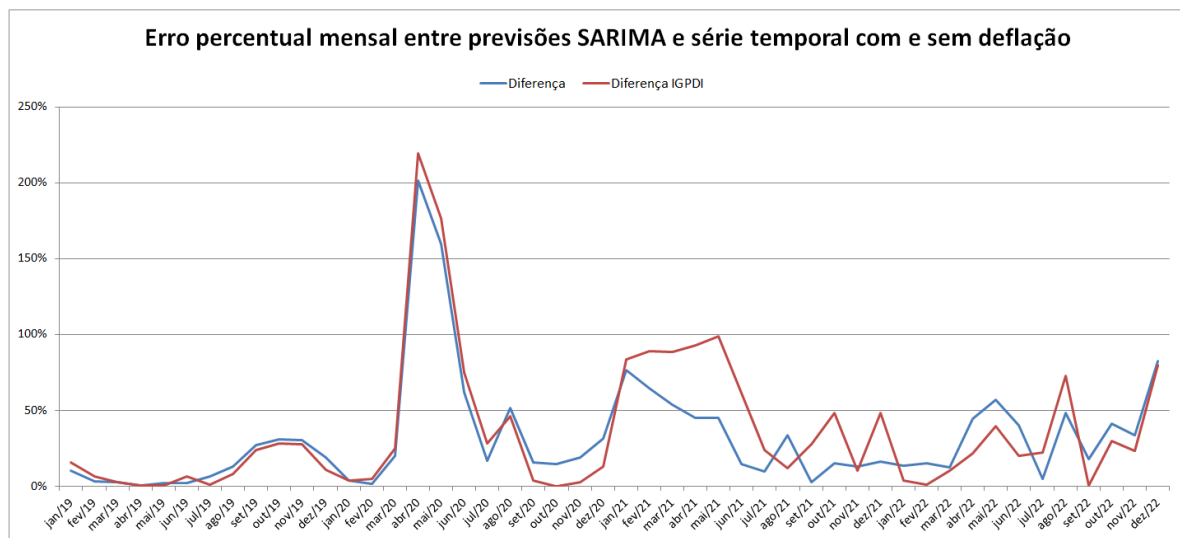


Figura 41 - Erro percentual mensal entre previsões SARIMA e série temporal com e sem deflação

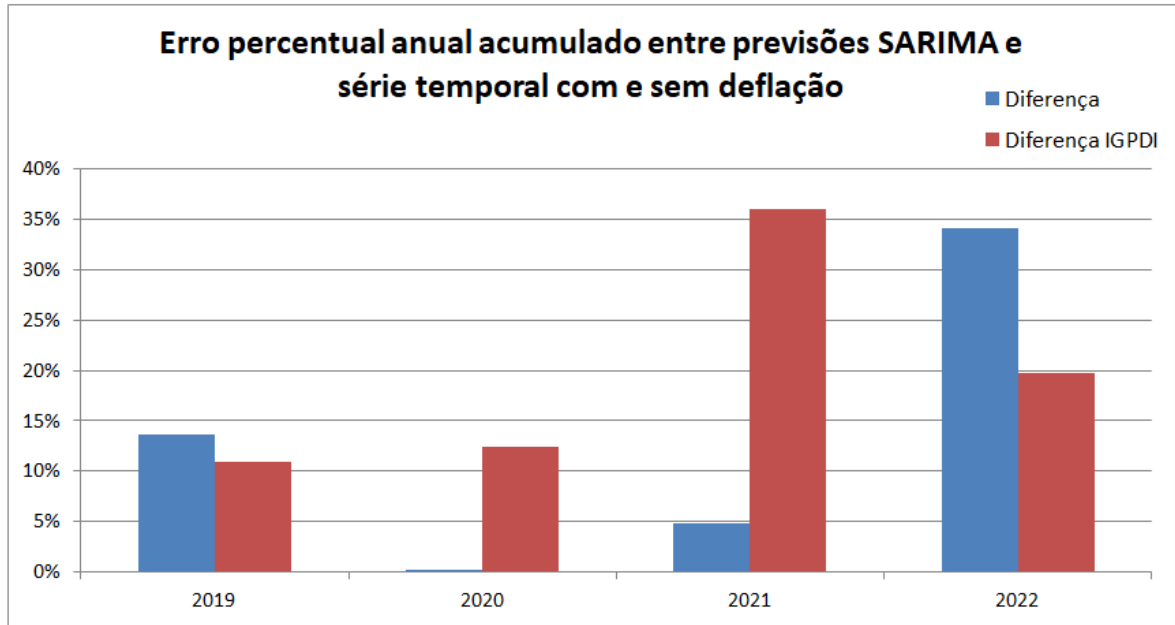


Figura 42 - Erro percentual anual acumulado entre previsões SARIMA e série temporal com e sem deflação