

**Secretaria da Fazenda do Estado do Rio Grande do Sul**  
**Tesouro do Estado**  
**Divisão de Estudos Econômicos e Fiscais e Qualidade do Gasto**

**Modelo SARIMA de previsão para o ICMS no Rio Grande do Sul**

Clarissa Black<sup>1</sup>

**TEXTOS PARA DISCUSSÃO TE/RS Nº 20**

**janeiro/2022**

Publicação cujo objetivo é divulgar resultados de estudos direta ou indiretamente desenvolvidos pelo Tesouro do Estado, ou de interesse da instituição, os quais, por sua relevância, levam informações para profissionais especializados e estabelecem um espaço para sugestões e debates de ideias. Todas as contribuições recebidas passam, necessariamente, por avaliação de admissibilidade e por análise dos pares. As opiniões emitidas nesta publicação são de exclusiva e inteira responsabilidade do(s) autor(es), não exprimindo, necessariamente, o ponto de vista do órgão.

---

<sup>1</sup>Doutora em Economia pela Universidade Federal do Rio Grande do Sul (UFRGS). Pesquisadora em Economia da Fundação de Economia e Estatística (FEE), no período 2011-2018. Atualmente no Tesouro do Estado, Secretaria da Fazenda do Rio Grande do Sul. Contato: [clarissablack@gmail.com](mailto:clarissablack@gmail.com)

## **Resumo**

Este artigo apresenta a modelagem autorregressiva e de médias móveis com sazonalidade (SARIMA) para a previsão da arrecadação do Imposto de Circulação de Mercadorias e Serviços (ICMS), no Estado do Rio Grande do Sul. Uma das justificativas para o trabalho é melhorar a previsibilidade da principal receita tributária do Estado para garantir, entre outros, o cumprimento dos percentuais mínimos constitucionais a serem gastos em saúde e educação. A estratégia empírica envolveu trabalhar com dois modelos SARIMA em diferentes subamostras, os quais registraram um razoável desempenho nas previsões fora da amostra, mesmo considerando um horizonte temporal relativamente amplo, em 36 meses.

## 1. Introdução

O objetivo deste trabalho é aplicar a modelagem autorregressiva e de médias móveis com sazonalidade (SARIMA) para a previsão da arrecadação do Imposto de Circulação de Mercadorias e Serviços (ICMS), no Rio Grande do Sul.

Os modelos econométricos do tipo SARIMA consistem em uma modelagem univariada, portanto com a ausência de variáveis explicativas exógenas. Esses modelos são muito utilizados quando o intuito é fazer previsão de qualquer tipo de série temporal, pois são modelos nos quais “a variável fala por si”, sendo necessário somente o conhecimento do seu histórico passado e presente para projetar o futuro.

A justificativa para o trabalho é melhorar as ferramentas de previsão da arrecadação para aumentar a eficiência da gestão das finanças públicas estaduais, em geral, e, particularmente, garantir o cumprimento dos gastos mínimos constitucionais com saúde e educação, por exemplo.

Vale lembrar que o ano de 2021 representou um período de expressivo crescimento da arrecadação de tributos, com destaque para o desempenho do ICMS, com uma taxa aproximada de 30%, em comparação com 2020, no Rio Grande do Sul. O cenário foi marcado pela recuperação econômica, pela inflação de dois dígitos e pela operação de privatização da Companhia Estadual de Distribuição de Energia Elétrica (CEEE-D).

Nesse período, um aspecto desafiador à gestão das finanças públicas estaduais envolveu o cumprimento dos gastos mínimos constitucionais em saúde e educação, os quais são calculados com base na Receita Líquida de Impostos e Transferências (RLIT)<sup>2</sup>. Quando a arrecadação ocorre muito acima das expectativas, como ocorreu em 2021, a qualificação dos instrumentos de previsão permite uma maior clareza no horizonte de previsibilidade, para melhorar a gestão fiscal e garantir que esses percentuais mínimos constitucionais sejam alcançados.

Diante da importância desta temática para o Tesouro do Estado e para atender ao objetivo proposto, este artigo tem mais duas seções, além desta introdução, das considerações finais e referências bibliográficas consultadas. Na seção 2 é discutido o método econométrico, enquanto a seção 3 apresenta a aplicação empírica e os principais resultados encontrados. A estratégia inicial envolveu trabalhar com dois modelos SARIMA. O primeiro foi encontrado

---

<sup>2</sup> A RLIT é calculada com base na soma das receitas de impostos diretamente arrecadados pelo Estado, sendo o ICMS o mais relevante, das transferências recebidas de outros entes da federação, deduzidas as transferências aos municípios.

através da inspeção visual das funções de autocorrelação e autocorrelação parcial. O segundo tem a mesma configuração do conhecido modelo dos passageiros de avião.

Os dois modelos geraram previsões próximas e um desempenho razoável, mesmo considerando um horizonte temporal relativamente amplo, em 36 passos fora da amostra. Em outra subamostra com dados até janeiro de 2021, em um passo a frente, o primeiro modelo foi melhor, enquanto em seis passos acumulados, o *airline model* registrou menor desvio. Por fim, foram realizadas previsões em 12 passos para cada ano do intervalo 2012-2020. Os dois modelos se saíram bem, mas, na maior parte do período, o modelo  $(0,1,1)(0,1,1)_{12}$  registrou maior acerto.

## 2. Metodologia

Na econometria de séries temporais há basicamente dois tipos de modelagem, a univariada e a multivariada. A depender do objetivo proposto e das características da série, uma ou outra abordagem se mostra mais apropriada. Quando o intuito é explicar os determinantes da série a ser estudada, um modelo com variáveis explicativas (exógenas) é a melhor estratégia, portanto, nesse caso, a abordagem multivariada é recomendada. Outra situação na qual se utiliza esse tipo de modelagem é no desenho de cenários.

Por outro lado, se o objetivo é fazer previsão, especialmente no curto prazo, a modelagem multivariada se mostra mais sujeita à má especificação. Isso porque ao utilizar esse tipo de modelo para fazer previsão são necessárias algumas etapas, as quais tendem a piorar a capacidade preditiva dos modelos.

A primeira etapa de um método multivariado é a elaboração de um modelo teórico, o qual sempre envolve algum grau de abstração da realidade. Um segundo passo consiste em encontrar indicadores que representem as variáveis explicativas definidas no modelo teórico, as quais não se observam diretamente e significam outro tipo de abstração. Por fim, há ainda que se fazer uma previsão das variáveis explicativas, para somente depois disso realizar a previsão final, portanto, tem-se uma previsão com base em outras previsões, o que tende a acumular e aumentar os erros de previsão.

Diante disso, quanto o intuito é realizar uma previsão fora da amostra, em geral, modelos univariados são mais recomendados, por serem também mais parcimoniosos. Nesse tipo de modelo, a amostra de dados contém apenas a própria série que se deseja fazer previsão, o que dispensa a adição de variáveis explicativas. Assim, o método prescinde de um modelo teórico e dos diferentes níveis de abstração citados anteriormente.

Em perspectiva histórica, os primeiros métodos de previsão univariada surgiram com o modelo conhecido como Exponentially Weighted Moving Average (EWMA) de Holt (1957) e Winters (1960). Apesar do EWMA trazer uma ideia intuitiva ao decompor a série em nível, inclinação e sazonalidade, esses modelos não contêm uma solidez estatística, pois não fazem diferenciação entre a série estimada e o seu resíduo. Portanto, não é possível fazer inferência nem teste de hipótese, por exemplo, pois os EWMA são considerados determinísticos.

Frente a essa lacuna, surgiram os modelos autorregressivos, integrados e de médias móveis (ARIMA) de Box e Jenkins (1976), como uma importante contribuição em modelagem de previsão. Apesar desse tipo de modelo ser menos intuitivo do que o EWMA, ele dispõe de uma base estatística robusta.

Diante disso, esta seção discute inicialmente a metodologia SARIMA, a qual representa a extensão do modelo ARIMA para o caso sazonal, e apresenta o conhecido modelo dos passageiros de avião. Na sequência são discutidas as principais medidas de diagnóstico para modelos de previsão.

## 2.1 Modelos SARIMA

Os modelos ARIMA (p,d,q) quando contêm uma parcela sazonal (autorregressiva e de média móvel) se transformam em SARIMA (Seasonal Autorregressive Integrated Moving Average) (p,d,q)(P,D,Q)<sub>s</sub>, o qual considera a sazonalidade do tipo multiplicativa<sup>3</sup>.

Ao analisar a notação dos modelos SARIMA, a parte “AR” representa a quantidade “p” de termos autorregressivos, ou seja, se for um AR (1), há um termo defasado em um período, ou seja,  $y_{(t-1)}$ , e “p” é igual a 1. A parte “I” significa que a série precisa ser diferenciada “d” vezes para tornar-se estacionária. Já “MA” representa a quantidade de termos de média móvel “q”.

No caso da sazonalidade, pode ser necessária ainda uma diferenciação sazonal “D” e a inclusão de termos autorregressivos sazonais “P” e de médias móveis sazonais “Q”, com  $s=12$ , para o caso mensal, usualmente.

O modelo geral de sazonalidade multiplicativa SARIMA (p,d,q)(P,D,Q)<sub>s</sub> é dado pela equação a seguir, com base em Nelson, (p. 174, 1973), na qual se supõe que os resíduos  $\varepsilon_t$  tenham média zero, variância constante, sejam normais e não autocorrelacionados serialmente:

---

<sup>3</sup> Um procedimento alternativo para a sazonalidade é trabalhar com *dummies*, o que forneceria um caráter aditivo para a sazonalidade.

$$\underbrace{\varphi_p(L)}_{\text{AR}} \underbrace{\phi_p(L^s)^d}_{\text{SAR}} \underbrace{\Delta_d \Delta_s}_{\text{dif.}} y_t = \underbrace{\theta_q(L)}_{\text{MA}} \underbrace{H_Q(L^s)}_{\text{SMA}} \varepsilon_t \quad (1)$$

Na qual:

- $L$  indica a utilização de um operador de defasagens e  $L^s$  é a defasagem sazonal;
- $L^n y_t = y_{t-n}$ ;
- $\varphi_p(L)$  representa o polinômio:  $\varphi_p(L) = (1 - \varphi_1 L^1 - \varphi_2 L^2 - \dots - \varphi_p L^p)$  para um AR(p). No caso de um AR(1),  $y_t = \varphi_1 y_{(t-1)} + \varepsilon_t$ ;
- $\phi_p(L^s)$  representa o polinômio SAR(P), se for um SAR(1),  $y_t = \phi_{1s} y_{(t-1s)} + \varepsilon_t$ , com  $s$  representando a sazonalidade, no caso mensal, geralmente  $s=12$ ;
- $\Delta = (1 - L)$  e  $\Delta_s = (1 - L^s)$  representam a primeira diferença e a diferença sazonal, respectivamente;
- $\theta_q(L)$  exibe o polinômio de MA(q), no caso de um MA(1),  $y_t = \varepsilon_t - \theta_1 \varepsilon_{(t-1)}$ ;
- $H_Q(L)$  é a parte SMA(Q), sendo  $y_t = \varepsilon_t - \theta_1 \varepsilon_{(t-1s)}$ , se for um SMA(1), por exemplo;

Um exemplo clássico de representação da modelagem SARIMA é o modelo de previsão das vendas de passagens aéreas, cuja série temporal estudada é o total de passageiros de avião, nos Estados Unidos, de janeiro de 1949 a dezembro de 1960. O modelo que melhor se ajusta à série é um SARIMA (0,1,1)(0,1,1)12, exemplo usual em livros-texto de econometria básica. Outra forma de escrever o modelo (0,1,1)(0,1,1)12, a partir da equação 1 é dado pela equação 2:

$$(1 - L)(1 - L^{12})y_t = (1 - \theta_1 L)(1 - \theta_{12} L^{12})\varepsilon_t \quad (2)$$

Vale observar ainda que, a modelagem do tipo SARIMA tem uma relação com modelos estruturais em espaço de estado, conforme Harvey (p. 511, 1989), os quais decompõem a série em componentes não observáveis, como a tendência, o ciclo e a sazonalidade. Como exemplo, modelos ARIMA (0,1,1) podem ser relacionados aos modelos de nível local, enquanto modelos de tendência e ciclo podem ser modelados como um ARIMA (2,2,4).

O caso do modelo estrutural básico, com tendência e sazonalidade (e sem o componente cíclico), pode ser aproximado ao *airline model*, por hipótese, em alguns casos, conforme Harvey (p. 511, 1989), em séries temporais com sazonalidade proeminente, como parece ser o caso da série do ICMS mensal.

A estimação dos modelos SARIMA ocorre por Mínimos Quadrados Ordinários (MQO) e as principais etapas consistem em:

**a)** testar se a série é estacionária, o que pode ser feito por meio de um teste de Dickey-Fuller Aumentado (DICKEY; FULLER, 1979; SAID; DICKEY, 1984), por exemplo. Caso a série seja não estacionária, ou, em outras palavras, tenha raiz unitária, é necessária a diferenciação da série para torná-la estacionária e evitar o fenômeno conhecido como regressão espúria.

**b)** No caso da sazonalidade, pode ser necessária ainda uma diferenciação sazonal, caso a média e a variância não sejam constantes na frequência sazonal. Há três formas de investigar a presença de raiz unitária sazonal. A primeira é através de um teste de raiz unitária sazonal, o mais conhecido é o teste HEGY, o qual remete às iniciais dos autores Hylleberg, Engle, Granger e Yoo. Outra forma é observar se a função de autocorrelação apresenta decaimento lento na parte sazonal ( $t=12, 24, 36, \dots$ ). Por fim, a inspeção visual do gráfico que plota os valores mensais da série pode indicar sinais de não estacionariedade sazonal da série.

**c)** identificar a quantidade de parâmetros autorregressivos, de médias móveis e os componentes sazonais do modelo. Para isso, é feita a análise das funções de autocorrelação, para encontrar os termos autorregressivos, e de autocorrelação parcial, para encontrar os componentes de médias móveis. Podem ser utilizados ainda, critérios de informação como o de Akaike e Schwarz, por exemplo;

**d)** verificar se os resíduos são homocedásticos, normais e não autocorrelacionados, conforme as premissas dos modelos de MQO;

**e)** Por fim, é feita a previsão fora da amostra e verificado o seu desempenho em termos da sua capacidade preditiva. A subseção a seguir, apresenta as principais medidas de *performance* em modelos de previsão.

Diante disso, este artigo envolveu a aplicação empírica de dois modelos SARIMA. O primeiro foi identificado através das funções de autocorrelação e autocorrelação parcial e o modelo encontrado foi  $(1,1,1)(1,1,2)_{12}$ . O segundo modelo estimado trata-se do  $(0,1,1)(0,1,1)_{12}$ , configuração do conhecido *airline model*.

Na prática, a busca pelo modelo ideal é alcançada através da experimentação, haja vista as combinações de parâmetros serem infinitas. Assim, o processo não é apenas a aplicação da ciência, mas também envolve a arte no desenho do modelo e na busca pela configuração mais apropriada.

A expectativa é a de que as combinações de parâmetros propostas não esgotem a discussão, mas sirvam de ponto de partida para a melhoria dos métodos de previsão do ICMS, no Estado. Para isso é fundamental testar outras combinações possíveis de parâmetros e acompanhar os erros de previsão fora da amostra, assunto da próxima subseção, a qual exhibe as principais medidas de desempenho preditivo.

## 2.2 Análise de desempenho da previsão fora da amostra

Estatísticas como o R-quadrado, não são apropriadas para o diagnóstico de um modelo de previsão. Isso porque envolve o cálculo dos desvios da previsão dentro da amostra, ou seja, entre os valores estimados e os valores observados, considerando uma amostra  $T=t, t+1, \dots, t+n$ . O ideal, no caso de modelos de previsão, é analisar o seu desempenho fora da amostra  $T$ , ou seja, em  $T+1, T+2, \dots, T+h$ , com a quantidade de passos previstos fora da amostra dado por  $h$ .

Ao considerar  $y_t$  como os valores observados e  $\hat{y}_t$  os valores previstos, uma medida usual de *performance* na previsão é a estatística MAPE, que são as iniciais de *mean absolute percentage error*, calculada pela fórmula a seguir, a qual apresenta a soma dos erros percentuais em módulo:

$$MAPE = \left(\frac{1}{n}\right) \sum_{t=T+1}^{T+n} (|y_t - \hat{y}_t| / y_t) \quad (3)$$

Alternativamente, para o caso da previsão do ICMS, analisar a soma dos valores previstos em relação aos valores observados é uma medida interessante, pois permite que desvios de previsão positivos em um mês sejam compensados por desvios negativos em outro mês. A fórmula para o erro percentual acumulado é dada por:

$$\text{Erro acumulado \%} = \frac{\sum_{t=T+1}^{T+h} (\hat{y}_t - y_t)}{\sum_{t=T+1}^{T+h} (y_t)} \quad (4)$$

Outra medida usual de desvio de previsão é o coeficiente de desigualdade de Theil (TIC), cujo resultado é um número entre zero e um. Quanto mais próximo de zero, melhor é a previsão. O cálculo do chamado U de Theil é descrito como:



$$\text{TIC} = \frac{\sqrt{\frac{\sum_{t=T+1}^{T+h} (\hat{y}_t - y_t)^2}{h}}}{\sqrt{\frac{\sum_{t=T+1}^{T+h} \hat{y}_t^2}{h} + \frac{\sum_{t=T+1}^{T+h} y_t^2}{h}}} \quad (5)$$

Dentre outras medidas de desempenho preditivo podem ser citadas ainda a raiz do erro quadrado médio (RSME) e o erro médio absoluto (MAE), dadas pelas equações (6) e (7) a seguir:

$$\text{RSME} = \sqrt{\frac{\sum_{t=T+1}^{T+h} (\hat{y}_t - y_t)^2}{h}} \quad (6)$$

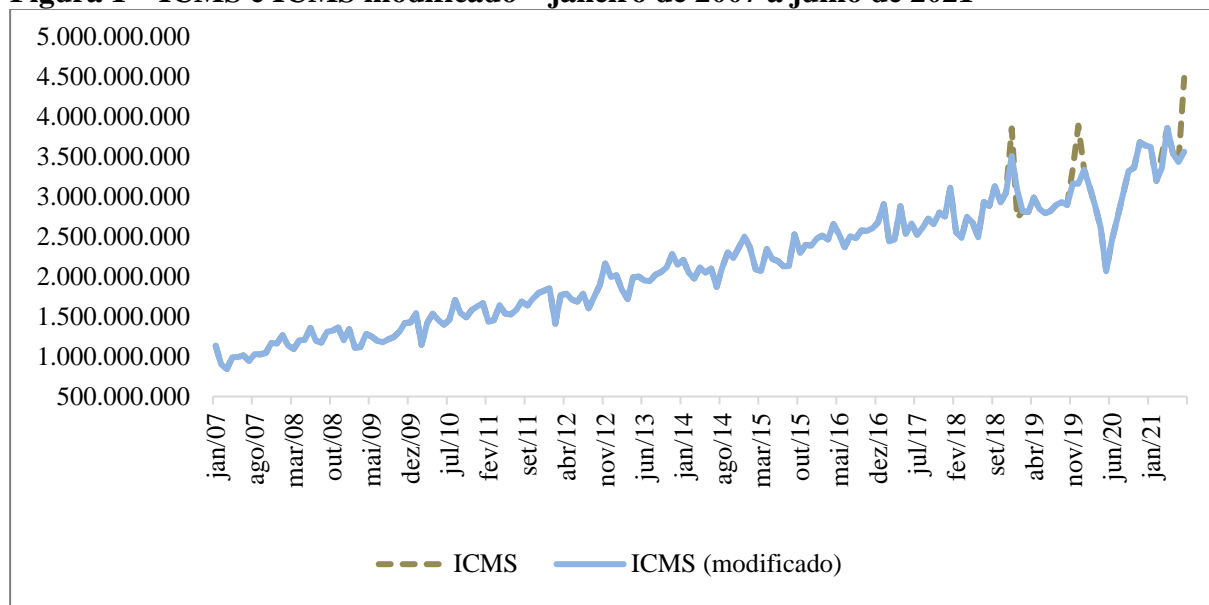
$$\text{MAE} = \frac{\sum_{t=T+1}^{T+h} |\hat{y}_t - y_t|}{h} \quad (7)$$

Quanto menores forem os valores encontrados nessas estatísticas, melhor é a capacidade preditiva dos modelos. Neste artigo foram analisadas as estatísticas dadas por (3), (4) e (5). Após a exibição das principais medidas de erro de previsão, a subseção a seguir apresenta a aplicação empírica e a análise dos resultados.

### 3. Aplicação empírica e análise dos resultados

A primeira etapa da aplicação empírica consistiu na realização de alguns ajustes na série do ICMS, cuja fonte é a Receita Estadual, conforme o gráfico abaixo. A amostra contém dados do ICMS nominal em reais (R\$), de janeiro de 2007 a julho de 2021. As modificações referem-se à eventos extraordinários, especialmente a partir do final de 2018.

**Figura 1 – ICMS e ICMS modificado – janeiro de 2007 a julho de 2021**



Fonte dos dados brutos: Rio Grande do Sul (2021a).

Nota: A série do ICMS modificado subtraiu do mês de dezembro de 2018 e somou no mês de janeiro de 2019 R\$ 347 milhões, devidos a antecipações do pagamento do ICMS. A série ainda exclui de novembro (R\$ 192 milhões) e de dezembro de 2019 (R\$ 728 milhões) os valores referentes a programas de parcelamento extraordinário e decisões judiciais. Excluem de julho de 2021 os R\$ 922 milhões referentes à CEEE-D. Acrescenta em fevereiro de 2021 e deduzem de março de 2021 os R\$ 170 milhões referentes à postergação de pagamentos de energia elétrica (RIO GRANDE DO SUL, 2021b, 2021c).

O segundo passo envolveu a verificação da estacionariedade da série. Para isso foi realizado um teste de Dickey-Fuller Aumentado, com constante, cuja estatística de teste ficou calculada em 0,158551 e o p-valor foi de 0,9399. Assim, a hipótese nula de raiz unitária não é rejeitada. Já a primeira diferença da série é estacionária, cuja estatística de teste é -8,735785. No caso de dados mensais, torna-se necessário analisar ainda se a série é estacionária na frequência sazonal. Por meio da inspeção visual dos valores mensais da série, verificam-se indícios de não estacionariedade sazonal.

Na sequência foi realizada a inspeção da função de autocorrelação e de autocorrelação parcial da série diferenciada (de primeira ordem e sazonalmente)<sup>4</sup> para identificar a quantidade de termos autorregressivos e de médias móveis. Com algum esforço, a análise das funções parecem indicar, inicialmente, um modelo SARIMA (1,1,1)(1,1,2)<sub>12</sub>.

Em outras palavras, de acordo com a função de autocorrelação, há termos autorregressivos significativos no *lag* 1 e no *lag* 12, ou seja AR(1) e SAR(12). Já a função de autocorrelação parcial permite identificar termos significativos no *lag* 1, no *lag* 12 e no *lag* 24, ou seja, MA(1), SMA(12) e SMA(24), com outros *lags* em zonas de indefinição. Observa-

<sup>4</sup>  $\Delta_1 \Delta_{12} y_t = (1 - L)(1 - L^{12})y_t$ .

se que a estimação do modelo SARIMA (1,1,1)(1,1,2)12, em uma subamostra de dados até julho de 2018, resultou em alguns coeficientes não significativos, conforme a tabela a seguir.

**Tabela 1 – Coeficientes estimados - SARIMA (1,1,1)(1,1,2)12**

Variável	Coefficiente	Desvio-padrão	Estatística t	p-valor
AR(1)	-0.039824	0.101636	-0,39183	0.6960
SAR(12)	-0.676964	0.081400	-8,31656	0.0000
MA(1)	-0.880361	0.039073	-22,53122	0.0000
SMA(12)	-0.008614	0.031785	-0,27100	0.7869
SMA(24)	-0.823749	0.029309	-28,10523	0.0000

Fonte: elaborado pela autora.

Ao retirar o termo constante, o termo AR(1) e o termo SMA(24), a estimativa resultou no parâmetro SAR(12) não significativo. Ao retirá-lo do modelo, os termos significativos passaram a ser MA(1) e SMA(12), o que resulta na configuração do conhecido modelo dos passageiros de avião, ou seja (0,1,1)(0,1,1)12.

Portanto, são estabelecidos dois modelos potenciais, quais sejam: (a) (1,1,1)(1,1,2)12; (b) (0,1,1)(0,1,1)12, chamado na literatura de *airline model*. O modelo (a) exibido na Tabela 1 anterior resultou em resíduos sem correlação serial, conforme o teste de Breusch-Godfrey (estatística F de 1,366694 e p-valor de 0,2594). O R-quadrado foi calculado em 0,690367. O modelo também apresentou resíduos normais (estatística de Jaque-Bera de 2,874215 e p-valor de 0,237614).

Com relação às previsões em 36 passos (T+36), a tabela abaixo reporta as seguintes medidas: erro (valor previsto – valor observado); erro percentual (erro/valor observado); erro acumulado percentual (somatório dos erros/somatório dos valores observados, conforme equação 4); erro percentual médio absoluto MAPE (média do somatório da razão entre a diferença em módulo entre o valor observado e o valor previsto em relação ao valor observado), conforme equação (3) na seção 2.2.

**Tabela 2 – Previsões em 36 passos à frente, considerando estimativas calculadas em uma subamostra de dados de janeiro de 2007 a julho de 2018 – SARIMA (1,1,1)(1,1,2)12**

Passos	Período	Previsões	ICMS (modificado)	Erro	Erro %	Erro acumulado %	MAPE (%)
1	ago/18	2.759.389.898	2.880.273.792	-120.883.894	-4,20%	-4,20%	4,20%
2	set/18	2.892.354.190	3.126.414.425	-234.060.235	-7,49%	-5,91%	5,84%
3	out/18	2.844.536.065	2.926.666.495	-82.130.430	-2,81%	-4,89%	4,83%
4	nov/18	2.958.556.004	3.047.325.169	-88.769.165	-2,91%	-4,39%	4,35%
5	dez/18	2.908.948.743	3.503.477.023	-594.528.280	-16,97%	-7,24%	6,87%

6	jan/19	3.088.873.551	3.076.899.082	11.974.469	0,39%	-5,97%	5,79%
7	fev/19	2.787.759.122	2.814.176.075	-26.416.953	-0,94%	-5,31%	5,10%
8	mar/19	2.804.431.611	2.803.894.230	537.381	0,02%	-4,69%	4,46%
9	abr/19	2.962.902.850	2.987.288.337	-24.385.487	-0,82%	-4,27%	4,06%
10	mai/19	2.894.428.691	2.839.760.343	54.668.348	1,93%	-3,68%	3,85%
11	jun/19	2.805.364.860	2.791.424.001	13.940.859	0,50%	-3,32%	3,54%
12	jul/19	2.954.458.730	2.818.447.445	136.011.285	4,83%	-2,68%	3,65%
13	ago/19	2.937.808.832	2.894.912.023	42.896.809	1,48%	-2,37%	3,48%
14	set/19	3.019.765.050	2.930.610.422	89.154.628	3,04%	-1,98%	3,45%
15	out/19	3.024.719.799	2.895.240.042	129.479.757	4,47%	-1,56%	3,52%
16	nov/19	3.137.265.243	3.160.537.766	-23.272.523	-0,74%	-1,51%	3,34%
17	dez/19	3.142.411.431	3.156.623.194	-14.211.763	-0,45%	-1,44%	3,17%
18	jan/20	3.240.154.120	3.334.674.001	-94.519.881	-2,83%	-1,53%	3,16%
19	fev/20	2.920.690.438	3.123.355.070	-202.664.632	-6,49%	-1,80%	3,33%
20	mar/20	2.935.849.276	2.887.472.732	48.376.544	1,68%	-1,63%	3,25%
21	abr/20	3.177.268.338	2.606.217.214	571.051.124	21,91%	-0,65%	4,14%
22	mai/20	3.053.027.885	2.066.995.110	986.032.775	47,70%	0,89%	6,12%
23	jun/20	3.063.574.719	2.454.386.207	609.188.512	24,82%	1,77%	6,93%
24	jul/20	3.046.065.131	2.729.325.271	316.739.860	11,61%	2,15%	7,13%
25	ago/20	3.107.784.919	3.016.264.056	91.520.863	3,03%	2,19%	6,96%
26	set/20	3.217.055.196	3.319.074.031	-102.018.835	-3,07%	1,96%	6,81%
27	out/20	3.186.572.007	3.358.904.152	-172.332.145	-5,13%	1,66%	6,75%
28	nov/20	3.300.104.186	3.675.444.420	-375.340.234	-10,21%	1,14%	6,87%
29	dez/20	3.268.184.718	3.635.784.668	-367.599.950	-10,11%	0,67%	6,99%
30	jan/21	3.421.561.721	3.618.514.343	-196.952.622	-5,44%	0,42%	6,93%
31	fev/21	3.114.519.823	3.192.427.368	-77.907.545	-2,44%	0,32%	6,79%
32	mar/21	3.130.703.347	3.357.156.462	-226.453.115	-6,75%	0,08%	6,79%
33	abr/21	3.315.969.723	3.854.845.127	-538.875.404	-13,98%	-0,46%	7,01%
34	mai/21	3.229.481.041	3.534.051.892	-304.570.851	-8,62%	-0,73%	7,05%
35	jun/21	3.172.595.046	3.428.958.710	-256.363.664	-7,48%	-0,95%	7,06%
36	jul/21	3.267.869.993	3.558.349.766	-290.479.773	-8,16%	-1,18%	7,10%

Fonte: elaborado pela autora.

Com base na tabela anterior é possível observar que os erros em termos acumulados ficaram inferiores a -7,24%, enquanto com base na estatística MAPE, os erros foram iguais ou menores do que 7,10%, durante o período. O coeficiente de Theil, dado pela equação (5) foi calculado em 0,048729. Vale notar que, esses resultados incluem os grandes desvios provocados pela crise econômica provocada pela Covid-19 (entre os passos 21 e 24). Ao considerar os erros acumulados até a o passo 17 (dezembro de 2019), que corresponde ao período anterior à crise, a estatística MAPE ficou calculada em 3,17% e o erro acumulado foi de -1,44%.

Por sua vez, o segundo modelo a ser estimado é um *airline model*, dado por  $(0,1,1)(0,1,1)12$ , no mesmo intervalo entre janeiro de 2007 e julho de 2018, conforme a Tabela 3 a seguir.

**Tabela 3 – Coeficientes estimados – SARIMA (0,1,1)(0,1,1)12**

Variável	Coefficiente	Desvio-padrão	Estatística t	p-valor
MA(1)	-0.827537	0.050251	-16,46794	0.0000
SMA(12)	-0.844881	0.031681	-26,66861	0.0000

Fonte: elaborado pela autora.

Os resíduos não apresentaram correlação serial, de acordo com o teste de Breusch-Godfrey (estatística F de 0,337731 e p-valor de 0,7141). O R-quadrado foi calculado em 0,597625, o coeficiente MA(1) foi estimado em -0,827537\* (estatística t=-16,46794) e o SMA(12) foi calculado em -0,844881\* (estatística t=-26,66861), ambos significativos estatisticamente, em um intervalo de confiança de 99%. As previsões em 36 passos para o modelo estão exibidas na tabela a seguir:

**Tabela 4 – Previsões em 36 passos à frente, considerando estimativas calculadas em uma subamostra de dados de janeiro de 2007 a julho de 2018 – SARIMA (0,1,1)(0,1,1)12**

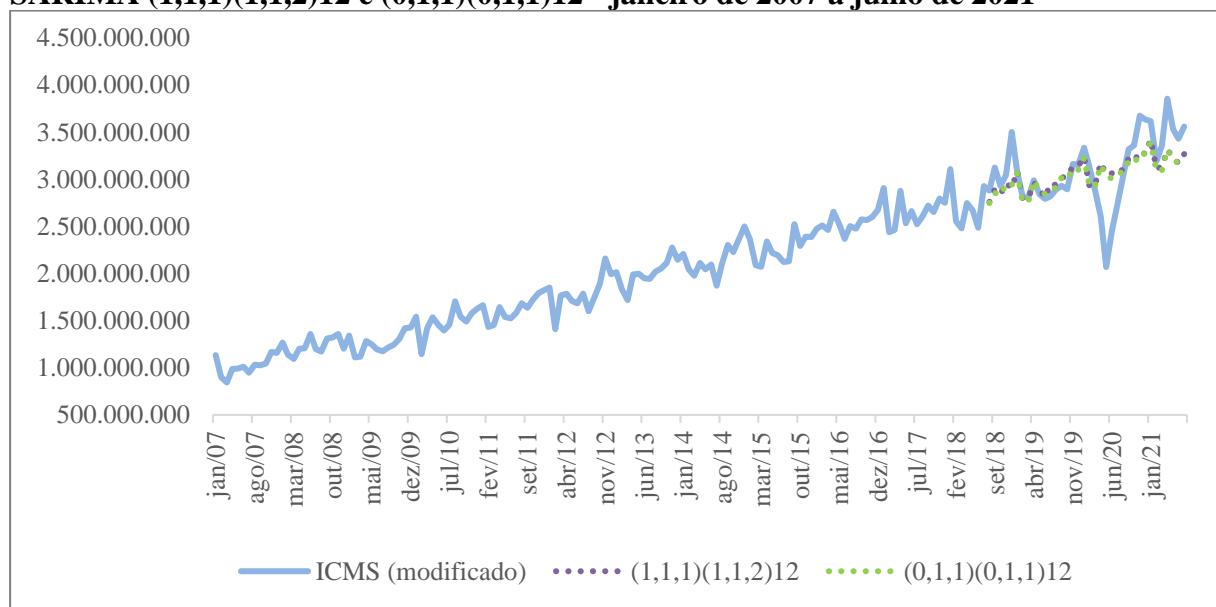
Passos	Período	Previsões	ICMS (modificado)	Erro	Erro %	Erro acumulado %	MAPE (%)
1	ago/18	2.747.949.041	2.880.273.792	-132.324.751	-4,59%	-4,59%	4,59%
2	set/18	2.856.916.450	3.126.414.425	-269.497.975	-8,62%	-6,69%	6,61%
3	out/18	2.834.139.159	2.926.666.495	-92.527.336	-3,16%	-5,53%	5,46%
4	nov/18	2.951.217.579	3.047.325.169	-96.107.590	-3,15%	-4,93%	4,88%
5	dez/18	2.937.571.903	3.503.477.023	-565.905.120	-16,15%	-7,47%	7,14%
6	jan/19	3.062.976.563	3.076.899.082	-13.922.519	-0,45%	-6,31%	6,02%
7	fev/19	2.761.359.058	2.814.176.075	-52.817.017	-1,88%	-5,72%	5,43%
8	mar/19	2.764.548.904	2.803.894.230	-39.345.326	-1,40%	-5,22%	4,93%
9	abr/19	2.985.919.216	2.987.288.337	-1.369.121	-0,05%	-4,65%	4,38%
10	mai/19	2.876.831.095	2.839.760.343	37.070.752	1,31%	-4,09%	4,08%
11	jun/19	2.842.725.650	2.791.424.001	51.301.649	1,84%	-3,58%	3,87%
12	jul/19	2.891.999.178	2.818.447.445	73.551.733	2,61%	-3,09%	3,77%
13	ago/19	2.912.181.979	2.894.912.023	17.269.956	0,60%	-2,82%	3,52%
14	set/19	3.021.149.387	2.930.610.422	90.538.965	3,09%	-2,40%	3,49%
15	out/19	2.998.372.096	2.895.240.042	103.132.054	3,56%	-2,01%	3,50%
16	nov/19	3.115.450.516	3.160.537.766	-45.087.250	-1,43%	-1,97%	3,37%
17	dez/19	3.101.804.841	3.156.623.194	-54.818.353	-1,74%	-1,96%	3,27%
18	jan/20	3.227.209.501	3.334.674.001	-107.464.500	-3,22%	-2,03%	3,27%
19	fev/20	2.925.591.995	3.123.355.070	-197.763.075	-6,33%	-2,27%	3,43%
20	mar/20	2.928.781.841	2.887.472.732	41.309.109	1,43%	-2,09%	3,33%
21	abr/20	3.150.152.153	2.606.217.214	543.934.939	20,87%	-1,14%	4,17%
22	mai/20	3.041.064.033	2.066.995.110	974.068.923	47,12%	0,41%	6,12%
23	jun/20	3.006.958.588	2.454.386.207	552.572.381	22,51%	1,22%	6,83%
24	jul/20	3.056.232.116	2.729.325.271	326.906.845	11,98%	1,64%	7,05%
25	ago/20	3.076.414.917	3.016.264.056	60.150.861	1,99%	1,65%	6,84%
26	set/20	3.185.382.325	3.319.074.031	-133.691.706	-4,03%	1,40%	6,74%
27	out/20	3.162.605.034	3.358.904.152	-196.299.118	-5,84%	1,10%	6,70%
28	nov/20	3.279.683.454	3.675.444.420	-395.760.966	-10,77%	0,57%	6,85%
29	dez/20	3.266.037.779	3.635.784.668	-369.746.889	-10,17%	0,12%	6,96%
30	jan/21	3.391.442.439	3.618.514.343	-227.071.904	-6,28%	-0,13%	6,94%
31	fev/21	3.089.824.933	3.192.427.368	-102.602.435	-3,21%	-0,24%	6,82%

32	mar/21	3.093.014.779	3.357.156.462	-264.141.683	-7,87%	-0,50%	6,85%
33	abr/21	3.314.385.091	3.854.845.127	-540.460.036	-14,02%	-1,02%	7,07%
34	mai/21	3.205.296.971	3.534.051.892	-328.754.921	-9,30%	-1,30%	7,13%
35	jun/21	3.171.191.526	3.428.958.710	-257.767.184	-7,52%	-1,50%	7,15%
36	jul/21	3.220.465.054	3.558.349.766	-337.884.712	-9,50%	-1,75%	7,21%

Fonte: elaborado pela autora.

Para o modelo  $(0,1,1)(0,1,1)12$ , os erros em termos acumulados ficaram inferiores a -7,47%, ao passo que a estatística MAPE registrou 7,21% como o percentual máximo. O coeficiente de Theil foi calculado em 0,048702. Ao analisar os erros acumulados até a previsão para dezembro de 2019, a estatística MAPE ficou calculada em 3,27% e o erro acumulado de -1,96%. O Gráfico a seguir ilustra os valores observados e aqueles previstos pelos dois modelos, os quais apresentam resultados relativamente próximos.

**Figura 2 - ICMS modificado e previsão fora da amostra em trinta e seis passos – SARIMA  $(1,1,1)(1,1,2)12$  e  $(0,1,1)(0,1,1)12$  - janeiro de 2007 a julho de 2021**



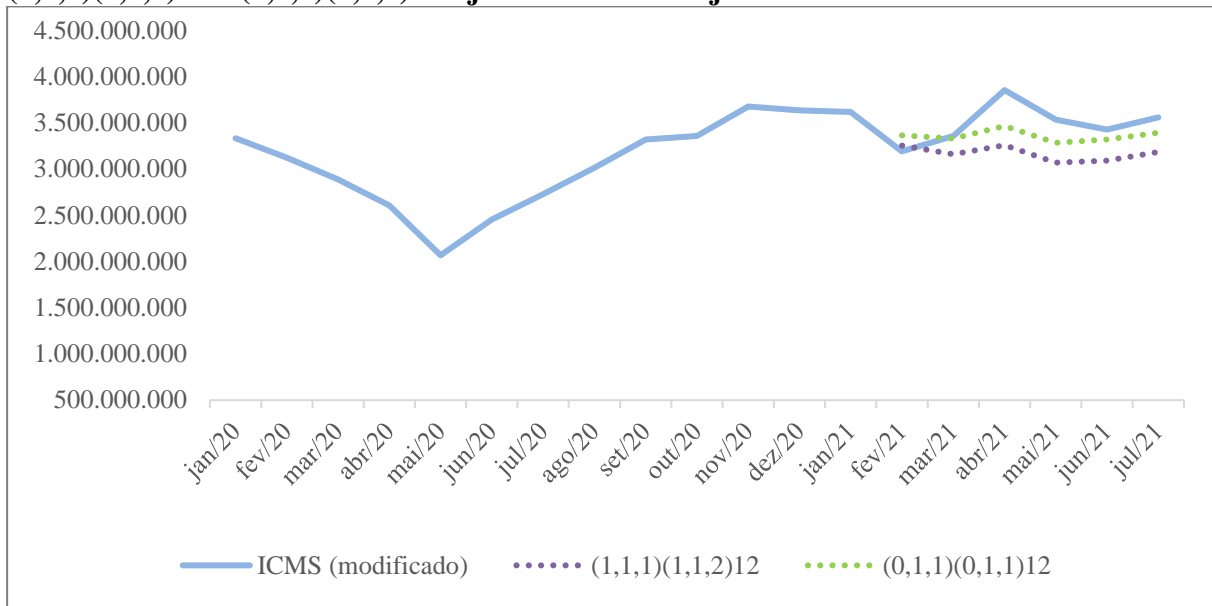
Fonte: elaborado pela autora.

Na sequência, considerou-se outra subamostra, com dados entre janeiro de 2007 a janeiro de 2021, para realizar previsões para os dois modelos SARIMA, em seis passos. Com base no gráfico a seguir, é possível observar novamente que os dois modelos produzem estimativas dentro do intervalo de confiança, mas o *airline model* se aproximou mais dos valores observados.

Quanto aos desvios de previsão, em um passo à frente, o modelo SARIMA  $(1,1,1)(1,1,2)12$  registrou desvio de 1,9% em relação ao valor observado, enquanto para o modelo  $(0,1,1)(0,1,1)12$  o erro foi de 5,5%. Em termos acumulados, nos seis passos da

previsão entre fevereiro e julho de 2021, o modelo  $(1,1,1)(1,1,2)_{12}$  resultou em um erro acumulado de -9,1%, já para o modelo  $(0,1,1)(0,1,1)_{12}$  esse percentual ficou em -3,6%.

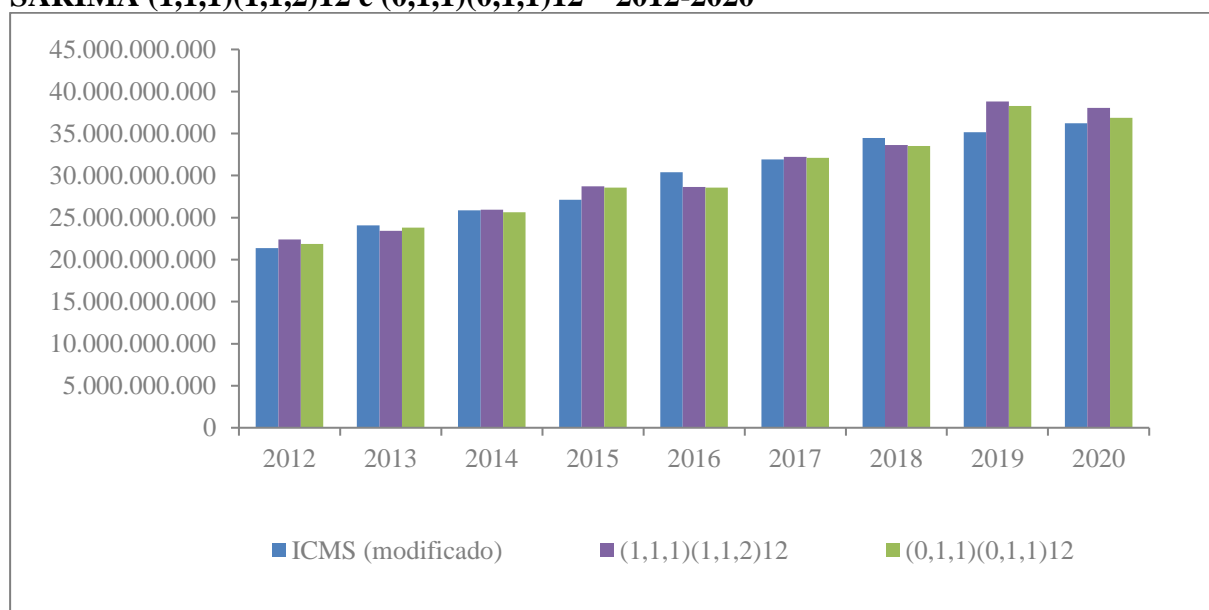
**Figura 3 - ICMS modificado e previsão fora da amostra em seis passos – SARIMA  $(1,1,1)(1,1,2)_{12}$  e  $(0,1,1)(0,1,1)_{12}$  - janeiro de 2020 a julho de 2021**



Fonte: elaborado pela autora.

Por fim, o último exercício realizado foi o somatório das previsões em 12 passos fora da amostra para cada ano do intervalo 2012-2020, sempre com dados até dezembro do ano anterior. Por exemplo, para o ano de 2012 foram consideradas as previsões com base em uma amostra com 60 observações, de janeiro de 2007 a dezembro de 2011. O resultado das previsões com base nos dois modelos SARIMA, para o período 2012-2020, está exibido na figura a seguir.

**Figura 4 - ICMS modificado e somatório da previsão fora da amostra em 12 passos – SARIMA (1,1,1)(1,1,2)12 e (0,1,1)(0,1,1)12 – 2012-2020**



Fonte: elaborado pela autora.

Com base nesse exercício, é possível verificar que os modelos ora subestimam, ora superestimam o ICMS, uma constatação importante para a gestão fiscal. Os anos em que a arrecadação ocorreu acima das expectativas foram: 2013, 2014, 2016 e 2018. Consequentemente, 2012, 2015, 2017, 2019 e 2020 foram anos em que as previsões ficaram acima do realizado. Ao comparar os dois modelos, a configuração (1,1,1)(1,1,2)12 foi melhor nos anos 2014, 2016 e 2018, enquanto nos demais, o modelo (0,1,1)(0,1,1)12 registrou o melhor desempenho.

#### 4. Considerações finais

Este artigo aplicou a modelagem SARIMA para a previsão da arrecadação de ICMS, no Estado do Rio Grande do Sul. Foram estimados dois modelos SARIMA, quais sejam, (1,1,1)(1,1,2)12 e (0,1,1)(0,1,1)12. Inicialmente, foram analisados os resultados da previsão dos dois modelos em duas subamostras, com previsões em 36 passos e em seis passos, respectivamente.

Os dois modelos apresentaram um desempenho razoável, com estimativas dentro do intervalo de confiança. Na primeira subamostra, com dados até julho de 2018, os erros dos dois modelos ficaram muito próximos. Na segunda subamostra com dados até janeiro de 2021, a configuração (1,1,1)(1,1,2)12 gerou previsões com valores inferiores em relação ao



(0,1,1)(0,1,1)<sup>12</sup>. O primeiro foi melhor na previsão em um passo a frente, enquanto o segundo registrou menor erro acumulado em seis passos.

O último exercício consistiu em realizar previsões em 12 passos para cada ano do intervalo 2012-2020, com cada previsão anual baseada em uma subamostra contendo dados até dezembro do ano antecedente. Verificou-se que os modelos subestimaram a arrecadação em alguns anos e superestimaram em outros, o que indica a ausência de viés, do ponto de vista estatístico, e se coloca como um aspecto relevante, sob o prisma da gestão fiscal. Os dois modelos se saíram bem, mas, na maior parte do período, o modelo (0,1,1)(0,1,1)<sup>12</sup> registrou maior acerto, exceto nos anos de 2014, 2016 e 2018.

Como desenvolvimentos futuros, espera-se continuar monitorando e aprimorando os métodos de projeção da principal fonte de receita do Estado e testar outras combinações possíveis de parâmetros, dado que as possibilidades são infinitas. Portanto, espera-se que os modelos propostos não esgotem a discussão, mas estimulem outros trabalhos a propor melhorias nos métodos de previsão do ICMS, no Estado, de modo a tornar a gestão fiscal mais eficiente, especialmente no que tange à garantia de cumprimento dos mínimos constitucionais.

## 5. Referências Bibliográficas

BOX, G. E. P.; JENKINS, G. M. **Time series analysis: forecasting and control**. San Francisco: Holden-Day, 1976.

DICKEY, D. A.; FULLER, W. A. Distribution of the estimators for autoregressive time series with a unit root, **Journal of the American Statistical Association**, New York, v. 74, n. 366, p. 427–431, jun. 1979.

NELSON, C., R. **Applied time series analysis for managerial forecasting**. Holdey-Day, Inc., 1973.

HOLT, C. C. **Forecasting seasonals and trends by exponentially weighted moving averages**. Pittsburgh: Carnegie Institute of Technology, 1957.

RIO GRANDE DO SUL. Secretaria da Fazenda. **Receita dados**. Porto Alegre: Secretaria da Fazenda do Rio Grande do Sul, 2021a. Disponível em: <<http://receitadados.fazenda.rs.gov.br/>>. Acesso em: 25 ago. 2021.

\_\_\_\_\_. Secretaria da Fazenda. Impactos covid-19, **Boletim Receita Estadual**, Porto Alegre, n. 41, 26 jul. 2021b.

\_\_\_\_\_. Secretaria da Fazenda. **Primeiro bimestre de 2020 mostrou melhora fiscal antes da pandemia**. Porto Alegre: Secretaria da Fazenda do Rio Grande do Sul, 2020c. Disponível

em: <<https://www.fazenda.rs.gov.br/conteudo/12924/primeiro-bimestre-de-2020-mostrou-melhora-fiscal-antes-da-pandemia/termosbusca=antecipa%C3%A7%C3%A3o>>. Acesso em 26 ago. 2021.

SAID, S. E.; DICKEY, D. A. Testing for unit roots in autoregressive moving average models of unknown order, **Biometrika**, Oxford, v. 71, n.3, p. 599–607, dec. 1984.

WINTERS, P. R. Forecasting sales by exponentially weighted moving averages. **Management Science**, Providence, v. 6, n. 3, p. 324-342, apr. 1960.