

Modelo estrutural de previsão para o ICMS no Rio Grande do Sul

Clarissa Black¹

Resumo

Este artigo tem como objetivo aplicar o método econométrico univariado estrutural em espaço de estado para a previsão da arrecadação do Imposto de Circulação de Mercadorias e Serviços (ICMS), no Estado do Rio Grande do Sul. A qualificação dos instrumentos de previsão de receita é importante, por exemplo, para a garantia do cumprimento dos mínimos constitucionais a serem gastos em saúde e educação. A estratégia metodológica consistiu em trabalhar com diferentes subamostras e verificar o erro de previsão n-passos à frente. Foram testados seis modelos com componentes fixos e estocásticos. Dentre eles, os modelos com nível estocástico, inclinação fixa e sazonalidade fixa ou estocástica apresentaram os menores erros de previsão fora da amostra.

Palavras-Chave: ICMS; modelo estrutural de séries temporais em espaço de estado; filtro de Kalman.

Structural forecasting model for ICMS in Rio Grande do Sul

This article aims to apply the structural univariate econometric method in state space to forecast the collection of the Tax on the Circulation of Goods and Services (ICMS) in the State of Rio Grande do Sul. The qualification of revenue forecasting instruments is important, for example, to guarantee compliance with the constitutional minimum to be spent on health and education. The methodological strategy consisted of working with different subsamples and verifying the n-steps ahead forecast error. Six models with fixed and stochastic components were tested. Among them, the models with stochastic level, fixed slope and fixed or stochastic seasonality presented the smallest out-of-sample forecast errors.

Key-words: ICMS; structural state space time series model; Kalman filter.

JEL: C22; C53; H71.

Área 8 – Econometria

¹ Doutora em Economia pela Universidade Federal do Rio Grande do Sul (UFRGS).

1. Introdução

O objetivo deste trabalho é apresentar um modelo estrutural em espaço de estado univariado para a previsão do Imposto de Circulação de Mercadorias e Serviços (ICMS), no Rio Grande do Sul, a principal fonte de receita do Estado. A principal justificativa envolve a qualificação dos instrumentos de previsão da receita estadual para uma gestão do orçamento público mais eficiente e que atenda aos preceitos legais e constitucionais.

Como exemplo, o ano de 2021 representou um período de expressivo crescimento da arrecadação de tributos no Rio Grande do Sul, com destaque para o ICMS. Uma combinação de fatores levou a esse resultado, a citar: (a) o efeito estatístico da reduzida base de comparação de 2020, o pior momento da crise econômica provocada pela pandemia da Covid-19; (b) a recuperação econômica, especialmente no primeiro semestre de 2021; (c) a operação de privatização da Companhia Estadual de Distribuição de Energia Elétrica (CEEE-D), a qual levou à contabilização de recursos extraordinários; (d) à dinâmica inflacionária, uma vez que o ICMS é uma variável nominal, com destaque para o crescimento do preço dos combustíveis, ao longo de 2021.

Nessa conjuntura, uma questão sensível às finanças públicas estaduais envolveu o cumprimento dos gastos mínimos em saúde e educação, fixados pela Constituição Federal, os quais são calculados com base na Receita Líquida de Impostos e Transferências (RLIT)². Com a arrecadação acima das expectativas, a gestão fiscal no Tesouro do Estado se tornou desafiadora para o atingir as regras definidas na Constituição do País. Nesse sentido, a qualificação dos instrumentos de previsão permite uma maior clareza no horizonte de previsibilidade e a minimiza os riscos de que esses percentuais mínimos não sejam alcançados.

Para atender ao objetivo proposto, este artigo dispõe de mais duas seções além desta introdução, das considerações finais, das referências bibliográficas consultadas e de um Anexo. A próxima seção apresenta a metodologia, com base em Harvey (1989). Na sequência, a aplicação empírica em diferentes subamostras e a análise dos resultados de previsão. Vale notar que este é o primeiro trabalho, considerando o alcance do conhecimento sobre o tema, a aplicar a método estrutural em espaço-estado para a previsão do ICMS, no Rio Grande do Sul, o que caracteriza uma contribuição inovadora à literatura.

Em resumo, foram testados seis modelos com componentes fixos e estocásticos. As configurações que apresentaram os menores erros de previsão, acumulados em seis passos, em uma subamostra de dados até janeiro de 2021, consideram o nível estocástico, a inclinação fixa e a sazonalidade fixa ou estocástica, nessa ordem.

Outro teste envolveu realizar a previsão em 12 meses para os últimos cinco anos. Os dois modelos citados subestimaram a arrecadação em 2016 e 2018; praticamente acertaram em 2017 (erro de 0,1%) e superestimaram o ICMS de 2019 e 2020. Vale notar que, esse padrão das previsões, em alguns momentos acima e em outros abaixo do realizado, indica a ausência de viés, do ponto de vista estatístico.

No último exercício, em uma subamostra de dados até julho de 2018, o primeiro modelo com o nível estocástico, a inclinação fixa e a sazonalidade fixa apresentou um bom desempenho, em geral, na previsão em 36 passos à frente. Os maiores erros ocorreram em 2020, durante os meses da crise sem precedentes provocada pela pandemia da Covid-19.

² 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 da União, deduzidas as transferências aos municípios e as restituições aos contribuintes.

2. Metodologia

A metodologia adotada neste trabalho consiste na escolha da melhor configuração entre seis modelos univariados estruturais em espaço-estado (HARVEY, 1989) na previsão fora da amostra para a arrecadação do ICMS, no Estado do Rio Grande do Sul. A fonte para os dados consultados é a Receita Estadual do Rio Grande do Sul, que disponibiliza séries mensais do ICMS a partir de janeiro de 2007. Neste artigo, serão considerados os dados atualizados até o mês de julho de 2021. A estratégia metodológica consiste em trabalhar com subamostras e verificar o erro de previsão.

Em termos históricos, os primeiros procedimentos *ad hoc* de previsão univariada surgiram com o modelo conhecido como *exponentially weighted moving average* (EWMA) de Holt (1957) e Winters (1960). A ideia básica nesses modelos era decompor a série em componentes não observáveis, como o nível, a inclinação e a sazonalidade. No entanto, como não havia diferenciação entre sinal e ruído nesse método, uma base estatística sólida está ausente. Em outras palavras, não é possível fazer inferência, realizar o teste de hipótese e estabelecer um intervalo de confiança.

Diante dessa lacuna, ganhou proeminência os modelos ARIMA de Box e Jenkins (1976), os quais passaram a separar o resíduo da série estimada, portanto, com maior solidez estatística. Porém, explicar o comportamento de uma série em termos autorregressivos e de médias móveis torna a narrativa mais abstrata e menos intuitiva. Além disso, a necessidade de diferenciar a série para torná-la estacionária, implica em perdas dos componentes dinâmicos da série.

Por sua vez, os modelos estruturais resgataram a ideia inicial dos modelos de suavização exponencial, juntamente com um referencial estatístico tão robusto quanto a modelagem ARIMA, mas com vantagens ao considerar os parâmetros variáveis no tempo e prescindir da discussão sobre estacionariedade e raiz unitária. O método também é mais sofisticado, haja vista a incorporação do filtro de Kalman (1960) à estatística, um algoritmo cuja aplicação original envolvia a engenharia de controle (HARVEY, 1989).

A justificativa para a utilização de modelos univariados tem relação com a dificuldade em encontrar as variáveis explicativas adequadas em um modelo multivariado de previsão, independentemente de elas mesmas serem previsões ou variáveis antecedentes. Essa é uma tarefa trabalhosa e que exige muito estudo e conhecimento acerca do objeto. Adicionalmente, “[...] a badly misspecified model with explanatory variables may lead to worse forecast than a naive time series model. [...] [A] multivariate model is more vulnerable to misspecification than a univariate one” (HARVEY, p. 14-15, 1989).

Outra questão que envolve a seleção de variáveis em modelos multivariados tem relação com um dilema entre fazer previsão dentro ou fora da amostra. Em outras palavras, se o objetivo é explicar o comportamento da série, a melhor opção é pela inclusão de todas as variáveis relevantes, do ponto de vista teórico, e com parâmetros significativos, do ponto de vista estatístico, incluindo quebras estruturais e *dummies*. Por outro lado, se o intuito for o de fazer projeções, então o modelo mais parcimonioso e simples tende a ter um erro de previsão menor. Já para o desenho de cenários, a modelagem univariada é menos recomendada, enquanto a abordagem multivariada se torna mais interessante.

Diante disso, e no que tange aos modelos estruturais, de acordo com Harvey (1989), qualquer série temporal pode ser decomposta em componentes não observáveis diretamente, como a tendência (nível e inclinação), o ciclo, a sazonalidade e o termo irregular. Vale lembrar que esse tipo de modelo permite que os coeficientes variem no tempo (e não apenas as variáveis), uma vez que as séries mudam suas características com o passar dos anos. Desta forma, os parâmetros podem ser estocásticos ou fixos.

Dentre os possíveis modelos estruturais propostos por Harvey (1989), foram testadas seis configurações, quais sejam:

- a) Nível estocástico + inclinação fixa + sazonalidade fixa;
- b) Nível estocástico + inclinação fixa + sazonalidade estocástica;
- c) Nível fixo + inclinação estocástica + sazonalidade estocástica;
- d) Nível estocástico + sazonalidade fixa;
- e) Nível estocástico + sazonalidade estocástica;
- f) Nível fixo + inclinação estocástica + sazonalidade fixa;

Dois são as justificativas para a essa escolha. A primeira tem relação com a remoção do componente cíclico, haja vista a possibilidade de complicações computacionais que decorrem da simultaneidade de dois componentes com variação periódica. Em segundo lugar, deve-se levar em conta o critério da parcimônia, quando o objetivo é fazer projeções. Por esta razão, optou-se por escolher apenas um dos componentes, ou nível ou a inclinação, para ser estocástico, em cada um dos três modelos.

Conforme Harvey (1989) e Commandeur e Koopman (2007), as equações do modelo estrutural em espaço de estado univariado de tendência (nível e inclinação) e sazonalidade por *dummies* são dadas por:

$$y_t = \mu_t + \gamma_t + \varepsilon_t \quad \varepsilon_t \sim \text{NID}(0, \sigma_\varepsilon^2) \quad (1)$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t \quad \eta_t \sim \text{NID}(0, \sigma_\eta^2) \quad (2)$$

$$\beta_t = \beta_{t-1} + \zeta_t \quad \zeta_t \sim \text{NID}(0, \sigma_\zeta^2) \quad (3)$$

$$\gamma_t = 0 - \sum_{j=1}^{s-1} \gamma_{t-j} + \omega_t \quad \omega_t \sim \text{NID}(0, \sigma_\omega^2) \quad (4)$$

Ao analisar os componentes da primeira equação, μ_t é a tendência, γ_t é o componente sazonal e ε_t representa o termo irregular. Na segunda equação, a tendência é dada pela soma do nível μ_{t-1} e da inclinação β_{t-1} , mais o erro η_t . A equação (3) apresenta a inclinação β_t como função de um termo autorregressivo e de um distúrbio ζ_t . Por sua vez, quanto ao componente sazonal dado pela equação (4), γ_t exige (s-1) equações de estado, na qual s representa a frequência sazonal. Se os dados forem mensais, s é igual a 12.

Quanto aos termos ε_t , η_t , ζ_t e ω_t , considera-se que são normais, independentes (não autocorrelacionados serialmente), identicamente distribuídos (com variância homocedástica) e não apresentam correlação entre si. Vale observar que, quando σ_η^2 , σ_ζ^2 e σ_ω^2 são diferentes de zero, o nível, a inclinação e a sazonalidade se tornam estocásticos. Por outro lado, quando são igualados a zero, deixam de variar no tempo, portanto passam a ser fixos.

A metodologia para estimar os parâmetros e os hiperparâmetros (variâncias do termo irregular e dos erros do vetor de estados) consiste na utilização do filtro de Kalman (1960). O filtro é aplicado em dois passos, através de equações de predição e de atualização, com base em todas as observações passadas e presentes. Portanto, esse método consiste em um algoritmo recursivo e inteligente, uma vez que ele aprende com os próprios erros (CUTHBERTSON; HALL; TAYLOR, 1992). Uma terceira etapa é a de suavização. Se o intuito for o de fazer previsão, o ideal é usar a série filtrada. Por outro lado, se o objetivo é estudar as características dinâmicas da série, o mais indicado é usar a série suavizada.

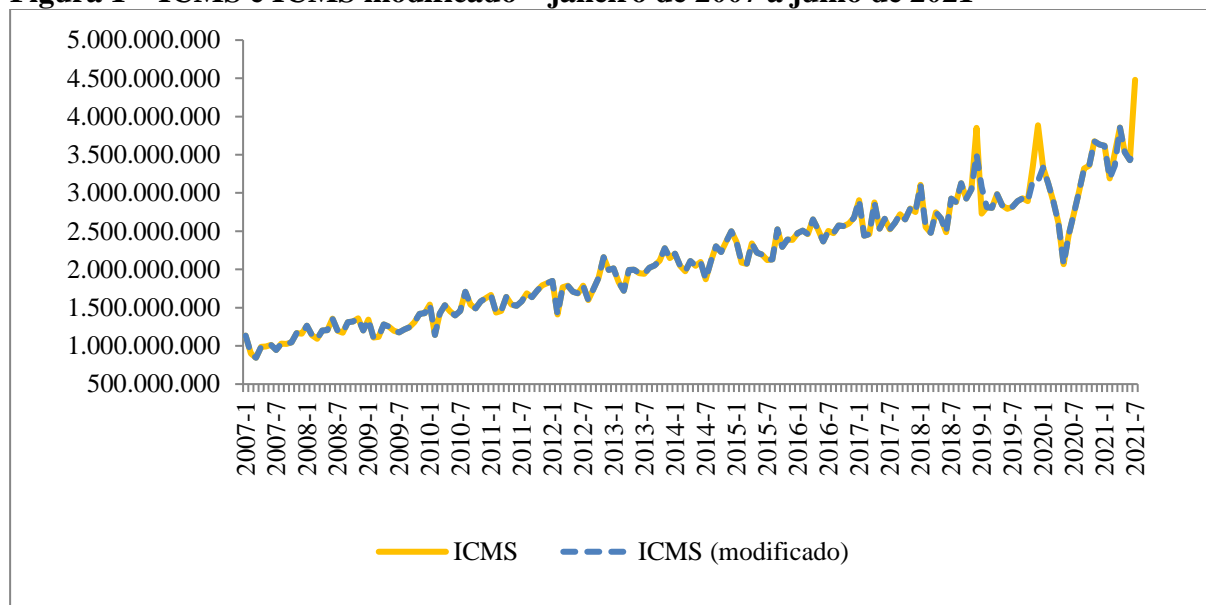
Após discutir a metodologia envolvida nos modelos estruturais em espaço de estado, a seção a seguir apresenta a aplicação empírica e analisa os principais resultados encontrados.

3. Aplicação empírica e análise dos resultados

Antes da estimação dos modelos, o primeiro passo envolveu a realização de alguns ajustes na série do ICMS, disponibilizada pela Receita Estadual, conforme o gráfico abaixo. As modificações referem-se à desconsideração de eventos extraordinários, especialmente a partir do final de 2018.

A justificativa para isso decorre do fato de que incluir *dummies* para esses eventos implica em adição de parâmetros para explicar apenas uma observação. Como o objetivo é reduzir a dimensionalidade de parâmetros em modelos de previsão, o ajuste manual se mostra como a estratégia mais apropriada.

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).

Depois de fazer esses ajustes pontuais na série, o próximo passo consistiu em analisar o erro de previsão em seis passos fora da amostra, ao considerar as estimações com base em uma subamostra de dados entre janeiro de 2007 e janeiro de 2021, para os modelos com:

- a) Nível estocástico + inclinação fixa + sazonalidade fixa;
- b) Nível estocástico + inclinação fixa + sazonalidade estocástica;
- c) Nível fixo + inclinação estocástica + sazonalidade estocástica;
- d) Nível estocástico + sazonalidade fixa;
- e) Nível estocástico + sazonalidade estocástica;
- f) Nível fixo + inclinação estocástica + sazonalidade fixa;

A tabela a seguir reporta as seguintes medidas: erro (valor previsto – valor observado); erro percentual (erro/valor observado); erro acumulado percentual (somatório dos erros/somatório dos valores observados); erro percentual médio absoluto MAPE³ (média do somatório da razão entre a diferença em módulo entre o valor observado e o valor previsto em relação ao valor observado).

Todos os modelos superestimaram os valores previstos no início. Em termos acumulados, o menor erro percentual foi apresentado pelo primeiro modelo, com taxa

³ Mean absolute percentage error (MAPE) é calculado como: $MAPE = \left(\frac{1}{h}\right) \sum_{t=T+1}^{T+h} (|y_t - \hat{y}_t| / y_t)$, com h representando a quantidade de passos na previsão fora da amostra T.

estimada em -1,8%, enquanto a estatística MAPE foi calculada em 3,6%. O segundo modelo com a sazonalidade estocástica ao invés de fixa, também apresentou um bom resultado, com erro acumulado de -2,8% e MAPE de 4,5%. A Tabela a seguir exibe os resultados das demais combinações testadas, em ordem crescente, conforme o erro acumulado, em seis passos.

Tabela 1 – Previsões em seis passos à frente, considerando estimativas calculadas em uma subamostra de dados de janeiro de 2007 a janeiro de 2021

	Previsões	ICMS (modificado)	Erro	Erro (%)	Erro acumulado (%)	MAPE (%)
Nível Estocástico + Inclinação Fixa + Sazonalidade Fixa						
1	3.374.421.022	3.192.427.368	181.993.654	5,7%	5,7%	5,7%
2	3.375.202.327	3.357.156.462	18.045.865	0,5%	3,1%	3,1%
3	3.531.921.267	3.854.845.127	-322.923.860	-8,4%	-1,2%	4,9%
4	3.409.923.260	3.534.051.892	-124.128.632	-3,5%	-1,8%	4,5%
5	3.419.479.937	3.428.958.710	-9.478.773	-0,3%	-1,5%	3,7%
6	3.448.583.579	3.558.349.766	-109.766.187	-3,1%	-1,8%	3,6%
Nível Estocástico + Inclinação Fixa + Sazonalidade Estocástica						
1	3.378.290.881	3.192.427.368	185.863.513	5,8%	5,8%	5,8%
2	3.352.632.406	3.357.156.462	-4.524.056	-0,1%	2,8%	3,0%
3	3.507.363.018	3.854.845.127	-347.482.109	-9,0%	-1,6%	5,0%
4	3.315.283.734	3.534.051.892	-218.768.158	-6,2%	-2,8%	5,3%
5	3.357.040.513	3.428.958.710	-71.918.197	-2,1%	-2,6%	4,7%
6	3.429.827.673	3.558.349.766	-128.522.093	-3,6%	-2,8%	4,5%
Nível Fixo + Inclinação Estocástica + Sazonalidade Estocástica						
1	3.400.238.736	3.192.427.368	207.811.368	6,5%	6,5%	6,5%
2	3.344.888.273	3.357.156.462	-12.268.189	-0,4%	3,0%	3,4%
3	3.437.697.160	3.854.845.127	-417.147.967	-10,8%	-2,1%	5,9%
4	3.183.784.565	3.534.051.892	-350.267.327	-9,9%	-4,1%	6,9%
5	3.342.879.963	3.428.958.710	-86.078.747	-2,5%	-3,8%	6,0%
6	3.545.642.016	3.558.349.766	-12.707.750	-0,4%	-3,2%	5,1%
Nível Estocástico + Sazonalidade Fixa						
1	3.351.811.751	3.192.427.368	159.384.383	5,0%	5,0%	5,0%
2	3.337.795.693	3.357.156.462	-19.360.769	-0,6%	2,1%	2,8%
3	3.479.717.272	3.854.845.127	-375.127.855	-9,7%	-2,3%	5,1%
4	3.342.921.903	3.534.051.892	-191.129.989	-5,4%	-3,1%	5,2%
5	3.337.681.218	3.428.958.710	-91.277.492	-2,7%	-3,0%	4,7%
6	3.351.987.497	3.558.349.766	-206.362.269	-5,8%	-3,5%	4,9%
Nível Estocástico + Sazonalidade Estocástica						
1	3.353.926.664	3.192.427.368	161.499.296	5,1%	5,1%	5,1%
2	3.321.522.321	3.357.156.462	-35.634.141	-1,1%	1,9%	3,1%
3	3.467.919.552	3.854.845.127	-386.925.575	-10,0%	-2,5%	5,4%
4	3.278.744.209	3.534.051.892	-255.307.683	-7,2%	-3,7%	5,8%
5	3.296.774.695	3.428.958.710	-132.184.015	-3,9%	-3,7%	5,4%
6	3.343.125.980	3.558.349.766	-215.223.786	-6,0%	-4,1%	5,5%

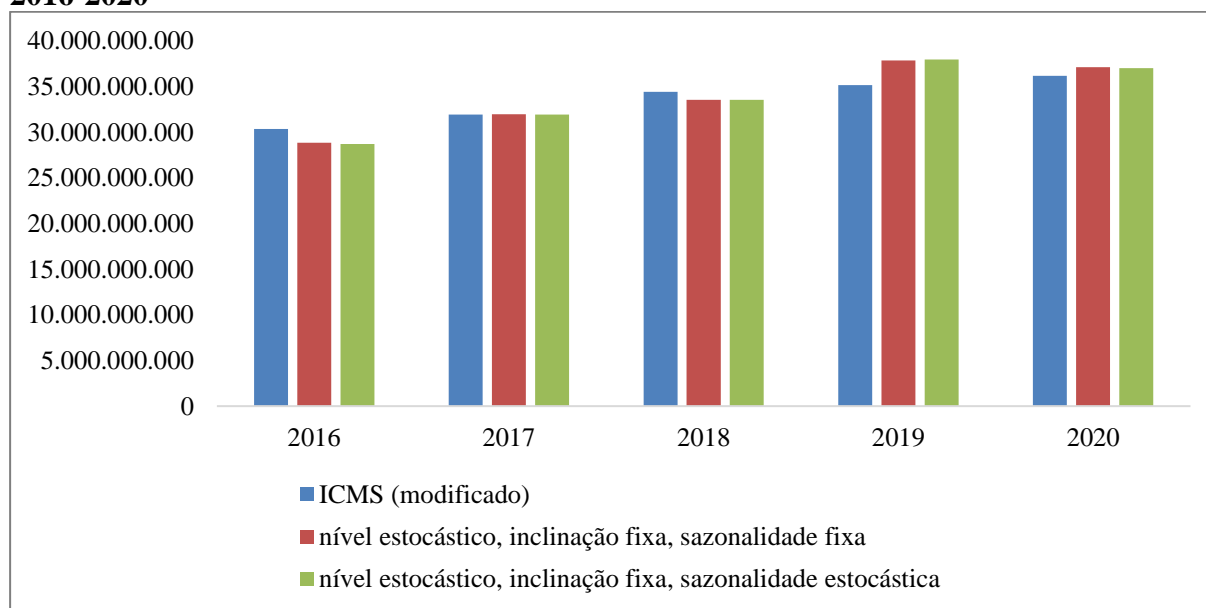
Nível Fixo + Inclinação Estocástica + Sazonalidade Fixa						
1	3.492.238.542	3.192.427.368	299.811.174	9,4%	9,4%	9,4%
2	3.575.202.024	3.357.156.462	218.045.562	6,5%	7,9%	7,9%
3	3.813.610.639	3.854.845.127	-41.234.488	-1,1%	4,6%	5,7%
4	3.772.809.806	3.534.051.892	238.757.914	6,8%	5,1%	5,9%
5	3.863.071.155	3.428.958.710	434.112.445	12,7%	6,6%	7,3%
6	3.972.386.967	3.558.349.766	414.037.201	11,6%	7,5%	8,0%

Fonte: elaborado pela autora.

Após verificar que o modelo de nível estocástico, inclinação fixa e sazonalidade fixa e o modelo com mesma configuração, mas com sazonalidade estocástica, apresentaram os menores erros acumulados, foi realizado um segundo exercício. O teste consistiu no somatório das previsões em 12 passos fora da amostra para os últimos cinco anos, sempre com dados até dezembro do ano anterior.

Dessa forma, para o ano de 2016 foram consideradas as previsões com base em uma amostra de dados até dezembro de 2015. Para o ano seguinte, as estimativas foram realizadas com base em dados até dezembro de 2016 e, assim por diante. As previsões com base nos dois modelos com o nível estocástico, a inclinação fixa e a sazonalidade fixa ou estocástica pode ser observada no gráfico a seguir:

Figura 2 - ICMS modificado e somatório da previsão fora da amostra em 12 passos – 2016-2020



Fonte: elaborado pela autora.

A partir do gráfico exibido é possível observar que os modelos subestimaram a arrecadação em 2016 e 2018; praticamente acertaram em 2017 (erro de 0,1%) e superestimaram o ICMS de 2019 e 2020. As estimativas ficaram muito próximas, mas a configuração que considera a sazonalidade fixa registrou desvios ligeiramente menores, na maior parte do período, exceto em 2020.

No terceiro exercício foi utilizada outra subamostra com dados até julho de 2018, para realizar a previsão em 36 passos para o modelo com o nível estocástico, a inclinação fixa e a sazonalidade fixa. O modelo foi estimado pelo método de máxima verossimilhança, com o

algoritmo de escore exato. O estado estacionário foi encontrado, com convergência incompleta. A análise do vetor de estados em julho de 2018 indica que os parâmetros são significativos a 1% (*p-valores* próximo à zero). Os valores estimados para o primeiro modelo estão dispostos na Tabela 3, em anexo, ao final do artigo.

Quanto à análise de diagnóstico, os resíduos do modelo não apresentam correlação serial, conforme o teste de Box-Ljung, que tem como hipótese nula a ausência de autocorreção. Os *p-valores* calculados foram de 0,1910 (em quatro *lags*) e de 0,8016 (em 36 *lags*). A suposição de normalidade dos resíduos não é rejeitada ao nível de significância de 1% e de 5%, conforme o teste de Bowman-Shenton, cuja estatística foi calculada em 5,9366 e o *p-valor* foi de 0,0514. Por sua vez, o coeficiente de determinação R-quadrado foi estimado em 0,97464.

Com base nesse modelo, foi realizada a previsão fora da subamostra e as estimativas foram comparados aos valores do ICMS modificado (conforme o Gráfico 1) realizados. A Tabela 2 também reporta as mesmas estatísticas de erro presentes na Tabela 1.

Tabela 2 – Previsões em trinta e seis passos à frente, considerando estimativas calculadas em uma subamostra de dados de janeiro de 2007 a julho de 2018

		Previsões	ICMS (modificado)	Erro	Erro %	Erro acumulado %	MAPE (%)
1	ago/18	2.754.848.790	2.880.273.792	-125.425.002	-4,4%	-4,4%	4,4%
2	set/18	2.838.600.853	3.126.414.425	-287.813.572	-9,2%	-6,9%	6,8%
3	out/18	2.834.863.350	2.926.666.495	-91.803.145	-3,1%	-5,7%	5,6%
4	nov/18	2.951.377.185	3.047.325.169	-95.947.984	-3,1%	-5,0%	5,0%
5	dez/18	2.930.770.065	3.503.477.023	-572.706.958	-16,3%	-7,6%	7,2%
6	jan/19	3.027.275.662	3.076.899.082	-49.623.420	-1,6%	-6,6%	6,3%
7	fev/19	2.755.210.880	2.814.176.075	-58.965.195	-2,1%	-6,0%	5,7%
8	mar/19	2.776.636.084	2.803.894.230	-27.258.146	-1,0%	-5,4%	5,1%
9	abr/19	2.967.629.966	2.987.288.337	-19.658.371	-0,7%	-4,9%	4,6%
10	mai/19	2.882.528.133	2.839.760.343	42.767.790	1,5%	-4,3%	4,3%
11	jun/19	2.865.423.026	2.791.424.001	73.999.025	2,7%	-3,7%	4,2%
12	jul/19	2.874.213.732	2.818.447.445	55.766.287	2,0%	-3,2%	4,0%
13	ago/19	2.913.387.002	2.894.912.023	18.474.979	0,6%	-3,0%	3,7%
14	set/19	2.997.139.065	2.930.610.422	66.528.643	2,3%	-2,6%	3,6%
15	out/19	2.993.401.562	2.895.240.042	98.161.520	3,4%	-2,2%	3,6%
16	nov/19	3.109.915.397	3.160.537.766	-50.622.369	-1,6%	-2,2%	3,5%
17	dez/19	3.089.308.277	3.156.623.194	-67.314.917	-2,1%	-2,2%	3,4%
18	jan/20	3.185.813.874	3.334.674.001	-148.860.127	-4,5%	-2,3%	3,5%
19	fev/20	2.913.749.092	3.123.355.070	-209.605.978	-6,7%	-2,5%	3,6%
20	mar/20	2.935.174.296	2.887.472.732	47.701.564	1,7%	-2,3%	3,5%
21	abr/20	3.126.168.178	2.606.217.214	519.950.964	20,0%	-1,4%	4,3%
22	mai/20	3.041.066.345	2.066.995.110	974.071.235	47,1%	0,1%	6,3%
23	jun/20	3.023.961.238	2.454.386.207	569.575.031	23,2%	1,0%	7,0%
24	jul/20	3.032.751.944	2.729.325.271	303.426.673	11,1%	1,4%	7,2%
25	ago/20	3.071.925.214	3.016.264.056	55.661.158	1,8%	1,4%	7,0%
26	set/20	3.155.677.277	3.319.074.031	-163.396.754	-4,9%	1,1%	6,9%
27	out/20	3.151.939.774	3.358.904.152	-206.964.378	-6,2%	0,8%	6,8%
28	nov/20	3.268.453.609	3.675.444.420	-406.990.811	-11,1%	0,3%	7,0%
29	dez/20	3.247.846.489	3.635.784.668	-387.938.179	-10,7%	-0,2%	7,1%
30	jan/21	3.344.352.086	3.618.514.343	-274.162.257	-7,6%	-0,5%	7,1%
31	fev/21	3.072.287.304	3.192.427.368	-120.140.064	-3,8%	-0,6%	7,0%
32	mar/21	3.093.712.508	3.357.156.462	-263.443.954	-7,8%	-0,8%	7,1%
33	abr/21	3.284.706.390	3.854.845.127	-570.138.737	-14,8%	-1,4%	7,3%
34	mai/21	3.199.604.557	3.534.051.892	-334.447.335	-9,5%	-1,6%	7,4%

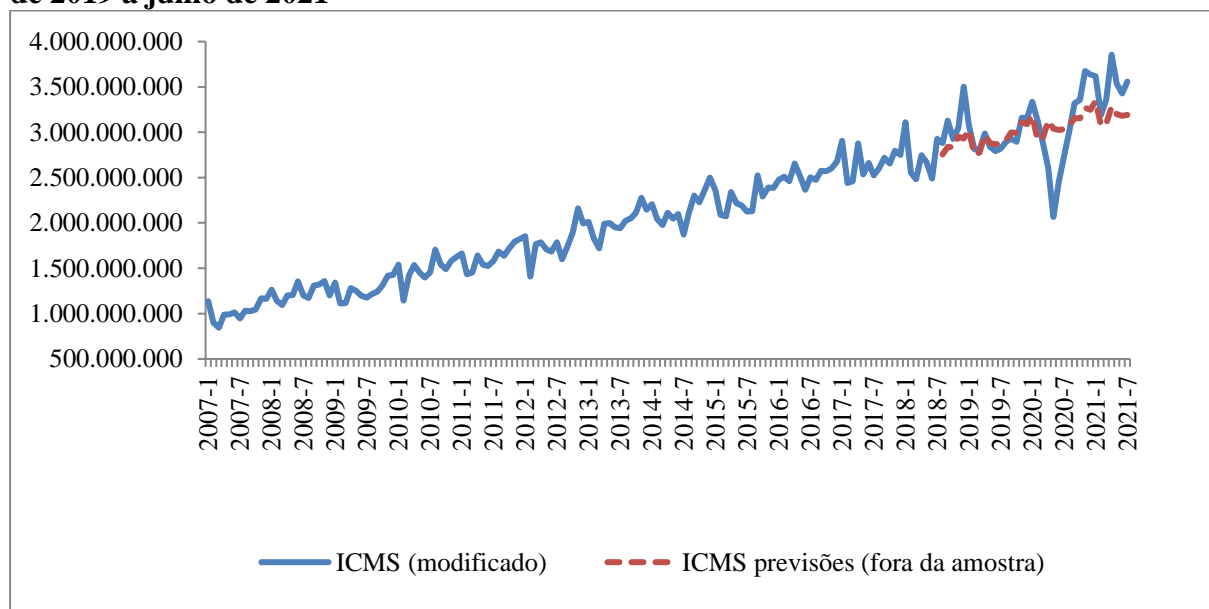
35	jun/21	3.182.499.450	3.428.958.710	-246.459.260	-7,2%	-1,8%	7,3%
36	jul/21	3.191.290.156	3.558.349.766	-367.059.610	-10,3%	-2,1%	7,4%

Fonte: elaborado pela autora.

Com base na tabela anterior, é possível observar que o modelo subestima a arrecadação na maior parte do tempo, exceto a partir do passo 20 até o passo 25, que corresponde ao período mais agudo da crise da covid-19, no qual o modelo passa a superestimar os valores.

As maiores discrepâncias entre os valores previstos e os observados (erro %) ocorreram justamente nesse período, que corresponde ao intervalo que compreende os meses de março a julho de 2020. Mesmo assim, em termos acumulados e com base na estatística MAPE, os erros ficaram inferiores a 7,6% durante todo o período. Ao considerar as estatísticas de erro no período até dezembro de 2019, o período anterior à crise, o erro acumulado é de -2,2%, enquanto a estatística MAPE é de 3,4%. O Gráfico a seguir ilustra os valores observados e aqueles previstos pelo modelo.

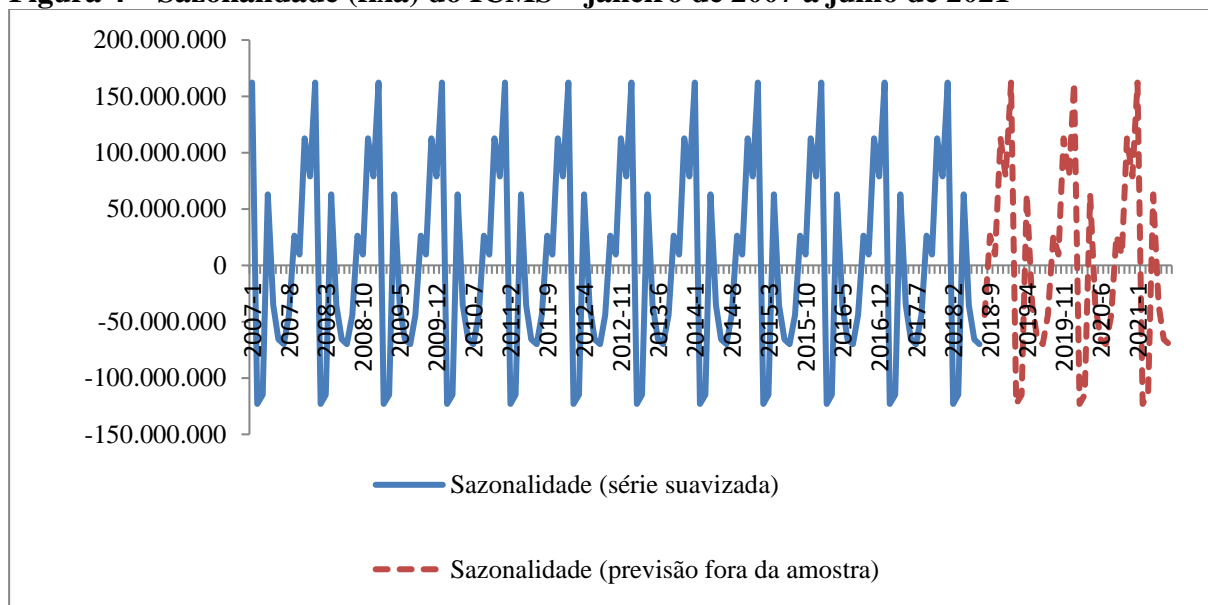
Figura 3 - ICMS modificado e previsão fora da amostra em trinta e seis passos - janeiro de 2019 a julho de 2021



Fonte: elaborado pela autora.

Por sua vez, os dois gráficos a seguir exibem as previsões apresentadas no Gráfico 2 desmembradas em seu componente sazonal e de nível. Vale destacar que a estimativa da sazonalidade considera a série suavizada, enquanto as previsões são obtidas com base na série filtrada. No que concerne à sazonalidade, é possível observar um padrão em formato de “W”, de acordo com o Gráfico 3 abaixo.

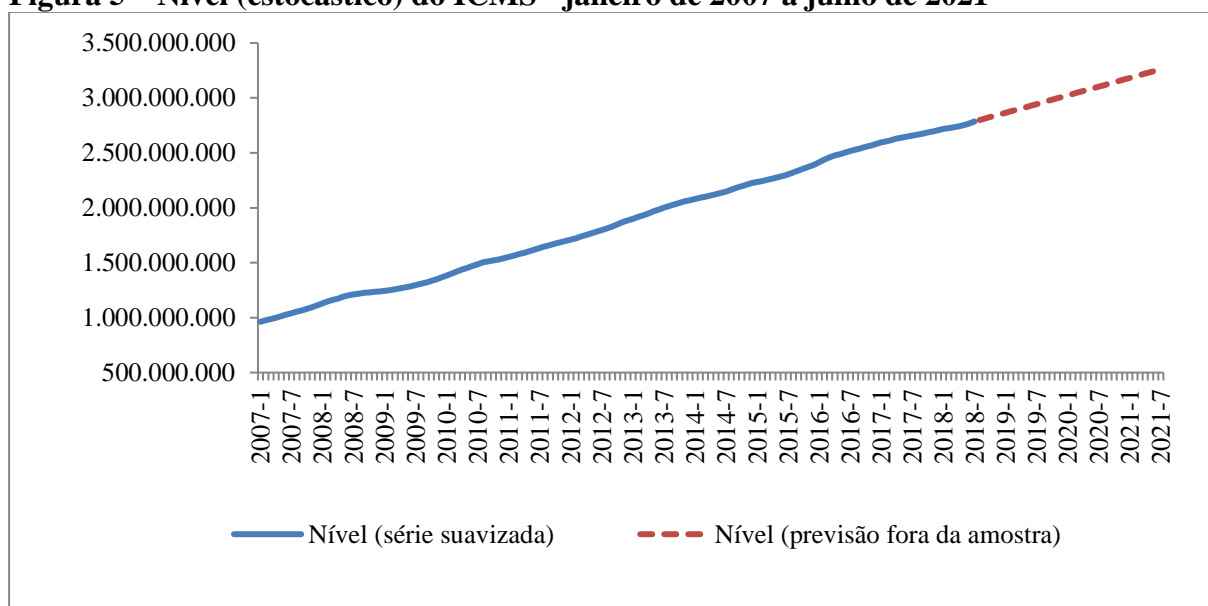
Figura 4 – Sazonalidade (fixa) do ICMS – janeiro de 2007 a julho de 2021



Fonte: elaborado pela autora.

O comportamento sazonal do ICMS indica um movimento negativo nos meses de fevereiro, março, maio, junho, julho e agosto. Os melhores meses são janeiro, novembro e dezembro, nessa ordem. Vale lembrar que, como a arrecadação é baseada nos fatos geradores do mês anterior, a arrecadação elevada de janeiro pode ser relacionada às vendas de final de ano, por hipótese. Por sua vez, os piores meses (março e fevereiro, nessa ordem) para a arrecadação do ICMS podem ser explicados pela menor quantidade de dias úteis e aos períodos de férias. Uma análise mais aprofundada da dinâmica sazonal do ICMS pode ser objeto de estudos futuros.

Figura 5 – Nível (estocástico) do ICMS - janeiro de 2007 a julho de 2021



Fonte: elaborado pela autora.

Já o nível, diferentemente do componente sazonal, é estocástico no modelo estimado. Portanto, ele não é fixo, e apresenta um aumento ao longo do tempo, o que é usual em séries

nominais. Da mesma forma, novamente a estimativa do nível é calculada com base na série suavizada, enquanto as previsões são estimadas com a série filtrada.

4. Considerações finais

Este artigo propôs a utilização do modelo estrutural em espaço de estado para a previsão da arrecadação de ICMS, no Estado do Rio Grande do Sul, em uma contribuição pioneira à literatura. Dentre os seis modelos testados, o de nível estocástico, inclinação fixa e sazonalidade fixa apresentou o menor erro de previsão, em seis passos à frente, ao considerar as estimativas com base em na subamostra de dados de janeiro de 2007 a janeiro de 2021.

Em seis passos, o erro médio acumulado do modelo foi de -1,8%, ao passo que a medida de erro MAPE foi de 3,6%. Vale notar que a configuração com o nível estocástico, a inclinação fixa e a sazonalidade estocástica também apresentou um bom resultado, com percentuais de -2,8% e 4,5%, respectivamente.

Em outro exercício, com os dois modelos supracitados, realizou-se o somatório das previsões em 12 meses para os últimos cinco anos, sempre com dados até dezembro do ano anterior. Os dois modelos subestimaram a arrecadação em 2016 e 2018; acertaram a arrecadação de 2017 (erro de 0,1%); e superestimaram nos dois últimos anos observados. As estimativas ficaram muito próximas, mas a configuração que considera a sazonalidade fixa registrou desvios ligeiramente menores, em quatro dos cinco períodos analisados.

Por fim, ao considerar outra subamostra, com dados até julho de 2018, o modelo de nível estocástico, inclinação fixa e sazonalidade fixa apresentou um bom desempenho na previsão em 36 passos à frente. No entanto, a sua *performance* ficou prejudicada durante a crise da Covid-19. Mesmo assim, em termos acumulados e com base na estatística MAPE, os erros ficaram inferiores a 7,6%. Ao analisar os desvios no período anterior à crise, o erro acumulado ficou em -2,2% e a estatística MAPE em 3,4%.

Espera-se que o trabalho realizado sirva de instrumento para qualificar a gestão dos recursos públicos do Estado do Rio Grande do Sul e estimule outros estudos dentro desta temática e em outras unidades da federação. Nesse sentido é fundamental testar outras configurações de modelos, entre componentes fixos e estocásticos, e acompanhar os desvios de previsão. Assim, com uma melhor previsibilidade das receitas, torna-se mais factível a sua aplicação nas áreas prioritárias e o cumprimento dos percentuais mínimos, definidos no marco constitucional do país.

5. Referências Bibliográficas

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

COMMANDEUR, J. J. F.; KOOPMAN, S. J. **An Introduction to State Space Time Series Analysis**. New York: Oxford University Press, 2007.

CUTHBERTSON, K., HALL, S. G. E.; TAYLOR, M. P. **Applied Econometric Techniques**. London: Harvester Wheatsheaf, 1992.

HARVEY, A. C. **Forecasting, Structural Time Series Models and the Kalman Filter**. Cambridge: Cambridge University Press, 1989.

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

KALMAN, R. E. A new approach to linear filtering and prediction problems. **Journal of Basic Engineering**, New York, v. 82, n. 1, p. 35-45, mar. 1960.

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.

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