

PREVISÃO DA ARRECADAÇÃO DE ICMS PARA O ESTADO DE SERGIPE: MODELAGEM A PARTIR DA METODOLOGIA DE BOX E JENKINS

SERGIPE'S ICMS COLLECTION FORECASTING: AN APPLICATION OF BOX AND JENKINS METHODOLOGY

Wilton Bernardino ¹Tiago de Moura Soeiro ²João Gabriel Nascimento de Araújo ³Rodrigo Vicente dos Prazeres ⁴

Data de recebimento: 14/12/2020

Data de aceite: 22/06/2021

Resumo

O presente trabalho teve por objetivo avaliar e prever o comportamento da arrecadação tributária em Impostos sobre Circulação de Mercadorias e Serviços (ICMS) no Estado de Sergipe, Brasil. Com esse propósito utilizou-se a modelagem proposta por Box e Jenkins (1976), a qual permite a seleção de um modelo de séries temporais considerando dinâmicas autoregressiva e de médias móveis, bem como a incorporação de padrões sazonais ao comportamento da série temporal analisada. A base de dados foco do estudo compreende o período de janeiro de 2005 a agosto de 2017, contemplando informações mensais. Os modelos SARIMA foram estimados no ambiente de programação e análise de dados, R, com o uso dos pacotes o “*forecast*” e “*seasonal*”, que contém algoritmos automatizados úteis à estimação e previsão de series temporais seguindo a proposta de especificação SARIMA. Alternativamente, com base na análise da estrutura de autocorrelação da série e na avaliação da presença do padrão de raiz unitária sazonal, dois outros modelos que consideram a presença de uma raiz unitária sazonal e defasagens em períodos específicas da série foram estimados. O trabalho busca contribuir para o avanço dos estudos sobre o comportamento da arrecadação em ICMS no Brasil, tema que tem sido explorado em alguns trabalhos científicos publicados na década corrente, tendo bastante relevância na área contábil. As análises evidenciaram que o modelo $SARIMA(1,1,1)(0,1,1)_{12}$ mostra-se como o mais adequado para a base de dados

¹ Graduação em Licenciatura em Matemática pela Universidade Federal de Pernambuco, mestrado em Estatística pela Universidade Federal de Pernambuco e doutorado em Economia pela Universidade Federal de Pernambuco. Professor Adjunto IV na Universidade Federal de Pernambuco - Departamento de Ciências Contábeis e Atuariais. E-mail: wiltonrecc@gmail.com

² Graduação em Ciências Contábeis pela Universidade Federal de Pernambuco, graduação em Ciências Contábeis pela Universidade Federal de Pernambuco e doutorado em Ciências Contábeis pela Universidade Federal de Pernambuco. Professor associado da Faculdade de Economia, Administração e Contabilidade da Universidade Federal de Alagoas (FEAC-UFAL). E-mail: mourasoeiro@gmail.com

³ Bacharel, Mestre e Doutorando em Ciências Contábeis pela UFPE. E-mail: j_gabriel90@hotmail.com

⁴ Graduação em Ciências Contábeis pela Universidade Católica de Pernambuco e mestrado e doutorado em Ciências Contábeis pela Universidade Federal de Pernambuco. Professor adjunto da Faculdade de Administração, Economia e Contabilidade da Universidade Federal de Alagoas (FEAC-UFAL). E-mail: rodrigovprazes@gmail.com

considerada. Nesse sentido, em todos os critérios de seleção considerados, o modelo se mostrou superior perante os demais modelos analisados. Modelos com especificação semelhante também foram evidenciados em alguns trabalhos recentes tendo ênfase na modelagem do ICMS no Brasil.

Palavras-chave: Previsão; ICMS; SARIMA.

Abstract

The aim of this paper was to evaluate and predict the behavior of the tax revenues on Circulation of Goods and Services (ICMS) in the Brazilian State of Sergipe. For this purpose, the Box and Jenkins (1976) modeling was used. This method allows the selection of a time series model by considering autoregressive and moving average dynamics, as well as the incorporation of seasonal patterns to the behavior of the analyzed time series. The analyzed database comprises the period from January 2005 to August 2017 in a monthly frequency. The SARIMA models were estimated in the R with the use of the *forecast* and *seasonal* packages. These packages contain automated useful algorithms that can be used in time series estimation and forecasting. Alternatively, based on the autocorrelation analysis and on an unit seasonal root testing, two other models which adjust a unit root seasonal pattern were proposed. The paper seeks to contribute to the advance of studies on the behavior of ICMS tax collection in Brazil, a subject that has been explored in some scientific papers that have been published in the current decade, having a significant relevance in the accounting area. The analyzes showed that the *SARIMA* (1,1,1)(0,1,1)₁₂ model seems to be the most suitable for the considered database. In this sense, in all the considered selection criteria, this model was superior to the other analyzed models. Models with SARIMA similar specifications have been pointed out in some recent papers which focus on the ICMS modeling in Brazil.

Keywords: Forecasting; ICMS; SARIMA.

Introdução

Dentro das atribuições do governo, como gestor dos recursos públicos, projetar a arrecadação de impostos constitui uma tarefa básica no processo de planejamento de orçamentos, possibilitando ao Estado fixar de forma mais adequada o montante a ser gasto com despesas e investimentos, com a finalidade de evitar ou minimizar possíveis déficits nas contas públicas, alocando os recursos de forma mais eficiente.

Na esfera estadual, o Imposto sobre Operações Relativas à Circulação de Mercadorias e de Serviços de Transporte Interestadual e Intermunicipal e de Comunicações (ICMS) responde pela maior parcela da arrecadação no que diz respeito às receitas tributárias Rezende (2009). Diante da representatividade do ICMS nas receitas tributárias estaduais, pesquisas envolvendo modelos de séries temporais têm sido feitas com a finalidade de prever de forma mais acurada a arrecadação do ICMS. Como exemplo, destacam-se os estudos de Arraes e Chumvichitra (1996), Passos e Ramos (2005), Marques e Uchôa (2006), Cruz (2007) e Camargos (2008), Passos, Ramos e Almeida (2011). Essas pesquisas convergem para a superioridade dos modelos de autorregressivos relativamente a outros modelos de previsão de arrecadação do ICMS.

O presente trabalho⁵ tem por objetivo analisar o comportamento da arrecadação de ICMS para o estado de Sergipe e comparar a previsão realizada pelo modelo empírico com a arrecadação real do ICMS. O estudo propõe a utilização dos modelos sazonais autoregressivos integrados de médias móveis (modelos SARIMA), seguindo a metodologia de Box e Jenkins (1976) para uma série temporal com observações mensais da arrecadação do ICMS do estado de Sergipe, considerando o período de janeiro de 2005 a dezembro de 2014. Com esse propósito são geradas previsões para a arrecadação do ICMS para os anos de 2015 e 2016.

⁵ Baseado em texto apresentado no XI Congresso da UFPE de Ciências Contábeis.

Assim, seguindo a linha dos estudos mencionados, através da modelagem de Box e Jenkins (1976), o trabalho objetiva conhecer o comportamento da arrecadação de ICMS para o estado de Sergipe, buscando por um modelo com boa capacidade preditiva na previsão das receitas de ICMS consideradas, as quais estão deflacionadas pelo índice de inflação IGP-DI, a preços de dezembro de 2017. Vale destacar a importância do estudo no que tange a sua contribuição na modelagem de séries de arrecadação em impostos no Brasil, em particular, das receitas de ICMS, que vêm sendo objeto de modelagem estatística em vários estudos quantitativos por vários estados brasileiros.

Além desta introdução, o trabalho está organizado da seguinte forma: a Seção 2 apresenta o referencial teórico; a Seção 3 descreve a metodologia empregada na pesquisa; a Seção 4 discute os resultados e, por fim, a Seção 5 contempla as considerações finais da pesquisa.

Referencial Teórico e métodos de previsão

Para Teti (2009), prever arrecadação é pressupor permitir antecipar resultados futuros, tomando como base dados históricos e modelando-os com o aparato de análises estatísticas. Existe uma grande variedade de modelos de previsão, podendo ser divididas em dois grandes grupos: modelos qualitativos e modelos quantitativos. Os métodos qualitativos baseiam-se na presunção, isso é, em hipóteses ou conjecturas, as chamadas aproximações não-extrapolativas. Esses métodos não apresentam uma formalidade matemática mínima para os resultados encontrados, como aponta Teti (2009).

Dentre os métodos qualitativos mais utilizados seguindo esta proposta estão o de previsão conjectural e o de previsão de consenso. Ambos os métodos consistem em análises baseadas na experiência prática do analista ou grupo responsável pela previsão. Esses modelos são mais indicados em casos com poucos dados históricos e em ambientes de muita instabilidade política, econômica e/ou administrativa, uma vez que os modelos quantitativos não são, por vezes, capazes de captar as principais informações relevantes que afetam a projeção. Entretanto, Siqueira (2002) afirma que há uma série de desvantagens sobre esses modelos, por serem baseados em métodos subjetivos.

Os métodos econométricos, em geral, se utilizam de dados históricos na tentativa de prever observações futuras, com a premissa de que as séries seguirão o mesmo padrão de comportamento registrado no histórico de dados, sendo essa sua principal desvantagem ou limitação. Tais métodos podem ser divididos em modelos casuais e não casuais.

Enquanto os modelos casuais se utilizam de variável dependente e variáveis explicativas (independentes), os não casuais consistem em metodologias que os dados contêm todas as informações necessárias e que, portanto, podem falar por si. De acordo com Siqueira (2002), os métodos quantitativos, em geral, apresentam resultados mais precisos que os qualitativos e que, dentre os métodos quantitativos, os não casuais, em geral, possuem maior acurácia que os casuais.

As previsões também podem ser classificadas em *ex-post* e *ex-ante*. As previsões *ex-post* se utilizam de informações já são realizadas em ajustes de modelos para um período de previsão considerado, enquanto a análise *ex-ante* utiliza variáveis ainda não conhecidas com a mesma proposta preditiva (Teti, 2009).

Método Box-Jenkins em estudos de arrecadação de ICMS no Brasil

O método desenvolvido por Box e Jenkins (1976) permite, num estudo de séries temporais, avaliar em qual modelo a série melhor se ajusta, dentre o processo autoregressivo puro (determinar o valor do parâmetro autoregressivo p), processo de média móvel genuíno (encontrar o valor do parâmetro de média móvel q), processo ARMA (determinando p e q), o modelo integrado ARIMA (encontrando os valores de p , q e a ordem de integração, d) e a modelagem incorporando padrão sazonal *SARIMA* (p , d , q)(P , D , Q), encontrando os parâmetros p , d , q e P , D , Q que descrevem as dinâmicas não sazonal e sazonal da série.

Com foco nos estudos envolvendo a modelagem Box-Jenkins envolvendo modelos sazonais para séries de arrecadação de ICMS pelos estados do Brasil, destacamos o trabalho de Liebel (2004), que propõe um modelo *SARIMA*(1,1,0)(2,1,0)¹², sendo o índice subscrito a ordem da defasagem associada ao termo autoregressivo ajustado para a dinâmica sazonal da série, para dados de arrecadação de ICMS no estado do Paraná, o de Passos et. al (2011) que elege modelo com especificação *SARIMA*(0,1,1)(0,1,1)¹², sendo o índice superescrito a ordem da defasagem associada ao termo de média móvel ajustado para a dinâmica sazonal da série, como o mais adequado para o ajuste da série de ICMS no estado do Pará, o estudo de dos Santos (2006), que ajusta um modelo *SARIMA*(12,1,12)(0,1,1)¹² em modelagem para dados de ICMS no estado de Minas Gerais, além de Scheffer et. al (2014), que ajustam um modelo *SARIMA* (0,1,1)(0,0,1)₁₂ para dados de ICMS no estado do Rio Grande do Sul.

Seguindo a mesma linha de pesquisa, Cruz (2007) utiliza dados de arrecadação de ICMS para o estado do Piauí tendo foco em modelos ARMA e desconsiderando o ajuste sazonal da série. Castanho (2011) modela um processo *ARIMA*(1,0,2) para a série de ICMS do estado do Espírito Santo. Pessoa et. al (2013) propõem os modelos *ARIMA*(1,0,1) e *AFIRMA* (0,0.36,1) (ver Granger e Joyeux, 1980) em análises de ICMS para o estado de Minas Gerais. De Azevedo et. al (2015) avaliaram a utilização da metodologia ARIMA na predição do ICMS nos estados de São Paulo, Minas Gerais, Rio de Janeiro, Rio Grande do Sul, Paraná e Bahia, os quais representam cerca de 70% da arrecadação de ICMS Nacional. Os autores destacam que a utilização da modelagem ARIMA é capaz de prover melhorias na acurácia da previsão da arrecadação de ICMS pelos Estados Brasileiros.

Metodologia

Para a realização deste estudo foram coletados no *site* do Instituto de Pesquisa Econômica Aplicada (www.ipeadata.gov.br) as séries (mensais) de ICMS do estado de Sergipe e do índice geral de preços IGP-DI, no período de janeiro de 2005 (2005.1) a agosto de 2017 (2017.8). No processo de análise da série de ICMS foi selecionado o período de 2005.1 a 2016.8, sendo o período de 2016.9 a 2017.8 a referência para avaliação preditiva dos modelos. Além disso, os valores do ICMS foram deflacionados pelo índice de inflação IGP- DI, sendo considerados os valores de ICMS a preços de agosto de 2017.

Na realização da modelagem SARIMA, plotagem dos gráficos e dos demais testes estatísticos foi utilizado o *software* R versão 3.3.2, com a interface RStudio 0.99.489.0. Os pacotes necessários à análise no R foram: (i) *tseries*, para a reconhecimento/plotagem de séries temporais; (ii) *uroots*, para a realização do teste de Dirkey-Fuller Aumentado, na avaliação da presença de raízes unitárias, (iii) *stats*, para a realização do teste de Ljung-Box (1978), com a finalidade de avaliar a estrutura de autocorrelação residual, (iv) *seasonal* para avaliação do ajuste sazonal da série de ICMS, e (v) *forecast*, para a estimação do modelo SARIMA e realização das análises *ex-post*. Na metodologia de análise temporal proposta por Box e Jenkins (1976), são considerados os passos elencados a seguir:

Identificação: esta etapa consiste em determinar os valores de p , d , q e P , D , Q , os quais descreverão a dinâmica da série em análise. Esses valores são encontrados através de análises na estrutura de autocorrelação (gráfico ACF), do autocorrelograma parcial (PACF), bem como da realização de testes de raiz unitária (sazonal e não sazonal) e análise de resíduos;

Estimação: Após a definição dos valores p , d , q e P , D , Q , estima-se o modelo pelo método da máxima verossimilhança. Este procedimento é feito para um certo número de modelos, elencados como competitivos para a escolha da melhor especificação SARIMA na análise de série temporal;

Diagnóstico: estimados os modelos deve se verificar qual o mais adequado com base no diagnóstico, fundamentado nos critérios de Akaike (1973), AIC e de Schwarz (1978), BIC e em medidas de qualidade de ajuste, como a raiz quadrada do erro quadrático médio, o erro relativo absoluto de previsão, dentre outras, descritas na Seção 3;

Realização das previsões: cumpridas as etapas anteriores, utiliza-se o modelo selecionado para prever os valores futuros da série temporal em uma janela preditiva.

Na seleção do modelo, portanto, deve-se fazer uso do princípio da parcimônia, recomendado por Box-Jenkins (1976), que tem como preceito minimizar a quantidade de parâmetros e defasagens garantindo um o poder preditivo razoável ao modelo escolhido. Outro ponto relevante é realizar uma avaliação criteriosa dos resíduos encontrados, os quais, em uma especificação SARIMA adequada, não devem apresentar estrutura de autocorrelação. As etapas da modelagem realizadas sobre a série de arrecadação de ICMS do estado de Sergipe são descritas na Figura 1, resumida nos itens a seguir: (1) Plotagem dos dados, correção dos valores pelo IGP-DI e identificação de observações discrepantes e padrões de tendência, estrutura autoregressiva e sazonalidade; (2) Transformação dos dados para a escala logarítmica. Utilização de um modelo próprio ou dos algoritmos automáticos implementados nas funções *auto.arima*, do pacote *forecast*, e *seas*, do pacote *seasonal*; (3) Avaliação sobre a estacionariedade e diferenciação da série; (4) Examinar o ACF/PACF, para avaliar o modelo apropriado. (5) Realizar a modelagem para encontrar o melhor modelo; (6) Análise dos resíduos do modelo utilizando o ACF e o teste de Ljung-Box (1976); (7) Realizam-se as previsões. Para avaliar o desempenho do modelo estimado são calculadas quatro medidas de avaliação, também utilizadas em Pessoa, et. al (2013).

i) Erro Absoluto Médio de Previsão (*mean absolute error*)

$$MAE = \sum_{t=T+1}^{T+h} \frac{|Y_t - \hat{Y}_t|}{h}, \quad (1)$$

ii) Erro Absoluto Médio Percentual de Previsão (*mean absolute percent error*)

$$MAPE = 100 \times \sum_{t=T+1}^{T+h} \frac{\left| \frac{Y_t - \hat{Y}_t}{Y_t} \right|}{h}, \quad (2)$$

iii) Erro Absoluto Mediano de Previsão (*median absolute error*)

$$MDAE = \text{mediana}(|Y_{T+1} - \hat{Y}_{T+1}|, \dots, |Y_{T+h} - \hat{Y}_{T+h}|), \quad (3)$$

Figura 1: Esquema da modelagem da série temporal



Fonte: Elaboração própria.

iv) Erro Relativo Absoluto de Previsão (*Relative Absolute error*)

$$RAE = \frac{\sum_{t=T+1}^{T+h} |Y_t - \hat{Y}_t|}{\sum_{t=T+1}^{T+h} |Y_t - \bar{Y}|}, \quad (4)$$

v) Raiz Quadrada do Erro Quadrado Médio de Previsão (*root mean squared error*)

$$RMSE = \sqrt{\frac{\sum_{t=T+1}^{T+h} (Y_t - \hat{Y}_t)^2}{h}}, \quad (5)$$

vi) Coeficiente de Desigualdade de Theil (*Theil's U index of inequality*), Theil (1958)

$$TIDX = \frac{\sqrt{\frac{\sum_{t=T+1}^{T+h} (Y_t - \hat{Y}_t)^2}{h}}}{\sqrt{\frac{\sum_{t=T+1}^{T+h} Y_t^2}{h} + \frac{\sum_{t=T+1}^{T+h} \hat{Y}_t^2}{h}}}, \quad (6)$$

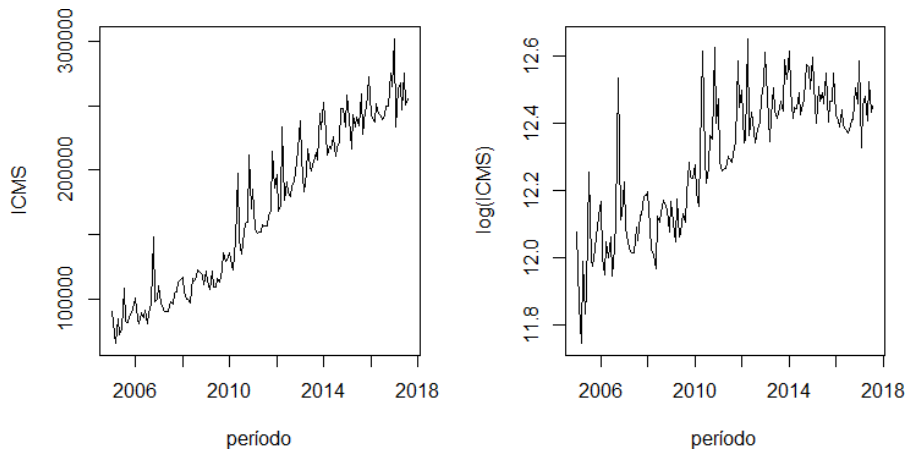
Sendo, respectivamente, Y_t e \hat{Y}_t os valores observados (reais) e previstos (pelo modelo) em um horizonte de previsão de $t = T + 1$ a $T + h$, em que T denota a última unidade de tempo em que os dados são observados para fins de previsão, o termo \bar{Y} denota a média de Y . Assim, o critério de avaliação do desempenho dos modelos estimados baseia-se na busca pelo menor valor dos indicadores descritos nas equações (1) a (6).

Apresentação dos resultados - análise da série de ICMS para o estado de Sergipe

A Figura 2 apresenta a série de ICMS do estado de Sergipe, deflacionada, em escala padrão (gráfico à esquerda) e logarítmica (gráfico à direita). O período gráfico compreende os meses de janeiro de 2005 a dezembro de 2018. Por uma simples inspeção gráfica é possível perceber uma tendência ascendente bem como picos regulares que sinalizam um possível padrão de sazonalidade. Também é possível destacar a forte flutuação ao longo da evolução da série, indicando volatilidade elevada. A modelagem de série temporal via estratégia Box e Jenkins (1976) é bastante indicada em séries com as características mencionadas.

Por conseguinte, com a finalidade de avaliar o modelo mais apropriado, foram plotados os gráficos ACF e PACF da série em escala logarítmica, mostrados na Figura 3. Como se observa nos gráficos da Figura 3, com base no ACF da série percebe-se um comportamento descendente e o PACF apresenta comportamento de decaimento oscilatório. Esses padrões sugerem a inclusão dos parâmetros p (autoregressivo) e q (média-móvel) em uma modelagem SARIMA. Além disso, o decaimento lento no ACF sugere que a série não seja estacionária. O resultado da realização do teste Dickey-Fuller Aumentado apresentado na Tabela 1 reforça a hipótese, sugerindo não rejeição da hipótese nula (presença de raiz unitária) e, assim, a necessidade de diferenciação da série.

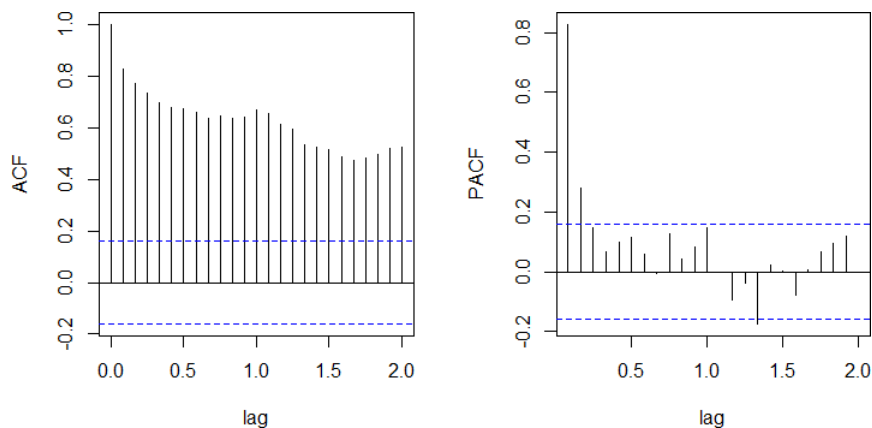
Figura 2: Séries do ICMS do estado de Sergipe em escala padrão (à esquerda) e logarítmica (à direita), no período de janeiro de 2005 a agosto de 2017.



Fonte: Dados da Pesquisa.

Ao diferenciar a série, a mesma comporta-se similarmente ao padrão gráfico de uma série estacionária, conforme mostra a Figura 4. Diante disso, foram plotados o ACF e PACF para a série diferenciada. Os gráficos são mostrados na Figura 5.

Figura 3: ACF (gráfico à esquerda) e PACF (gráfico à direita) da série de ICMS em escala logarítmica.

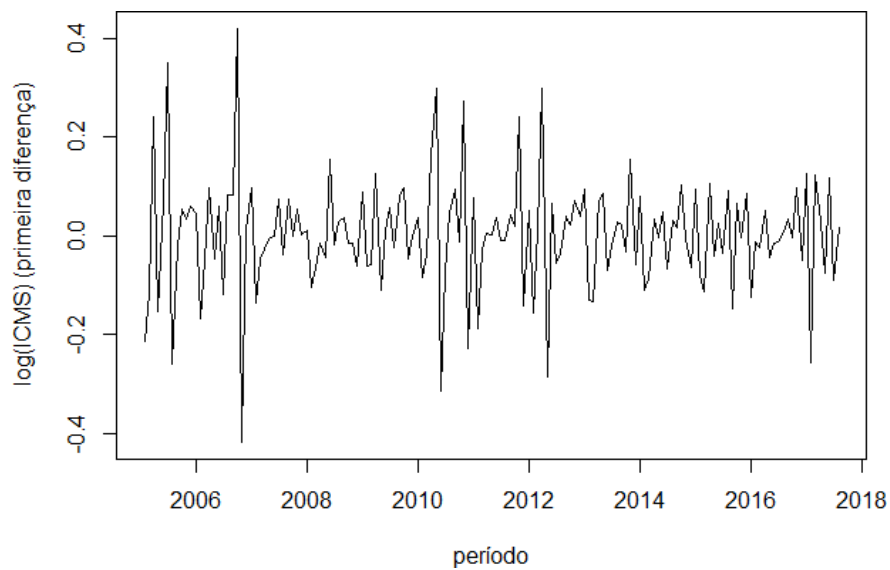


Fonte: Dados da Pesquisa.

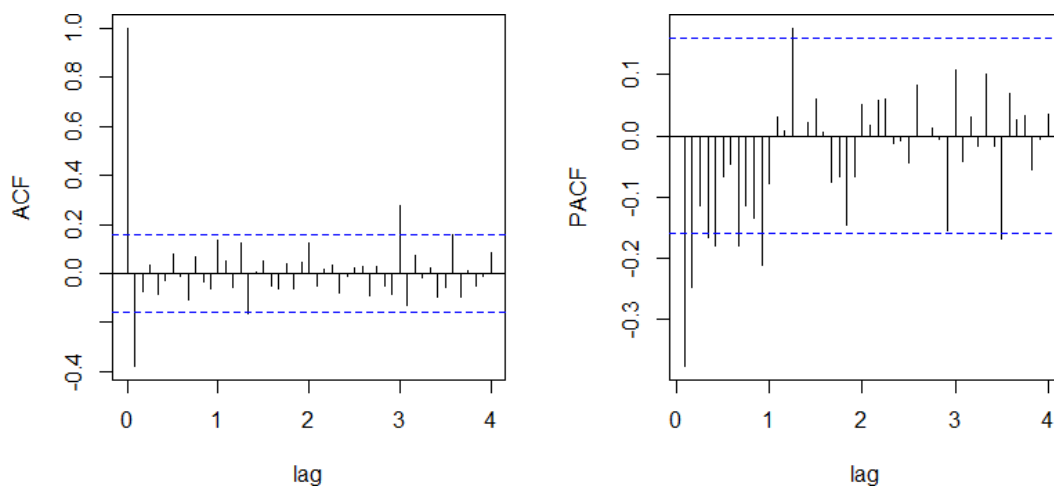
Tabela 1: Teste de raiz unitária da série ICMS em escala logarítmica.

Estatística Calculada	Valor Crítico a 1%
0.50	-2.58

Fonte: Dados da Pesquisa

Figura 4: Série logarítmica do ICMS em primeira diferença.

Fonte: Dados da Pesquisa

Figura 5: ACF (à esquerda) e PACF (à direita) da série de (log) ICMS em primeira diferença.

Fonte: Dados da Pesquisa

Como é possível observar, os gráficos apresentados na Figura 5 evidenciam um decaimento rápido no ACF e um comportamento de decaimento oscilatório no PACF. Adicionalmente, o teste ADF, exposto na Tabela 2, sugere que a série, em primeira diferença, apresenta padrão de estacionariedade.

Tabela 2: Teste de raiz unitária e auto-correlação da série ICMS transformada.

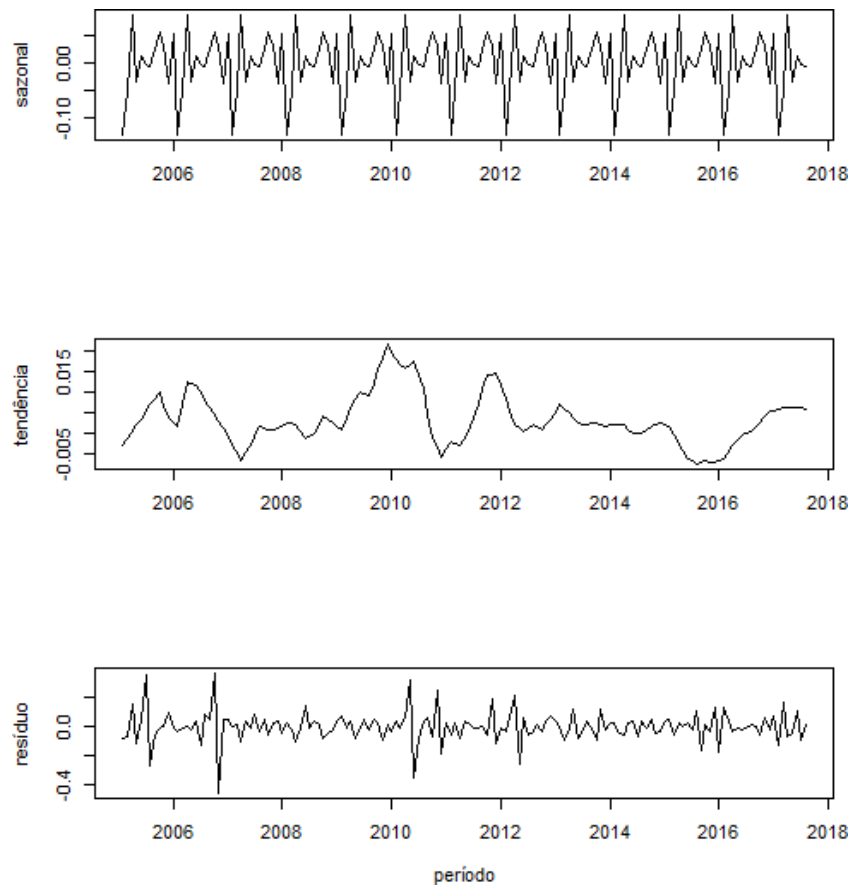
Estatística Calculada	Valor Crítico a 1%
-13.48	-2.58

Fonte: Dados da Pesquisa

A decomposição da série de ICMS (em primeira diferença e em escala log) nos fatores de resíduo, tendência e sazonalidade (Figura 6) sinaliza um padrão sazonal ao longo do período de análise. Além disso,

o fator de tendência comporta-se em concordância com um padrão de série estacionária. Com base nessas características apresentadas, justifica-se a aplicação da modelagem SARIMA.

Figura 6: Decomposição da Série (logarítmica) de ICMS nos fatores de residual (gráfico abaixo), tendência (gráfico do meio) e sazonalidade (gráfico acima).



Fonte: Dados da Pesquisa.

Elegendo o modelo SARIMA

A seleção de modelos ocorreu no ambiente de programação R, inicialmente com auxílio da função *auto.arima* do pacote *forecast*⁶. Utilizando o critério de Akaike (1973), critério AIC, a rotina *auto.arima* elegeu o modelo *SARIMA* (1,1,1)(2,0,0)¹ (modelo 1), em que o índice subscrito indica a ordem de defasagem relacionada à dinâmica sazonal da parte autoregressiva. A utilização da função *seas*, do pacote *seasonal*⁷ sugeriu o segundo modelo (modelo 2), tendo especificação *SARIMA* (0,1,1)(0,1,1)₁₂, em que o índice superescrito indica a ordem de defasagem relativa à dinâmica sazonal da parcela de médias móveis. O modelo 2 é conhecido por *modelo Airlines* (Box e Jenkins, 1986). Com base nos modelos 1 e 2, observando os correlogramas da Figura 5, podemos sugerir um terceiro modelo (modelo 3), propondo modelar a dinâmica sazonal na parcela relativa às médias móveis (ao invés da parte autoregressiva), utilizando o modelo 1 como base. O PACF, apresentando picos nos valores das autocorrelações parciais em períodos anuais (no eixo *x*, valores 1, 2 e 3), sugere a ordem 12 para a defasagem, valor também sugerido no modelo 2. Assim, o modelo 3 é proposto com a especificação *SARIMA* (1,1,1)(0,0,1)₁₂. O quarto modelo foi escolhido com base na realização do teste de raiz unitária proposto por Hylleberg et. al (1990), que não rejeita a hipótese de presença de raiz unitária sazonal a 1% de significância. O teste HEGY (Hylleberg-Engle-Granger-Yoo) foi realizado considerando como termos determinísticos uma constante (intercepto) e dummies sazonais. A estatística “*t*₂” calculada (ver Hylleberg et. al (1990)) apresenta o valor de -2.948, sendo o p-valor associado de 0.0316, sugerindo a presença de raiz unitária sazonal, com frequência anual, a 1% de significância.

⁶ disponível no endereço eletrônico <https://cran.r-project.org/web/packages/forecast/forecast.pdf>

⁷ ver <https://cran.r-project.org/web/packages/seasonal/seasonal.pdf>

Com isso, o modelo 4 é proposto com especificação $SARIMA(1,1,1)(0,1,1)_{12}$. A Tabela 3 mostra os p-valores do teste de Ljung-Box (1978), considerando os ajustes dos modelos 1 a 4 e um número de defasagens (*lags*) igual a 12, 24 e 36. Como é possível observar, a hipótese nula de resíduos não correlacionados é rejeitada para o ajuste do modelo 2, a 5% de significância, em todos as defasagens consideradas. Ao mesmo nível de significância, o ajuste do modelo 1 conduz à rejeição da hipótese nula apenas quando são considerados 36 defasagens. Nesse ponto, os modelos 3 e 4 se destacam, visto que sinalizam ausência de autocorrelação serial em todos as defasagens analisados. Os pvalores associados ao modelo 4 apresentam-se superiores em todos os casos.

Tabela 3: Pvalores do teste de Ljung-Box (1978) considerando o número de *lags* de 12, 24 e 36 para modelos 1 a 4

lags	pvalores		
	12	24	36
Modelo1	0.42 1	0.111	0.036
Modelo2	0.02 8	0.013	0.006
Modelo3	0.99 6	0.791	0.700
Modelo 4	0.98 1	0.988	0.932

Fonte: Dados da Pesquisa.

A Tabela 4 apresenta um resumo das medidas de qualidade de ajuste, calculadas após estimação dos modelos 1 a 4, considerando uma janela de treinamento compreendendo os períodos de 2005.1 a 2016.8 (janeiro de 2005 a agosto de 2016). Nesse sentido, foram realizadas previsões doze passos à frente, sendo comparados os valores preditos pelos modelos com os valores observados na série logarítmica de ICMS para os meses de 2016.9 a 2017.8.

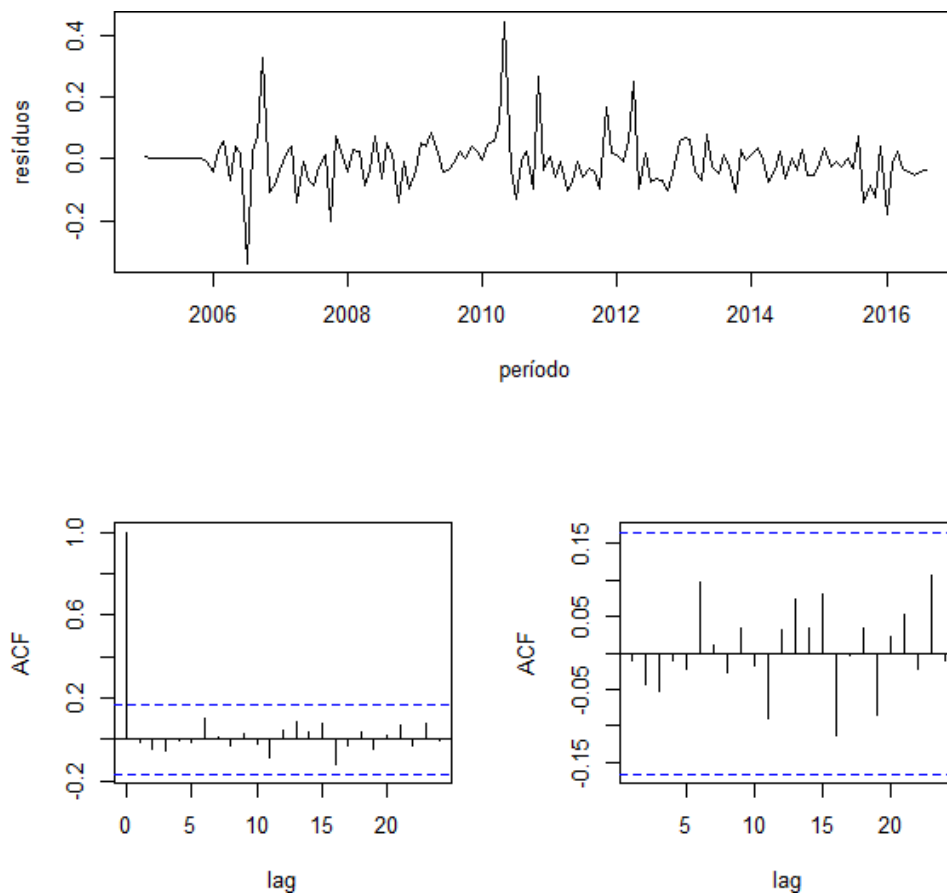
Tabela 4: Comparativo entre os modelos

Modelo 1	Modelo 2	Modelo 3	Modelo 4
Especificação (SARIMA)	(1,1,1)(2,0,0) ₁₂	(0,1,1)(0,1,1) ₁₂	(0,1,1)(0,1,1) ₁₂
	(1, 1, 1)(0, 1, 1) ₁₂		
AIC	-235.34	-224.70	-244,32
BIC	-220.67	-215.60	-232.58
MAE	0.0582	0.0812	0.0712
MAPE	0.0046	0.0065	0.0057
MDAE	0.0438	0.0698	0.0622
RAE	1.218	1.698	1.489
RMSE	0.918	0.097	0.087
TIDX	0.0061	0.0077	0.0070

Fonte: Dados da Pesquisa

Os resultados da análise comparativa entre os quatro modelos sugerem a especificação $SARIMA(1,1,1)(0,1,1)_{12}$ como a melhor escolha. O modelo 4 apresenta os menores valores para medidas AIB, BIC, bem como para o MAE, MAPE, MDAE, RAE, RMSE e o coeficiente TIDX. Dessa forma, o modelo 4 foi o escolhido. A Figura 7 mostra o gráfico dos resíduos (gráfico superior), bem como do ACF e PACF associados (gráficos inferiores), considerando o ajuste $SARIMA(1,1,1)(0,1,1)_{12}$ (modelo 4). Uma análise gráfica residual evidencia um ajuste adequado, em favor da hipótese de que os resíduos sejam bem-comportados, apresentando-se como ruídos brancos (*white noise*).

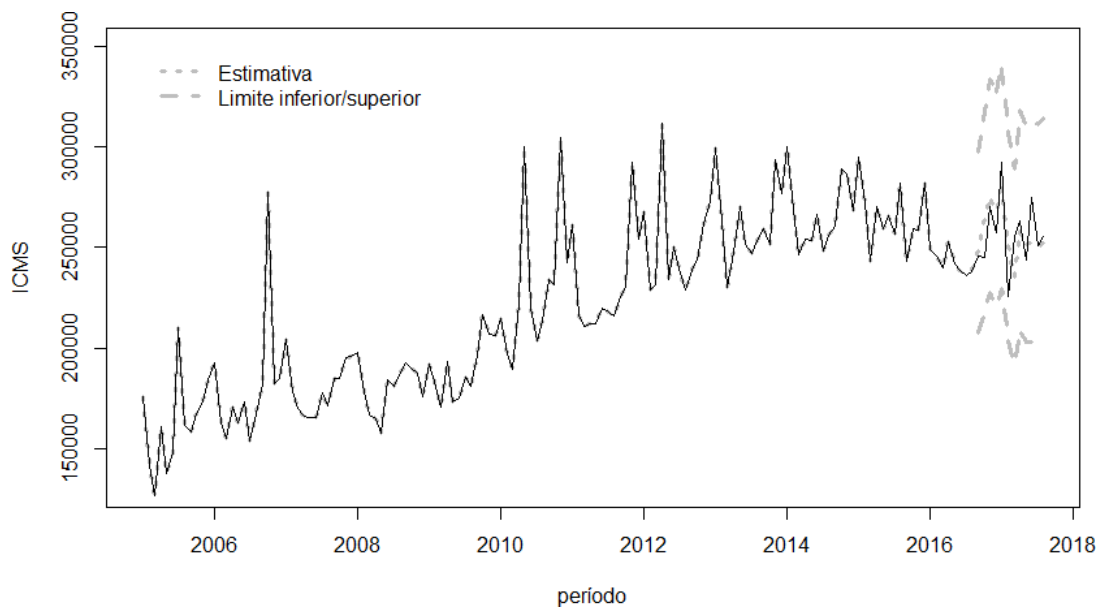
Figura 7: Resíduos dos modelos 4 (gráfico superior), ACF (gráfico inferior à esquerda) e PACF (gráfico inferior à esquerda) dos resíduos para o ajuste do modelo 4.



Fonte: Dados da Pesquisa

Previsão

Realizada a análise descrita na Seção 3, procederam-se com as previsões conforme podem ser vistas no gráfico apresentado na Figura 8, o qual mostra a série de ICMS completa e as estimativas obtidas pelo modelo 4 com as margens de confiança obtidas ao grau de cobertura de 95%. A Tabela 5 contém os valores estimados, observados e as estimativas intervalares obtidas ao grau de cobertura de 95%. Conforme observado, as previsões fornecidas pelo modelo *SARIMA* (1,1,1)(0,1,1)₁₂ mostram-se bastante satisfatórias. Vale destacar que a modelagem *SARIMA* obtida na presente análise é compatível com os estudos de Liebel (2004), Passos et. al (2011), dos Santos (2006) e Scheffer et. al (2014), os quais concluem sobre a especificação *SARIMA* considerando a presença de uma raiz unitária sazonal.

Figura 9: Previsão do ICMS pelos modelos 1 (esquerda) e 2 (direita). Unidade monetária em mil reais

Fonte: Dados da Pesquisa.

Tabela 5: Resultados da estimação considerando o modelo $SARIMA(1,1,1)(0,1,1)_{12}$
Unidade monetária em mil Reais

Mês	Predito	Observado	Limite inferior	Limite superior
2016.9	248341.1	245728.4	207509.9	297206.5
2016.10	263925.0	245065.1	218883.0	318235.7
2016.11	275751.3	270165.2	227765.3	333847.1
2016.12	267827.2	257394.9	220403.0	325455.6
2017.1	279602.4	292228.4	229287.9	340957.9
2017.2	250891.1	225732.3	205018.8	307027.1
2017.3	235516.4	255656.6	191789.3	289213.0
2017.4	258327.2	263177.9	209649.2	318307.7
2017.5	251309.4	244103.4	203270.9	310700.6
2017.6	252095.6	274896.4	203234.7	312703.3
2017.7	250100.0	251010.7	200971.4	311238.3
2017.8	252011.7	255166.0	201859.8	314623.8

Fonte: Dados da Pesquisa.

Considerações finais

Este trabalho teve por objetivo realizar uma avaliação preditiva para a série de arrecadação do Imposto sobre Circulação de Mercadorias e Serviços (ICMS) do Estado de Sergipe, tendo por base dados (mensais) do período de janeiro de 2005 a agosto de 2017. A metodologia de pesquisa consistiu em estimar modelos SARIMA com o uso de algoritmos automatizados, disponíveis nos pacotes “forecast”, que contém o algoritmo “auto.arima”, e o “seasonal”, que contém o algoritmo “seas”, ambas implementadas no ambiente de programação e análise de dados, R. Além disso, dois outros modelos que consideram a presença de raiz unitária sazonal foram analisados. A identificação de uma raiz unitária sazonal foi feita por meio do teste propostos HEGY (Hylleberg et. al, 1990).

A partir dos procedimentos metodológicos conclui-se que a especificação $SARIMA(1,1,1)(0,1,1)_{12}$ mostra-se como a mais adequada ao ajuste da base de arrecadação considerada. Vale destacar que conclusões semelhantes foram observadas em dados de arrecadação de ICMS de outros Estados brasileiros, como pode ser visto em Liebel (2004), Passos et. al (2011) e Scheffe et. al (2014).

Isso reforça a relevância do trabalho no sentido de contribuir com avanços no entendimento das aplicações da modelagem Box e Jenkins (1976) em dados de arrecadação de ICMS no Brasil. Nesse sentido, a pesquisa também contribui com sinalização de metodologias úteis à melhoria da qualidade das previsões de receitas de ICMS no Brasil.

Referências

- ARRAES, R. D. A., & CHUMVICHITRA, P.. **Modelos autoregressivos e poder de previsão: uma aplicação com o ICMS. Texto para discussão**, (152). 1996.
- AKAIKE, H. **Maximum likelihood identification of Gaussian autoregressive moving average models. Biometrika**, 60(2), 255-265. 1973.
- BOX, G. E., JENKINS, G. M., REINSEL, G. C., & LJUNG, G. M.. **Time series analysis: forecasting and control**. John Wiley & Sons. 2015.
- CAMARGOS, A. A. B. **Modelos de previsão de arrecadação tributária do estado de São Paulo: ICMS, IPVA, ICMD e Taxas**. 170 f. *Dissertação (Mestrado Profissional em Economia) - IBMEC*, São Paulo. 2008.
- CASTANHO, B. J. D. S. **Modelos para previsão de receitas tributárias: o ICMS do estado do Espírito Santo**. 2011.
- CRUZ, C. C.S.. **Análise de séries temporais para uma previsão mensal do ICMS: O caso do Piauí. Dissertação (Mestrado em economia)**. Universidade Federal do Ceará, Fortaleza, 2007.
- DE AZEVEDO, R. R., DA SILVA, J. M., & GATSIOS, R. C. **Comparação de Modelos de Previsão de Série Temporal com Base no ICMS Estadual**. XV Congresso USP de Controladoria e Contabilidade, 2015.
- DOS SANTOS, C. M., & DE LIMA, J. E. **Análise de previsões da arrecadação do ICMS no Estado de Minas Gerais**. *Revista de Economia e Administração*, 5(4). 2006
- GRANGER, C. W., & JOYEUX, R. **An introduction to long-memory time series models and fractional differencing**. *Journal of time series analysis*, 1(1), 15-29. 1980.
- HYLLEBERG, S., ENGLE, R. F., GRANGER, C. W., & YOO, B. S. **Seasonal integration and cointegration. Journal of econometrics**, 44(1-2), 215-238. 1990
- LIEBEL, M. J. **Previsão de receitas tributárias: o caso do ICMS no estado do Paraná**. *Dissertação (Mestrado Profissionalizante em Engenharia) - Escola de Engenharia, Universidade Federal do Rio Grande do Sul, Porto Alegre*. 2004.
- LJUNG, G. M., & BOX, G. E. **On a measure of lack of fit in time series models**. *Biometrika*, 65(2), 297-303. 1978.
- MARQUES, C. A. G.; UCHÔA, C. F. A. **Estimação e previsão do ICMS na Bahia**. *Revista Desenharia*, Salvador, v. 3, n. 5, p. 195-211. 2006
- PASSOS, J. J.; RAMOS, E.M.L.S.; ALMEIDA, S. S. **Utilização de modelos ARIMA para previsão da arrecadação de ICMS do estado do Pará**. *Reunião Regional Da Abe*, v. 37. 2011
- PASSOS, J. J. D., & RAMOS, E. **Modelagem estatística para previsão de arrecadação de ICMS do Estado do Pará**. *XI Reunion de Trabajo em Procesamiento de la Informacion y Control*, 248-253. 2005.
- PESSOA, F. D. M. C., CORONEL, D. A., & DE LIMA, J. E. **Previsão de arrecadação de ICMS para o estado de Minas Gerais: uma comparação entre modelos Arima e Arfima**. *Revista Brasileira de Gestão e Desenvolvimento Regional*, 9(2). 2013.
- REZENDE, F. **A reforma tributária e a Federação**. Rio de Janeiro: Fundação Getúlio Vargas. 2009

SIQUEIRA M. L.. **Modelos de séries temporais para a previsão da arrecadação tributária federal.** Dissertação (Mestrado em Economia). Universidade de Pernambuco, Recife, 2002.

SCHEFFER, D., SOUZA, A. M., & ZANINI, R. R.. **Utilização de modelos ARIMA para previsão da arrecadação de ICMS do Estado do Rio Grande do Sul.** XVII Simpósio de pesquisa Operacional e Logística da Marinha-SPOLM. Anais. São Paulo. 2014

SCHWARZ, G.. **Estimating the dimension of a model.** The annals of statistics, 6(2), 461-464. 1978.

TETI, A. C. C. **Modelo de Previsão da Receita Tributária: O Caso do ICMS no Estado de Pernambuco.** Dissertação (Mestrado em Economia). Universidade Federal de Pernambuco. Recife, 2009.

THEIL, H. **Economic Forecasts and Policy.** Amsterdam: North Holland. 1958.