



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**FACULDADE DE ECONOMIA, ADMINISTRAÇÃO, ATUÁRIA, CONTABILIDADE**  
**E SECRETARIADO EXECUTIVO - FEAAC**  
**CURSO DE CIÊNCIAS ECONÔMICAS**

**SAMUEL SANTOS CAMELO ARRAIS**

**MODELO DE PREVISÃO DO ICMS DO ESTADO DO CEARÁ: APLICAÇÃO DA**  
**METODOLOGIA BOX-JENKINS**

**FORTALEZA**

**2019**

SAMUEL SANTOS CAMELO ARRAIS

MODELO DE PREVISÃO DO ICMS DO ESTADO DO CEARÁ: APLICAÇÃO DA  
METODOLOGIA BOX-JENKINS

Monografia apresentada à Faculdade de Economia, Administração, Atuária, Contabilidade e Secretariado Executivo, da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de Bacharel em Economia.

Orientador: Prof. Dr. Francisco José Sales Rocha

FORTALEZA

2019

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Biblioteca Universitária  
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

---

A1m      ARRAIS, SAMUEL SANTOS CAMELO.  
          MODELO DE PREVISÃO DO ICMS DO ESTADO DO CEARÁ : APLICAÇÃO DA  
METODOLOGIA BOX-JENKINS / SAMUEL SANTOS CAMELO ARRAIS. – 2019.  
          33 f. : il.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Faculdade de  
Economia, Administração, Atuária e Contabilidade, Curso de Ciências Econômicas, Fortaleza, 2019.  
Orientação: Prof. Dr. Francisco José Sales Rocha.

1. BOX-JENKINS. 2. ICMS. 3. ESTADO DO CEARA. 4. PREVISÃO . I. Título.

CDD 330

---

SAMUEL SANTOS CAMELO ARRAIS

MODELO DE PREVISÃO DO ICMS DO ESTADO DO CEARÁ: APLICAÇÃO DA  
METODOLOGIA BOX-JENKINS

Monografia apresentada à Faculdade de Economia, Administração, Atuária, Contabilidade e Secretariado Executivo, da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de Bacharel em Economia.

Aprovada em: \_\_\_/\_\_\_/\_\_\_\_\_.

BANCA EXAMINADORA

---

Prof. Dr. Francisco José Sales Rocha (Orientador)  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. Glauber Marques Nojosa  
Universidade Federal do Ceará (UFC)

---

Heriberto Brito Pereira  
Universidade do Estado do Rio Grande do Norte (UERN)

A Deus.

Aos meus pais e esposa.

## **AGRADECIMENTOS**

Primeiramente, agradeço Deus por ter me dado saúde e forças para superar as dificuldades. Aos meus pais, Francisco Euricelio Camelo e Darlene Maciel Santos Camelo, pela educação e pelo incentivo dado até hoje. Agradeço aos meus filhos, Sarah Arrais Camelo, Alice Arrais Camelo e Isaac Arrais Camelo, por entenderem a ausência do pai por todo esse caminho e agradeço especialmente a minha esposa, Riane Arrais de Souza Catunda Camelo por ter me apoiado desde o início, sempre me dando forças nos momentos mais difíceis e não me deixando desistir no meio do caminho.

Ao Prof. Dr. Francisco José Sales Rocha, pela excelente orientação. Aos professores participantes da banca examinadora Glauber Nojosa e Heriberto Pereira pelo tempo, pelas valiosas colaborações e sugestões. E aos colegas de faculdade que estiveram ao meu lado nesta jornada. Por fim, a Universidade Federal do Ceará.

“Todo alicerce requer esforço e muito trabalho”.

Santos.

## RESUMO

O presente trabalho tem por objetivo analisar os resultados preditivos da principal fonte de receita do Estado do Ceará – o Imposto sobre Comércio e Circulação de Mercadorias e Serviços (ICMS), por meio dos modelos de série temporal da família Box-Jenkins. Os dados extraídos são de variação mensal e correspondem ao período de janeiro de 2008 a maio de 2019, coletados pelo site do Conselho Nacional de Política Fazendária (CONFAZ). A partir da análise dos critérios de informação AIC e SBC, três modelos foram selecionados e ajustados. Os resíduos não apresentaram autocorrelação serial e os valores das métricas de erro foram satisfatórios em todos os modelos. Desta forma, concluiu-se que os resultados indicaram que os modelos da família Box-Jenkins podem ser uma ferramenta de grande utilidade, trazendo, assim, informações relevantes ao planejamento fiscal.

**Palavras-chave:** Box-Jenkins. ICMS. Autocorrelação.



## ABSTRACT

This paper aims to analyze the predictive results of the main source of revenue of the State of Ceará - the Tax on Trade and Circulation of Goods and Services (ICMS), using Box-Jenkins family time series models. The data extracted are of monthly variation and correspond to the period from January 2008 to May 2019, collected by the website of the National Council of Finance Policy (CONFAZ). From the analysis of the AIC and SBC information criteria, three models were selected and adjusted. Residues did not show serial autocorrelation and error metric values were satisfactory in all models. Thus, it was concluded that the results indicated that the Box-Jenkins family models can be a very useful tool, thus bringing relevant information to fiscal planning.

**Keywords:** Box-Jenkins. ICMS. Autocorrelation.

## LISTA DE FIGURAS

Figura 1 – Funções da FAC e FACP de um AR(1).....	21
Figura 2 – Funções da FAC e FACP de um MA(1).....	21

## LISTA DE QUADROS E GRÁFICOS

Quadro 1 – Padrões teóricos das FAC e FACP.....	20
Gráfico 1 – Comportamento mensal da série do ICMS.....	26

## LISTA DE TABELAS

Tabela 1	– Teste de estacionariedade <i>Dickey-Fuller Aumentado(ADF)</i> .....	27
Tabela 2	– Modelos estimados e resultados do AIC e SBC .....	27
Tabela 3	– Resultados dos testes de autocorrelação dos resíduos.....	28
Tabela 4	– Estimativas dos parâmetros e valores-p dos modelos selecionados .....	28
Tabela 5	– Previsão da arrecadação do ICMS do Ceará “Por Fora”.....	29
Tabela 6	– Medidas de estatísticas de erro .....	29

## **LISTA DE ABREVIATURAS E SIGLAS**

ICMS	Índice Nacional de Preços ao Consumidor Amplo
IPCA	Imposto sobre Comércio e Circulação de Mercadorias e Serviços

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>14</b>
<b>2</b>	<b>REFERENCIAL TEÓRICO.....</b>	<b>15</b>
2.1	<i>O ICMS.....</i>	15
2.2	<i>Aplicabilidade da Metodologia Box-Jenkins .....</i>	16
<b>3</b>	<b>METODOLOGIA BOX – JENKINS.....</b>	<b>17</b>
3.1	<i>Considerações iniciais.....</i>	17
3.2	<i>Modelos ARMA e SARIMA.....</i>	18
3.3.1	<i>Etapa de especificação e identificação.....</i>	18
3.3.2	<i>Etapa de estimação.....</i>	22
3.3.3	<i>Etapa de verificação.....</i>	23
3.3.4	<i>Etapa de previsão.....</i>	24
<b>4</b>	<b>RESULTADOS E DISCUSSÃO.....</b>	<b>26</b>
<b>5</b>	<b>CONSIDERAÇÕES FINAIS.....</b>	<b>30</b>
	<b>REFERÊNCIAS .....</b>	<b>31</b>

## 1 INTRODUÇÃO

É de comum conhecimento na comunidade acadêmica que o Estado (na forma de instituição governamental – seja ela municipal, estadual ou federal) é um importante agente econômico. Os meios de intervenção econômica (políticas) podem ser de natureza *cambial, salarial, monetária e fiscal* (FORTUNA, 1992). As políticas cambial, salarial e monetária se concentram na alçada do governo federal; e aos governos estaduais e municipais estendem-se as políticas fiscais.

As políticas fiscais, de maneira muito simplista, resumem-se na configuração de gastos e subsídios do Estado que, por sua vez, estão condicionados ao que o governo arrecada. Portanto, para uma eficiente formulação de uma agenda de política fiscal, devem-se considerar métodos acurados para previsão de receitas. Pois, como sugere a estrutura do modelo macroeconômico desenvolvido por Keynes (1936), economia só alcançará o Pleno Emprego se houver equilíbrio orçamentário, ou seja, despesa igual à receita. Como planejar quanto gastar sem saber o quanto arrecadará?

Neste sentido, o presente trabalho tem por objetivo analisar os resultados preditivos da principal fonte de receita do Estado do Ceará – o Imposto sobre Comércio e Circulação de Mercadorias e Serviços (ICMS), por meio dos modelos de série temporal da família Box-Jenkins.

Fundamenta-se esta análise na amplamente disseminada e empregada metodologia Box-Jenkins (1976) para previsão, a qual identifica e modela o comportamento de variáveis aleatórias através da investigação de suas séries temporais. Para tanto, tem-se como objetivos específicos: i) determinar qual modelo se caracteriza como o mais adequado seguindo a metodologia Box-Jenkins e ii) gerar previsões “por fora” da amostra para a série em estudo, com o intuito de constatar a eficiência preditiva da metodologia proposta.

Os dados utilizados contam com 137 observações mensais dos valores agregados brutos arrecadados de ICMS no estado do Ceará durante o período de janeiro de 2008 a maio de 2019, coletados do Ministério da Economia através do site do Conselho Nacional de Política Fazendária – CONFAZ.

O trabalho está dividido em, além desta introdução, outras quatro seções. A segunda seção, discute-se sobre o referencial teórico. A terceira seção, por sua vez, explicita-se a estrutura básica da metodologia Box-Jenkins. A quarta seção, alude sobre os resultados obtidos. Por fim, na quinta seção, comentam-se as considerações finais.

## 2 REFERENCIAL TEÓRICO

### 2.1 O ICMS

O ICMS foi instituído na Constituição Federal de 1988, e é detalhado na Lei Complementar Nº 87, de 13 de setembro de 1996, também conhecida como a Lei Kandir. Suas características e suas peculiaridades estão todas descritas nessa lei. Seleciona-se para este trabalho apenas alguns de seus atributos mais importantes (BRASIL, 1996).

O Imposto sobre Circulação de Mercadorias e Serviços (ICMS) é um imposto indireto que incide, como sugere o nome, sobre operações com mercadorias e serviços. De competência estadual, é um imposto *não cumulativo*, no sentido de não incidir em mais de uma operação em um mesmo processo. Ou seja, de acordo com o artigo 19, compensa-se o que for devido em cada operação relativa à circulação de mercadorias, prestação de serviços e comunicação com o montante cobrado nas operações anteriores. O imposto permite (de acordo com o artigo 6º, parágrafo primeiro) a diferenciação de alíquotas quanto à essencialidade do bem e quanto à localidade de seu consumidor final – se dentro ou fora do Estado cobrador do imposto.

Intuitivamente, como se trata de um imposto sobre operações com mercadorias, há espaço para conjecturar que os maiores contribuintes são empresas comerciais e industriais. Isto é verificado por Sobieski, Silva e Souza (2015). Entretanto, considerando a proporção e não a ordem de grandeza, a penalidade do imposto recai de modo mais severo sobre os mais pobres, já que à medida que a renda aumenta a alíquota diminui. Por conta disso, Lima (2010) afirma ser um imposto regressivo, de efeito suave para o de renda mais alta e de resposta sensível para o de renda mais baixa.

Contudo, mesmo com sua complexidade e visíveis distorções, o ICMS é uma das fontes mais expressivas de receita para os Estados e Distrito Federal (SOUSA et al., 2019); e isto o configura como o tributo de maior arrecadação no Brasil (REZENDE, 2009). Já no Ceará, a importância do fisco oriundo do ICMS não é somente elevada, mas vital para o estado, pois representa aproximadamente 95% da receita tributária (RIBEIRO, 2010).

Dada essa importância, autores como Guaragna e Mello (2002), Corvalão (2002), Camargos (2008), Leibel (2004, 2005) e outros direcionaram esforços para identificar modelos que melhor se adequem à natureza da variável, para poder prever econometricamente sua arrecadação (CLEMENTE e CLEMENTE, 2011). Esta tarefa “tem tido um crescimento notável tanto no



ambiente acadêmico como a atenção de políticos, institutos de pesquisas e do setor público contribuinte” (SCHEFFER et al., 2014. p. 2).

## 2.2 Aplicabilidade da Metodologia Box-Jenkins

A metodologia Box-Jenkins introduzida ao campo acadêmico por George Box e Gwilym Jenkins em seu livro de 1976 – *Time Series Analysis: Forecasting and control*, grande referência na análise estatística de séries temporais.

O método é usualmente aplicado para delinear processos temporais através de modelos autorregressivos integrados e de médias móveis – ARIMA (SOUSA et al., 2019). Mas, como demonstrado por Chain *et al* (2015), a metodologia não se restringe a somente um tipo de modelo, mas abrange toda uma classe de modelos autorregressivos.

O método também é aplicado em modelos que consideram sazonalidade na série (SARIMA), e em modelos de longa dependência (ARFIMA). Os modelos sazonais caracterizam-se por possuírem correlação em instantes de tempo  $s$ , onde  $s$  é o período sazonal (SCHEFFER; SOUZA; ZANINI, 2014). De outra forma, uma série histórica com sazonalidade se destaca por algo que se repete em um determinado período (SOUZA et al., 2019).

Chain et al. (2015), ao utilizar a metodologia Box-Jenkins para obter a previsão do ICMS do estado de Minas Gerais, constatou que o SARIMAX (1,1,1) (1,1,1)<sub>12</sub> foi o modelo com melhor performance. Concluiu também que não incorporar os valores atípicos nas estimativas pode prejudicar o processo de previsão.

Fontenele (2017) mostrou através da metodologia de modelos autorregressivos integrados de médias móveis sazonal (SARIMA), que para o estado do Ceará - CE o modelo de melhor resposta para auferir valores futuros do ICMS para o período de janeiro à dezembro de 2016, foi o SARIMA (1,1,1) (1,0,1)<sub>12</sub>.

Para Sousa et al. (2019), o modelo preferido para explicar o comportamento futuro do ICMS do estado do Ceará - CE, determinado por meio dos modelos da família Box-Jenkins, foi o SARIMA (1,1,1). O resultado obtido por Sousa et al. (2019), demonstrou que o Ceará obterá um considerável nível de arrecadação para o período de novembro de 2018 a outubro de 2019, e com baixa probabilidade de erro.

Para o estado do Rio Grande do Sul – RS, ao utilizar a metodologia Box-Jenkins verificou-se que o melhor modelo foi o SARIMA (0,1,1) (0,0,1)<sub>12</sub>. O resultado encontrado prever uma arrecadação crescente para o estado (SCHEFFER; SOUZA; ZANINI, 2014).

### 3 METODOLOGIA BOX – JENKINS

#### 3.1 Considerações iniciais

Nesta seção serão exibidas as etapas da metodologia Box-Jenkins. Conforme Rocha (2002) a concepção básica dos modelos da família Box-Jenkins é a de que um processo estocástico pode ser explicado por valores passados da variável dependente e pelo seu termo erro. Enders (2010) ressalta que a metodologia Box-Jenkins impõe como condição necessária a estacionariedade da série, isto é, média e variância constantes.

De acordo com Morretin e Tolo (2006), séries temporais apresentam os seguintes componentes: sazonalidade, tendência, ciclos e ruído aleatório. A sazonalidade é um movimento ondulatório que se repete em um determinado período do tempo; já a tendência é o sentido de deslocamento da série; o ciclo é o movimento de oscilação que tende a ser periódico; por fim, o ruído aleatório ou erro compreende a variabilidade ou movimento que não pode ser identificado.

As etapas da metodologia são as seguintes: especificação, identificação, estimação, verificação e previsão. Como bem ponderam Clemente e Clemente (2011), na primeira etapa, um tipo de modelo econométrico é cogitado, delimitando o universo da análise. Na segunda etapa, identifica-se a correta configuração do modelo, por meio das funções de autocorrelação e autocorrelação parcial (FAC e FACP, respectivamente). A correta execução desta etapa é crucial para o sucesso do processo, pois uma vez que não seja bem identificado, o modelo será reprovado nas próximas etapas. Na terceira etapa, os parâmetros são estimados e, na etapa seguinte, o modelo é testado realizando-se um diagnóstico do modelo ajustado. Por fim, realiza-se a previsão.

O diagnóstico do modelo ajustado consiste na análise do resíduo da regressão estimada para verificar a capacidade de previsão do modelo. Clemente e Clemente (2011) completam que se o modelo for considerado adequado, será usado para previsão. Sendo, entretanto, considerado inadequado, descarta-se o modelo em questão sem antes verificar e comparar a acurácia entre os modelos. Para tanto, admitem-se várias identificações alternativas para realização de testes e comparações entre os modelos.

### 3.2 Modelos ARMA e SARIMA

Um modelo ARMA de ordem  $p, q$  –  $ARMA(p, q)$  – é um modelo que contém componentes autorregressivos e de médias móveis. Considerando uma série temporal  $y_t$ , pode ser representada algebricamente como segue:

$$(1 - \phi_1 B - \dots - \phi_p B^p)y_t = (1 - \theta_1 B - \dots - \theta_q B^q)\varepsilon_t$$

em que  $B$  é o operador de defasagem,  $\phi$  e  $\theta$  são os parâmetros do modelo e  $\varepsilon_t$  o termo de erro. A parte autorregressiva do modelo (AR) é representada pelas variáveis defasadas de  $y_t$ , enquanto que as defasagens do termo de erro pela parte de média móvel (MA). O modelo exposto acima é chamado de ARMA ( $p, q$ ), em que  $p$  é a ordem de defasagem do termo autorregressivo e  $q$  a ordem de defasagem do termo de média móvel (GUJARATI, 2011).

Conforme Rocha (2002) os modelos do tipo SARIMA, em regra, são compostos por uma parte não sazonal e outra sazonal. A definição algébrica de modelo misto sazonal – não sazonal é feita como segue:

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 - \phi_1 B^S - \dots - \phi_{pS} B^{pS})y_t = (1 - \theta_1 B - \dots - \theta_q B^q)(1 - \theta_1 B^S - \dots - \theta_{qS} B^{qS})\varepsilon_t$$

$\phi$  e  $\theta$  correspondem aos parâmetros da parte sazonal do modelo, enquanto que  $S$  é o período sazonal. Os termos  $P$  e  $Q$  pertencem, respectivamente, a ordem de defasagem do componente autorregressivo sazonal (SAR) e do componente de média móvel sazonal (SMA). Dessa forma temos o modelo misto sazonal – não sazonal.

#### 3.3.1 Etapa de especificação e identificação

A etapa de identificação consiste em verificar quais modelos são apropriados, onde se inicia com a preparação dos dados, e analisa-se o gráfico da série em estudo. A inspeção desse gráfico serve para verificar a existência de tendência ou alterações na variância, o que indicaria se a série é ou não estacionária (WERNER; RIBEIRO, 2003).

Uma vez que a série em estudo não seja estacionária, temos de torná-la estacionária antes de executar a metodologia Box-Jenkins (GUJARATI, 2011). Segundo Bueno (2012) é necessário que o processo estocástico seja estacionário para proceder às inferências estatísticas sobre os parâmetros estimados.

Seja o processo estocástico,  $\{Y_t, t \in \mathbb{Z}\}$ ,  $\mathbb{Z} = \{1, \pm 2, \pm 3, \dots\}$ , em que  $t$  são os diferentes períodos da série, diz-se que ele é fracamente estacionário se:

i. Média constante

$$E(Y_t) = \mu, \text{ para todo } t \in \mathbb{Z}$$

ii. Variância constante

$$E(Y_t)^2 < \infty$$

iii. Covariância independente de  $t$

$$E(Y_t - \mu)(Y_{t-j} - \mu) = \gamma_j$$

A primeira condição garante que a média seja a mesma para todo período, ainda que a distribuição da variável aleatória se modifique ao longo do tempo. A segunda condição assegura que a variância seja finita, ainda que distintas em diferentes períodos. A terceira condição implica que a variância seja constante e a autocovariância seja independente do tempo, mas não da distância temporal entre as observações (BUENO, 2012).

Para verificar quanto a estacionariedade da série pode-se recorrer às funções de autocorrelação (FAC) e autocorrelação parcial (FACP). Estas fornecem através da análise de seu correlograma a existência de raiz unitária. Contudo, nem sempre é possível identificar se a série é estacionária apenas pela inspeção do correlograma amostral, principalmente em séries que apresentam sazonalidade (QUEIRÓS, 2012).

Uma forma mais técnica de verificar a existência de raiz unitária é através do teste Dickey-Fuller Aumentado (ADF). O teste ADF é um procedimento estatístico que apresenta em sua hipótese nula a existência de raiz unitária (REIMBOLD et al., 2017).

Nesse caso, o teste de raiz unitária se baseia na regressão do modelo definido como segue:

$$\Delta Y_t = \beta_1 + \beta_2 t + \phi Y_{t-1} + \sum_i^p \alpha_i \Delta Y_{t-i} + \varepsilon_t$$

Trata-se de uma modelo com constante e tendência determinística. O termo de erro  $\varepsilon_t$  apresenta um processo de ruído branco puro e  $\Delta Y_{t-i} = (Y_{t-1} - Y_{t-2})$ ,  $\Delta Y_{t-2} = (Y_{t-2} - Y_{t-3})$ , e assim sucessivamente. O parâmetro  $\alpha_i$  corresponde ao número de coeficientes de  $\Delta Y_{t-i}$  com  $p$  representando as defasagens a serem usadas para que o termo de erro não seja serialmente correlacionado. Define-se  $\beta_1$  como intercepto, também denominado de constante,

$\beta_2$  é o coeficiente de tendência; e  $\phi$  é o coeficiente de presença de raiz unitária (GUJARATI, 2011).

O número de defasagens pode ser determinado conforme o critério de *Schwert* (1989), definida na expressão abaixo. Sendo T o número de dados da série e  $P_{m\acute{a}x}$  a quantidade máxima de defasagens definida pela fórmula.

$$P_{m\acute{a}x} = \left[ 12 \left( \frac{T}{100} \right)^{\frac{1}{4}} \right]$$

Estima-se a equação Dickey e Fuller pelo método de mínimos quadrados e examinam-se os valores críticos da estatística  $t$ , encontradas no teste. Se o resultado do teste ADF, for menor que a estatística tabulada por Dickey e Fuller, então se rejeita a hipótese nula e, portanto, a série é estacionária (BUENO, 2012).

A inspeção visual do correlograma das funções FAC e FACP de uma série estacionária, fornecem informações quanto ao número e o tipo dos parâmetros do modelo (QUEIRÓS, 2012). O procedimento comumente utilizado para identificar os parâmetros dos modelos da família Box-Jenkins é apresentado de forma resumida no quadro 1.

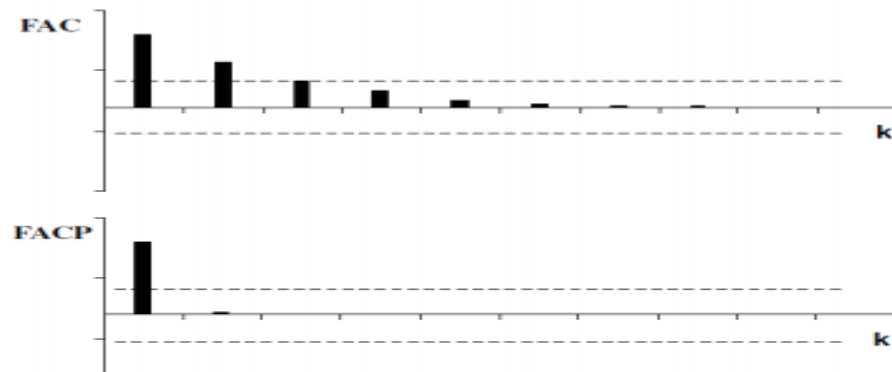
*Quadro 1 Padrões teóricos das FAC e FACP*

Modelo	FAC	FACP
AR(p)	Declina exponencialmente	Truncada na defasagem p
MA(q)	Truncada na defasagem q	Declina exponencialmente
ARMA(p,q)	Declinante	Declinante

*Fonte: Adaptada (BUENO, 2012)*

Para entendermos melhor o padrão teórico da tabela 1, iremos apresentar em seguida dois exemplos gráficos da função de autocorrelação FAC e FACP que tem como resultado os modelos AR(1) e MA(1), respectivamente.

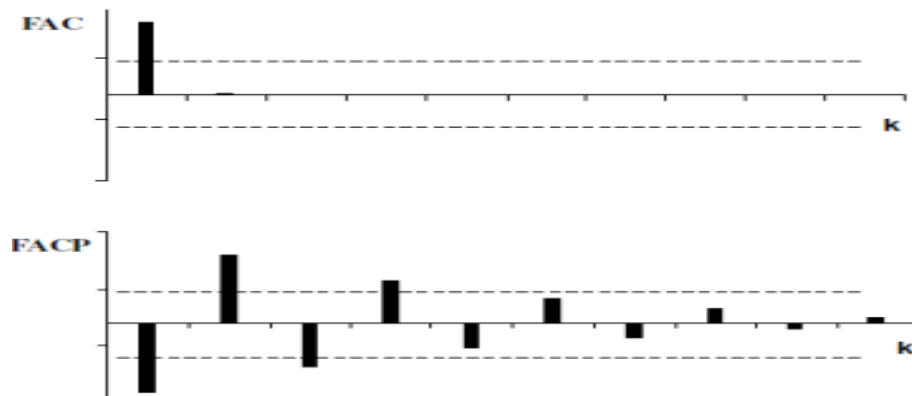
Figura 1: Funções da FAC e FACP de um AR(1)



Fonte: Baltar(2008)

O correlograma da FACP exposto na figura 1, claramente mostra um truncamento na primeira defasagem, ao mesmo tempo em que a FAC apresenta um decaimento exponencial. Este comportamento trata-se de um modelo AR(1).

Figura 2: Funções da FAC e FACP de um MA(1)



Fonte: Baltar (2008)

Ao passo em que o correlograma da FACP decai exponencialmente, corroborando com a ideia de que o componente AR é igual à zero, a FAC exibe um pico em sua primeira defasagem. Deste modo, pode-se constatar pela figura 2 um modelo MA(1).

Vale destacar que em modelos que apresentam sazonalidade, a metodologia Box-Jenkins sugere em geral modelos do tipo SARIMA. Em séries mensais com sazonalidade anual, as funções de autocorrelação FAC e FACP apresentam comumente picos com decaimento lento nas defasagens múltiplas de 12.

Para Morretin e Tolo (2006) a etapa de identificação da metodologia Box-Jenkins consiste em verificar a variância da série, diferenciá-la até se tornar estacionária e através da análise das funções de autocorrelação (FAC) e autocorrelação parcial (FACP) identificar a ordem do modelo. Segundo Rocha (2002, p.3), “a escolha do melhor modelo requer uma grande habilidade e arte, que só se adquire com a prática”.

Apesar dos correlogramas da FAC e FACP servirem como uma ferramenta de grande auxílio na seleção dos modelos, ainda assim, podem surgir dúvidas quanto ao número de parâmetros a serem incluídos na regressão. Desta forma, os critérios de informação *Akaike Information Criterion* (AIC) e *Schwarz Bayesian Criterion* (SBC) são colocados como uma alternativa de dirimir eventuais dúvidas quanto ao número de parâmetros a serem incluídos ao modelo. A principal ideia dos critérios de informação AIC e SBC é minimizar uma função baseada nos resíduos, penalizada pelo número de regressões (BUENO, 2012).

Rocha (2002) ressalta que o aumento de parâmetros no modelo reduz a soma do quadrado dos resíduos, porém torna o modelo menos parcimonioso. Conforme Box-Jenkins (1976) modelos menos parcimoniosos, isto é, sobre parametrizados, produzem resultados preditivos de menor eficiência.

A especificação do critério de informação AIC e SBC podem ser definidas pelas equações que seguem:

$$AIC = T \ln ( \text{soma do quadrado dos resíduos} ) + 2n$$

$$SBC = T \ln ( \text{soma do quadrado dos resíduos} ) + n \ln T$$

Em que  $n$  é o número de parâmetros do modelo; e  $T$  é o número de observações da série de tempo. O segundo termo das equações sobreditadas é o fator que penaliza a inclusão de parâmetros no modelo. A sistemática é que se escolhe o modelo com os menores valores de AIC e SBC (ROCHA, 2002). É com base nas melhores propostas de modelos que se estabelece o segundo passo da metodologia.

### 3.3.2 Etapa de estimação

Uma vez conhecido o modelo, pode-se utilizar o método de máxima verossimilhança ou mínimos quadrados ordinários para estimar seus parâmetros (GUJARATI, 2011). As máximas verossimilhanças condicional e exata, bem como o método de mínimos

quadrados ordinários, possuem as mesmas propriedades assintóticas quando as raízes do polinômio estão fora do círculo unitário (BUENO, 2012).

A escolha dos parâmetros na fase de identificação é de suma importância, pois identificar parâmetros de maneira parcimoniosa garantem melhores resultados na estimação do modelo (QEIRÓS, 2012). Todavia, existem casos atípicos em que coeficientes não significativos individualmente, levam a uma melhor previsão quando adicionados ao modelo, contrariando a metodologia utilizada neste trabalho (ROCHA, 2002).

Depois de estimados os parâmetros do modelo, a próxima fase constitui-se na checagem dos resíduos para determinar quais modelos serão escolhidos para a etapa de previsão.

### 3.3.3 Etapa de verificação

De acordo com Werne e Ribeiro (2013, p. 58) a etapa de verificação é utilizada para avaliar por meio da análise dos resíduos se o modelo estimado não apresenta dependência linear e, assim, os coeficientes de autocorrelação dos resíduos devem ser estatisticamente iguais a zero”.

Inicialmente, pode-se realizar a inspeção visual do correlograma dos resíduos do modelo estimado. Se os valores dos resíduos não extrapolarem os limites de confiança, então, não se rejeita a hipótese nula de que não há autocorrelação. Para uma análise de caráter mais técnico, pode-se utilizar o teste Ljung-Box para verificar a existência de autocorrelação serial dos resíduos (LJUNG-BOX, 1978).

O teste Ljung-Box, consiste em verificar se os resíduos são independentes e identicamente distribuídos. A estatística do teste é representada por:

$$Q(k) = n(n + 2) \sum_{j=1}^k \frac{\hat{\epsilon}_j^2}{n - j}$$

a estatística do teste possui distribuição Qui-Quadrado ( $\chi^2$ ) com  $k - n$  graus de liberdade,  $k$  é o número de defasagens utilizadas,  $n$  é o número de observações da série,  $\hat{\epsilon}$  é o valor estimado dos resíduos. Se a  $Q(k) > \chi_{G.L}^2$  e, neste caso, rejeita-se hipótese nula, ou seja, o resíduo apresenta autocorrelação serial. De modo alternativo ao teste Ljung-Box, pode-se ainda checar a correlação serial dos resíduos, por meio do teste Breush-Godfrey (QEIRÓS, 2012).



### 3.3.4 Etapa de previsão

Pode ser considerada como etapa final da metodologia Box-Jenkins a previsão ou projeção da série em análise. Nesta fase, realiza-se a previsão de até  $h$  passos a frente para os modelos selecionados, conforme os critérios estabelecidos nas etapas anteriores. No entanto, é necessário estar preparado para realizar ajustes ao modelo no caso de novas informações (QUEIRÓS, 2012).

Para se realizar a escolha mais acertada entre os modelos, deve-se realizar um estudo sobre os erros de previsão (BUENO, 2012). Segundo Campos, Clemente e Cordeiro (2006), modelos que apresentam todas as evidências estatísticas que o tornem consistentes, podem não gerar resultados satisfatórios, isto é, apresentam baixo desempenho de previsão. Para constatar se o modelo selecionado apresenta resultados satisfatórios, deve-se estimar o erro de previsão e o modelo que apresentar o menor erro é o escolhido (MARTIN et al., 2016).

Para verificar a acurácia dentre os modelos propostos existem alguns índices de desempenho. Segundo Bueno (2011), três deles são consideradas básicos: o Erro Absoluto Médio (EAM), a Raiz do Erro Quadrado Médio (REQM) e o Erro Absoluto Percentual Médio (EAPM).

$$EAM = \sum_{i=1}^n \left| \frac{y_{aju} - y_{real}}{n} \right| \quad REQM = \sqrt{\sum_{i=1}^n \frac{(y_{aju} - y_{real})^2}{n}} \quad EAPM = \left( \frac{100}{n} \right) \sum_{i=1}^n \left| \frac{y_{aju} - y_{real}}{y_{real}} \right|$$

onde  $n$  indica o número de observações,  $y_{aju}$  correspondem aos valores da série prevista e  $y_{real}$  aos valores da série observada.

Além das medidas supracitadas os critérios de informação AIC e o SBC, bem como o *U de Theil* que calcula a raiz quadrada da média do erro de previsão em termos relativo, podem ser utilizados como medidas de desempenho.

O coeficiente de desigualdade de *Theil* pode ser definido algebricamente de acordo com a seguinte formula:

$$U = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_{aju} - y_{real})^2}}{\sqrt{\frac{1}{n} \sum_{i=1}^n y_{aju}^2} + \sqrt{\frac{1}{n} \sum_{i=1}^n y_{real}^2}}$$

os valores de  $U$  variam entre 0 e 1, sendo que, os resultados que se aproximam de zero indica uma melhor previsão (MARTIN et al., 2016). Por fim, comparam-se os modelos selecionados através das estatísticas de erro, e escolhe-se o modelo com os melhores resultados.

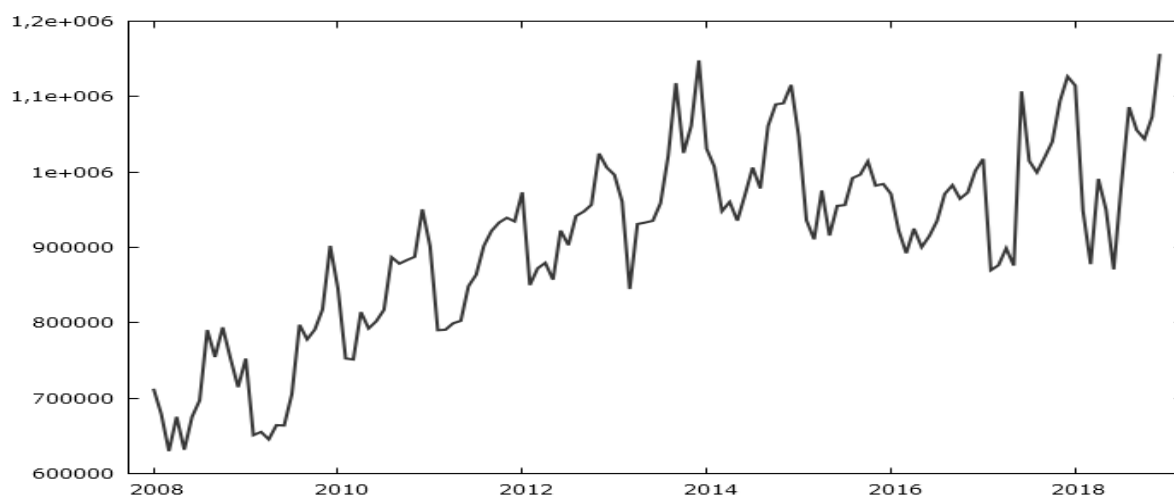
#### 4 RESULTADOS E DISCUSSÃO

A série temporal utilizada no presente trabalho compreende as informações financeiras da arrecadação de ICMS do estado do Ceará, em periodicidade mensal. Os dados utilizados são referentes ao período de janeiro de 2008 até maio de 2019, porém serão reservados os valores referentes ao primeiro semestre de 2019 para verificar a precisão do modelo ajustado.

A série em estudo foi deflacionada pelo Índice Nacional de Preços ao Consumidor Amplo (IPCA), em variação mensal (Base: dez de 2018). Para a análise dos dados foi utilizado o software estatístico e econométrico Gretl, pelo qual se realizou as etapas deste estudo.

Na figura 1 pode-se observar o gráfico da série de arrecadação de ICMS do estado do Ceará. Os dados estão deflacionados e corresponde ao período de janeiro de 2008 até dezembro de 2018. Pelo comportamento dos dados ao longo do tempo, é possível observar a existência de tendência e sazonalidade.

*Gráfico 1 Comportamento mensal da série do ICMS*



*Fonte: Elaborado pelo autor a partir de dados da CONFAZ (2019)*

Para prosseguir as etapas da metodologia Box-Jenkins é necessário garantir que a série em estudo seja estacionária. Para tanto, realizou-se o teste Dickey-Fuller Aumentado (ADF), onde se constatou que a hipótese nula não pôde ser rejeitada a um nível de 5% de confiança.

Para tornar a série estacionária foi necessário aplicar a primeira diferença em termos regular. A tabela 1 exibe os resultados do teste ADF com a série em nível e em primeira diferença.

Tabela 1 Teste de estacionariedade Dickey-Fuller Aumentado (ADF)

TESTE DE ESTACIONARIEDADE (ADF)				
Série			Estatística do teste	P-Valor
LnICMS	NÍVEL	COM CONSTANTE	-2.605495	0.0947
		SEM CONSTANTE	2.546092	0.9974
	1ª DIFERÊNCIA	COM CONSTANTE	-3.375061	0.0138
		SEM CONSTANTE	-2.825930	0.0050

Fonte: Elaborado pelo autor a partir de dados da CONFAZ (2019)

Após tornar a série estacionária, deve-se verificar por meio do correlograma das funções FAC e FACP quais os modelos preliminares. Para tanto, foram selecionados três modelos, cujo resultado dos critérios de informação AIC e SBC apresentaram os menores valores dentre os modelos testados. É possível verificar pela tabela 3, quais os candidatos selecionados, bem como seus valores de AIC e SBC.

Tabela 2 Modelos estimados e resultados do AIC e SBC

Ordem	Modelo	AIC	SBC
1	SARIMA (1,1,2)(1,0,1) <sub>12</sub>	- 383,7305	- 367,1064
2	SARIMA (1,1,2)(1,0,2) <sub>12</sub>	- 382,4891	- 363,0943
3	SARIMA (0,1,1)(1,0,1) <sub>12</sub>	- 374,9427	- 363,8262

Fonte: Elaborado pelo autor a partir de dados da CONFAZ (2019)

\* São considerados modelos concorrentes o SARIMA(1,1,2)(1,0,1) e SARIMA(0,1,1)(1,0,1)

Conforme exposto na tabela 3, dentre os modelos avaliados o que apresentou o menor valor pelo critério de informação AIC e SBA foi o SARIMA (1,1,2)(1,0,1)<sub>12</sub>. Enquanto que, entre os modelos concorrentes o SARIMA (1,1,2)(1,0,2)<sub>12</sub> conteve o menor valor pelo critério AIC e o SARIMA (0,1,1)(1,0,1) obteve o melhor resultado pelo critério SBC.

Depois de selecionados os modelos, deve-se verificar a existência de autocorrelação residual. Para isto, utilizou-se o teste Ljung-Box onde foi constatado que os três modelos selecionados rejeitam a hipótese nula de autocorrelação dos resíduos.

Os resultados que comprovam as condições de que não há autocorrelação a 5% de significância podem ser verificados pela tabela 4, que fornecem as estatísticas do teste, bem como os valores-p.

*Tabela 3 Resultados dos testes de autocorrelação dos resíduos*

<b>Modelo</b>	<b>Ljung-Box Q'</b>	<b>Valor-p</b>
SARIMA (1,1,2)(1,0,1) <sub>12</sub>	11,2182	0,9163
SARIMA (1,1,2)(1,0,2) <sub>12</sub>	12,1844	0,8376
SARIMA (0,1,1)(1,0,1) <sub>12</sub>	13,4206	0,8932

*Fonte: Elaborado pelo autor a partir de dados da CONFAZ (2019)*

Como exposto na tabela 4, os resultados obtidos pelo teste de autocorrelação residual atesta a validação dos três modelos. Dos modelos selecionados, são apresentados de forma resumida na tabela 5, os valores de seus coeficientes estimados e seus respectivos valores-p.

*Tabela 4 Estimativas dos parâmetros e valores-p dos modelos selecionados*

<b>Modelos/ Parâmetros</b>	<b>SARIMA (1,1,2)(1,0,1)<sub>12</sub></b>	<b>Valor-p</b>	<b>SARIMA (1,1,2)(1,0,2)<sub>12</sub></b>	<b>Valor-p</b>	<b>SARIMA (0,1,1)(1,0,1)<sub>12</sub></b>	<b>Valor-p</b>
Phi	-0,663699	7,92e-022	-0,666211	7,61e-021	N/A	N/A
Phi	0,942715	3,31e-131	0,936755	1,00e-116	0,947962	3,74e-135
Theta	0,248712	0,0218	0,256553	0,0263	-0,659489	1,18e-013
Theta	-0,615677	1,63e-011	-0,602382	1,51e-010	-0,669559	5,41e-024
Theta	-0,687477	4,42e-022	-0,754985	3,44e-020	N/A	N/A
Theta	N/A	N/A	0,106622	0,3803	N/A	N/A

*Fonte: Elaborado pelo autor a partir de dados da CONFAZ (2019)*

Pode-se verificar pela tabela 5, que todos os parâmetros exceto o Theta do SARIMA (1,1,2)(1,0,2) não foi estatisticamente significativo ao nível de 5% de confiança. Com isto, o próximo passo é verificar a qualidade preditiva de cada modelo.

Para verificar e validar a acurácia dos modelos selecionados optou-se por realizar uma estimação “por fora”, ou seja, foram reservadas as seis últimas observações da série de arrecadação de ICMS, para realizar testes de desempenho. Os resultados preditivos de cada modelo podem ser verificados pela tabela 6.

Tabela 5 Previsão da arrecadação do ICMS do Ceará “Por Fora”

PERÍODO S	SARIMA (1,1,2)(1,0,1) <sub>12</sub> (*)	SARIMA (1,1,2)(1,0,2) <sub>12</sub> (AIC)	SARIMA (0,1,1)(1,0,1) <sub>12</sub> (SBC)	VALOR REAL	DIF.% (*)	DIF.%(AIC )	DIF.%(SBC )
jan/19	1105708,65	1104065,7	1096937,76	1111331,28	0,51%	0,65%	1,30%
fev/19	988299,4324	990434,49	991291,582	1003614,98	1,53%	1,31%	1,23%
mar/19	964922,4361	972110,73	956696,038	957503,236	- 0,77%	-1,53%	0,08%
abr/19	1010321,343	1011374,6	1011230,03	1007325,8	- 0,30%	-0,40%	-0,39%
mai/19	987545,6475	988681,98	982964,22	1006608,38	1,89%	1,78%	2,35%

Fonte: Elaborado pelo autor a partir de dados da CONFAZ (2019)

Nota 1: Diferencial percentual (DIF.%)

Nota 2:(\*) modelo escolhido

Conforme a tabela 6 é possível afirmar de um modo geral, que os resultados de previsão obtidos pelos três modelos apresentaram bom desempenho. Contudo, o SARIMA (1,1,2)(1,0,1)<sub>12</sub> foi o de melhor capacidade preditiva.

Com o intuito de constatar e comparar a acurácia dos modelos ajustados optou-se ainda, por realizar os testes de EAM, EAPM e EQM. Para verificar a qualidade de previsão dos modelos selecionados, a tabela 7 exibe os resultados das estatísticas de erro.

Tabela 6 Medidas de estatísticas de erro

MODELOS	EAM	EAPM	REQM
SARIMA(1,1,2)(1,0,1) <sub>12</sub> (*)	0,032758	0,23827	0,04524
SARIMA(1,1,2)(1,0,2) <sub>12</sub> (AIC)	0,032895	0,23923	0,045095
SARIMA(0,1,1)(1,0,1) <sub>12</sub> (SBC)	0,033297	0,24235	0,048415

Fonte: Elaborado pelo autor a partir de dados da CONFAZ (2019)

Pelos critérios de EAM e EAMP, observa-se que o SARIMA (1,1,2)(1,0,1) foi o que apresentou os menores valores de perda. Entre os modelos concorrentes o modelo de melhor ajuste pelo critério AIC obteve melhores resultados pelas métricas de erro, porém não foi o modelo que forneceu previsões mais acuradas.

## 5 CONSIDERAÇÕES FINAIS

Neste trabalho, buscou-se identificar, por meio da metodologia Box-Jenkins, modelos de previsão para a série de arrecadação de ICMS do estado do Ceará. Os dados utilizados correspondem ao período de janeiro de 2018 a maio de 2019 e foram deflacionados pelo IPCA com variação mensal, tendo como período base dezembro de 2018.

Os modelos foram selecionados através dos critérios de informação AIC e SBC, e dentre os modelos identificados, foram escolhidos os três de melhor resultado, conforme os critérios mencionados. É importante ressaltar que os dados utilizados não eram estacionários, sendo assim, foi necessário realizar alguns procedimentos para tornar a série estacionária e dar continuidade às etapas da metodologia.

Ademais, foi realizado o teste Ljung-Box para verificar a existência de autocorrelação residual, o resultado obtido pelo teste não rejeitou a hipótese nula, ou seja, os modelos selecionados não apresentam dependência linear do termo de erro o que os tornam consistentes do ponto de vista estatístico.

Dentre os modelos escolhidos o que apresentou o melhor resultado preditivo e pelas métricas de desempenho adotadas neste trabalho foi o SARIMA  $(1,1,2)(1,0,1)_{12}$ . A identificação de sazonalidade na série de ICMS do estado do Ceará corrobora com o trabalho de Fontenele (2017). Pelas medidas de EAM, EAMP e REQM, pôde-se constatar que entre os modelos alternativos o de melhor ajuste pelo critério AIC foi o preferido, todavia o de melhor ajuste pelo critério SBC forneceu resultados semelhantes.

De uma forma geral, os modelos univariados da família Box-Jenkins podem ser considerados uma poderosa ferramenta para previsões de impostos como o ICMS. Na prática, a correta previsão de receitas tributárias serve como guia de decisão para os agentes econômicos, permitindo elaborações de políticas públicas de cunho social e econômico de forma eficiente.

A presente pesquisa limita-se em estudar apenas uma metodologia de previsão, deste modo, recomenda-se que outros métodos sejam testados e comparados com os resultados obtidos por este trabalho.

## REFERÊNCIAS

- BALTAR, Bruno de Paula; KLOTZLE, Marcelo Cabus (Advisor). **Temporal Analysis of Commodity Copper Price's Using the Box & Jenkins Model**. Rio de Janeiro, 2008. 63p. MSc. Dissertation - Departamento de Administração de Empresas. Pontifícia Universidade Católica do Rio de Janeiro.
- BOX, G.; JENKINS, G.; REINSEL, G. **Time Series Analysis: forecasting and control**. 4. ed. Nova Iorque: John Wiley & Sons, 2008.
- BUENO, R.L.S. **Econometria de séries temporais**. 2. ed. São Paulo: Cengage Learning, 2010.
- CAMPOS, P. A. C., Clemente, A., & DE CORDEIRO, A. A. L. (2006). Aplicação do modelo ARIMA para previsão do preço do frango inteiro resfriado no grande atacado do estado de São Paulo. In Anais do Congresso Brasileiro de Custos-ABC.
- CHAIN, C. P.; COSTA, D. F.; SANT'ANA, N. L. dos; BENEDICTO, G. C. de. Contribuição da modelagem de valores atípicos na previsão da arrecadação do ICMS do Estado de Minas Gerais. **Exacta**, v. 13, n. 2, 2015.
- CLEMENTE, Ademir; CLEMENTE, Leonel Toshio. Aplicação da metodologia Box-Jenkins para previsão do ICMS do estado do Paraná de agosto de 2011 a julho de 2012. **Economia e Tecnologia**, v.27, n. 4, 2011.
- DE PAULA BALTAR, Bruno. **Análise Temporal dos Preços da Commodity Cobre Usando o Modelo Box & Jenkins**, p. 24-38, 2009. Tese de Doutorado. PUC-Rio.
- QUEIRÓS, Emerson Oliveira. **Modelo de previsão para receita tributária estadual: aplicação da metodologia Box-Jenkins**. 2012.
- ENDERS, W. **Applied econometric time series**. 3. ed. New Jersey: Wiley, 2010.
- FONTENELE, S. V. P. **Arrecadação Tributária: Modelos de Previsão para o Estado do Ceará**. Dissertação (Mestrado em Economia). Universidade Federal do Ceará, Faculdade de Administração, Ciências Contábeis e Ciências Econômicas, Programa de Pós-Graduação em Economia (CAEN), Fortaleza, 2017.
- FORTUNA, Eduardo. **Mercado Financeiro**. 20ª ed. Editora Qualitymark, 2015.
- GUJARATI, D. N. **Econometria Básica**. Rio de Janeiro: 4ª Ed. AMGH Editora Ltda, 2006.
- KEYNES, J. M. **Teoria geral do emprego, do juro e da moeda**. São Paulo: Nova Cultura Editora. 1973.



LIMA, F. D. **A importância do ICMS na receita pública do Estado do Ceará**. Dissertação (Mestrado em Economia). Universidade Federal do Ceará, Faculdade de Administração, Ciências Contábeis e Ciências Econômicas, Programa de Pós-Graduação em Economia (CAEN), Fortaleza, 2010.

LJUNG, G.M.; BOX, G.E.P. On a measure of lack of fit in time series models. **Biometrika**, v.65, n.2, p.97-303, 1978.

MARTIN, A. C., HENNING, E., WALTER, O. M. F. C., & KONRATH, A. C. (2016). Análise de séries temporais para previsão da evolução do número de automóveis no Município de Joinville. **Revista ESPACIOS** | v. 37 (nº 6) Año 2016.

MORETTIN; P. A.; TOLOI, C. **Análise de Séries Temporais**. São Paulo: Edgar Blucher, 2006.

REIMBOLD, M. M. P., JOZIEL, D., VALER, L., & KIECKOW, F. **Aplicação de Teste de Raiz Unitária às Variáveis de Propulsores Eletromecânicos**.

REZENDE, F. ICMS: como era, o que mudou ao longo do tempo, perspectivas e novas mudanças. **Cadernos Fórum Fiscal**, Brasília, n. 10, p. 1-50, 2009.

RIBEIRO, Livio. **Modelos mensal e trimestral para projeção de arrecadação do ICMS para o estado do Espírito Santo**. Instituto Jones dos Santos Neves, Texto para discussão, ISBN 978-85-62509-14-8. Vitória, ES, 25p. Il, 2010. Disponível em: <http://www.ijsn.es.gov.br/component/attachments/download/900>.

ROCHA, Francisco José Sales. Metodologia Box-Jenkins de escolha de modelos de séries temporais sazonais: um estudo do comportamento temporal da produção física industrial (PFI) de Pernambuco. **Revista Economia & Gestão**, v.3, p.10-20, 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. Simpósio de Pesquisa Operacional e Logística da Marinha, 17. **Anais...** São Paulo: 2014.

SOBIESKI, B. C.; SILVA, D. S.; SOUZA, L. D. ICMS antecipado aplicado nas empresas do Simples Nacional. **Revista Conexão Eletrônica**, v. 12, n. 1, 2015.

SOUZA, K. R. *et al.* Previsão de receitas tributárias: uma análise do ICMS no Ceará por meio de séries temporais. **Revista Fatec Zona Sul**, v. 5, n. 5, 2019.

WERNER, Liane; RIBEIRO, José Luis Duarte. Previsão de demanda: uma aplicação dos modelos Box-Jenkins na área de assistência técnica de computadores pessoais. **Gestão & Produção**, São Carlos - SP, v. 10, n.1, p. 47-68, 2003.