



UNIVERSIDADE ESTADUAL DA PARAÍBA  
CENTRO DE CIÊNCIAS E TECNOLOGIA  
DEPARTAMENTO DE ESTATÍSTICA

Jonatha Felipe Farias

# **Uma aplicação de séries temporais na arrecadação do ICMS do Estado da Paraíba**

Campina Grande - PB

Agosto de 2017

Jonatha Felipe Farias

## **Uma aplicação de séries temporais na arrecadação do ICMS do Estado da Paraíba**

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Orientador: Prof. Dr. Sílvio Fernando Alves Xavier Júnior

Campina Grande - PB

Agosto de 2017

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

F224a Farias, Jonatha Felipe..  
Uma aplicação de séries temporais na arrecadação do ICMS do estado da Paraíba [manuscrito] / Jonatha Felipe Farias. - 2017.  
56 p. : il. colorido.  
Digitado.  
Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia , 2018.  
"Orientação : Não Informado , ."  
1. 1. Imposto Sobre Circulação de Mercadorias e Serviços – ICMS. 2. Séries temporais. 3. Modelagem Box-Jenkins..  
I. Título

21. ed. CDD 519.5



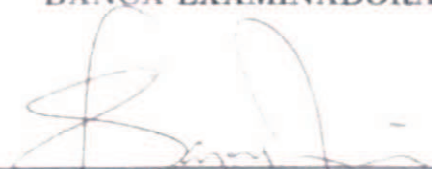
Jonatha Felipe Farias

## Uma aplicação de séries temporais na arrecadação do ICMS do Estado da Paraíba

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Trabalho aprovado em 10 de Agosto de 2017.

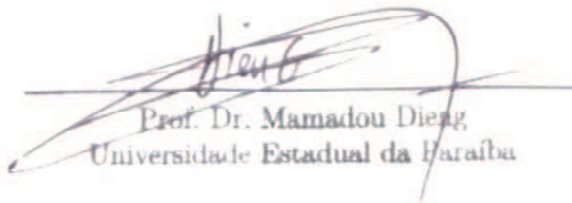
### BANCA EXAMINADORA



Prof. Dr. Sílvio Fernando Alves Xavier  
Júnior  
Universidade Estadual da Paraíba



Prof. Dr. Divanilda Maia Esteves  
Universidade Estadual da Paraíba



Prof. Dr. Mamadou Dieng  
Universidade Estadual da Paraíba

*Dedico este trabalho primeiramente a Deus, por ser essencial em minha vida, autor de meu destino, meu guia, socorro presente na hora da angústia, ao meu pai Edvaldo, minha mãe Maria e as minhas irmãs, ao meu professor pela paciência e incentivo na orientação, aos meus amigos que me apoiaram nessa caminhada universitária, aos meus parentes e a todos que mesmo longe deram aquela força.*

# Agradecimentos

Agradeço primeiramente a Deus por ter me dado forças para superar as dificuldades nessa vida e nessa longa caminhada acadêmica, por minha vida, família e amigos.

Ao professor Sílvio pela oportunidade, apoio, confiança e paciência na elaboração deste trabalho, e também a todos os professores por me proporcionarem o conhecimento não apenas racional, mas a manifestação do caráter e afetividade da educação no processo de formação profissional. A palavra mestre, nunca fará justiça aos professores dedicados aos quais sem nominar terão os meus eternos agradecimentos.

Agradeço a minha mãe Maria, heroína que me deu apoio, incentivo nas horas difíceis, de desânimo e cansaço. Ao meu pai Edvaldo, que apesar de todas as dificuldades me fortaleceu e que para mim foi muito importante. Obrigado! irmãs, primos e tias pela contribuição valiosa.

Meus agradecimentos aos amigos que fizeram parte da minha formação e que vão continuar presentes em minha vida com certeza.

A todos que direta ou indiretamente fizeram parte da minha formação, o meu muito obrigado.

*“Não ganhe o mundo e perca sua alma;  
sabedoria é melhor que prata e ouro.”  
(Bob Marley)*



# Resumo

O presente trabalho tem como intuito verificar as melhores maneiras de modelar os dados a uma série temporal e prever os valores de arrecadação do ICMS no Estado da Paraíba. Para o deflacionamento dos dados na série usamos o IGP-DI. As previsões foram calculadas usando o algoritmo de alisamento exponencial de Holt-Winters, tanto o aditivo como o multiplicativo, e também o método de modelagem de Box-Jenkins, SARIMA e SARIMAX. Todas as previsões foram geradas com 1 e 5 passos a frente, e os valores das somas dos quadrados erros foram a forma de avaliar quais os melhores valores previstos. Na modelagem SARIMAX foram acrescentadas algumas variáveis para verificar se estas afetariam de alguma forma em suas previsões, porém todos os modelos mostram uma boa capacidade de previsão, com intervalos de valores bem próximos, apresentando diferenças também não tão gritantes em seus erros relativos.

**Palavras-chaves:** ICMS, Holt-Winters, SARIMA, SARIMAX.

# Abstract

The present work aims to verify the best ways to model the data to a time series and to predict ICMS collection values in the State of Paraíba. For the deflation of the data in the series, we used the IGP-DI. To make the predictions, we use the exponential smoothing algorithm of Holt-Winters, both additive and multiplicative, as well as the Box-Jenkins, SARIMA and SARIMAX modelling method. All predictions were generated with 1 and 5 steps forward, and the values of the sums of the squared errors were the way to evaluate the best-predicted values. For the SARIMAX model, some variables were added to verify if they would affect in any way their predictions, but all the models show a good predictability, with intervals of values very close, presenting differences also not so glaring in their relative errors.

**Key-words:** ICMS, Holt-Winters, SARIMA, SARIMAX.

# Lista de ilustrações

Figura 1 – ICMS na Paraíba entre janeiro de 2003 até agosto de 2016. . . . .	35
Figura 2 – Gráficos do Correlograma e Correlograma Parcial dos dados reais. . . .	36
Figura 3 – Gráficos do Correlograma e Correlograma Parcial da primeira diferença dos dados logaritmados. . . . .	37
Figura 4 – Gráficos dos dados da série e seu ajuste de Holt-Winters. . . . .	38
Figura 5 – Gráficos da previsão dos dados sem e com atenuação de tendência. . .	39
Figura 6 – Gráficos de diagnóstico dos resíduos dos modelos I,III e VII para a série de arrecadação de ICMS na Paraíba . . . . .	43

# Lista de tabelas

Tabela 1 – Testes da raiz unitária de Dickey-Fuller . . . . .	30
Tabela 2 – Valores otimizados das constantes de suavização para Holt-Winters aditivo e Multiplicativo . . . . .	37
Tabela 3 – Previsões adquiridas pelos algoritmos Holt-Winters aditivo e Multiplicativo . . . . .	38
Tabela 4 – Erros de previsão . . . . .	39
Tabela 5 – Valores referentes ao critérios de escolha dos modelos da série . . . . .	41
Tabela 6 – Previsões com 5 passos a frente dos modelos SARIMA e SAMIRAX, Agosto - Outubro . . . . .	42
Tabela 7 – Previsões com 5 passos a frente dos modelos SARIMA e SAMIRAX, Novembro e Dezembro . . . . .	42
Tabela 8 – Erros relativos de previsão dos modelos . . . . .	42
Tabela 9 – Conjunto de Dado (2003 - 2009) . . . . .	49
Tabela 10 – Conjunto de Dado (2010 - 2016) . . . . .	49
Tabela 11 – Previsões intervalares do algoritmo Holt-Winters sem atenuação de tendência . . . . .	50
Tabela 12 – Previsões intervalares do algoritmo Holt-Winters com atenuação de tendência . . . . .	51

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>13</b>
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>15</b>
2.1	Conceitos Básicos	15
2.2	Processos Estocásticos	15
2.3	Funções de Autocovariância e Autocorrelação	17
2.4	Modelos de Suavização Exponencial	18
2.4.1	Alisamento Exponencial Simples (AES)	18
2.4.2	Alisamento Exponencial de Holt	19
2.4.3	Alisamento Exponencial de Holt-Winters	20
2.5	Modelagem Box-Jenkins	21
2.5.1	Modelo Autorregressivo (AR)	22
2.5.2	Modelo de Médias Móveis (MA)	23
2.5.3	Modelo autorregressivo de médias móveis (ARMA)	24
2.5.4	Modelo autorregressivo integrado de médias móveis (ARIMA)	24
2.5.5	Modelo autorregressivo de médias móveis sazonal (SARIMA)	25
2.5.6	Modelo autoregressivo de médias móveis sazonal com variáveis exógenas (SARIMAX)	26
2.5.7	Função de autocorrelação (FAC)	26
2.5.8	Função de autocorrelação parcial (FACP)	27
2.5.9	Etapas do modelo SARIMA	28
2.6	Avaliação de Previsão	33
<b>3</b>	<b>APLICAÇÃO</b>	<b>35</b>
3.1	Alisamento Exponencial	37
3.2	Modelagem SARIMA	40
<b>4</b>	<b>CONCLUSÃO</b>	<b>45</b>
	<b>REFERÊNCIAS</b>	<b>46</b>
	<b>APÊNDICES</b>	<b>48</b>
	<b>APÊNDICE A – CONJUNTO DE DADOS</b>	<b>49</b>
	<b>APÊNDICE B – PREVISÕES INTERVALARES DE HOLT-WINTERS</b>	<b>50</b>

APÊNDICE C – ROTINA DO R . . . . .	52
------------------------------------	----

# 1 Introdução

O principal meio de arrecadação por estados do nosso país é por meio do Imposto de Circulação de Mercadorias e Serviços (ICMS), com um percentual por volta de 20% do que se arrecada no Brasil, mostrando que é bastante importante no dimensionamento da receita. É de extrema importância que a estimativa das receitas sejam feitas corretamente, principalmente do ICMS, pois é de extrema importância nas contas do estado, e para isso é necessário um modelo de previsão de arrecadação muito preciso para o governo (MENDONÇA; MEDRANO, 2016). Tal estimativa é supervisionada pela Lei de Responsabilidade Fiscal (BRASIL, 2000), que no artigo 11 diz que tal responsabilidade da previsão, gestão e arrecadação de tributos é dos entes federativos.

Antes da existência do ICMS, em 1967, o meio de arrecadação existente era o IVC (Imposto sobre Vendas e Consignações), que foi substituído pelo ICM (Imposto sobre Circulação de Mercadorias). Porém em 1988, com inclusão de novos tributos que eram arrecadados separadamente, por exemplo energia elétrica, petróleo e derivados, surge o ICMS (REZENDE, 2009).

Um processo que estude diversos métodos na estimativa das receitas públicas deve ser bem preciso, uma vez que irá prever a quantidade que será gasta por seus entes, pois é o pilar no momento de separar as despesas na construção do orçamento (WORDBANK, 2000). Ciente de que os orçamentistas podem aumentar ou diminuir o valor das receitas com intuito de manter um equilíbrio em seu orçamento Rubin (2010) destacou que se um orçamento for superestimado, acarretando em gastos acima do previsto, no momento de sua arrecadação isso gerará um déficit, pois a devida arrecadação será insuficiente, além de que em outro ente pode-se faltar verba (CRUZ; SENHORAS, 2013). Por isso é de extrema importância, para o controle social, estudar um bom método de previsão orçamentária.

Sabendo da importância de uma excelente previsão das receitas em seus orçamentos Scheffer, Souza e Zanini (2014) estudaram métodos de séries temporais para tal previsão no Brasil, coisa que ainda é bem recente no âmbito nacional. O presente trabalho, tem por objetivo utilizar séries temporais e suas metodologias para encontrar um modelo que melhor consiga prever estimativas de arrecadação do ICMS no Estado da Paraíba.

Para tal estudo, vamos analisar os dados de arrecadação de ICMS da Paraíba no período de janeiro de 2003 até agosto de 2016 adquiridos por IPEADATA (b), com intuito de prever valores futuros dessa tal arrecadação. Para isso utilizaremos as metodologias de previsão de Holt-Winters aditivo e multiplicativo, e também a modelagem Box-Jenkins por meio de previsões no modelo SARIMA(p, q, d) x (P, Q, D). Vamos comparar tais modelagens para ver qual delas apresenta um melhor modelo de previsão para arrecadação

do ICMS da Paraíba.

No intuito de evitar uma dinâmica inflacionária nas previsões da série, utilizou-se os dados do IGP-DI (Índice Geral de Preços - Disponibilidade Interna) no mesmo período do ICMS também adquiridos por IPEADATA (a). Tal índice mede a variação mensal de preços. A utilização desse índice fará com que as previsões sejam mais precisas e não sofram com os efeitos do deflacionamento.

O presente trabalho é subdividido: capítulo 1 - Introdução; Capítulo 2 - Fundamentação Teórica, onde está presente todo assunto de séries temporais que se aborda o trabalho; Capítulo 3 - Aplicação da análise dos dados; Capítulo 4 - Conclusão. Por fim, referências bibliográficas e apêndices.



## 2 Fundamentação Teórica

### 2.1 Conceitos Básicos

Séries temporais são um conjunto de observações sequenciais num determinado tempo  $t$ . Estas podendo ser discretas, quando as observações ocorrem em tempo específico ( $T = t_1, \dots, t_n$ ), ou contínuas, quando as observações ocorrem continuamente no tempo ( $T = \{t: t_1 < t < t_2\}$ ). De acordo com Morettin e Tolo (2006) em uma série contínua limitada em um específico intervalo de tempo, dela pode ser retirada uma série discreta por uma amostragem.

Para Box, Jenkins e Reinsel (2008) a série temporal possui uma característica típica, onde as observações mais próximas possuem uma dependência, que é bem interessante, pois as análises de séries temporais utilizam de técnicas, com intuito de analisar tal dependência entre as observações. Morettin e Tolo (2006) apresentam dois destaques no uso da análise de séries temporais, onde o primeiro destaque é que a análise se encontra no domínio temporal e seus modelos são paramétricos, enquanto o segundo está no domínio de frequências e seus modelos são não paramétricos, porém têm o propósito de construir modelos para suas respectivas séries.

Segundo Morettin e Tolo (2006) para compreensão do comportamento das séries temporais é necessário ter conhecimento dos seguintes componentes característicos: sazonalidade, tendência, ciclo e erro aleatório.

A *sazonalidade* é algo que se repete em um determinado espaço de tempo, a *tendência* é o comportamento da série ao seu desenvolver, ou seja, o movimento que ela faz, o *ciclo* é a oscilação dos dados em torno da tendência, já o *erro aleatório* são flutuações (movimentos) que não são identificáveis.

Para entender melhor o contexto de séries temporais, é necessário saber o que são processos estocásticos, uma vez que o processo de séries temporais é originado de um processo estocástico.

### 2.2 Processos Estocásticos

Seja  $T$  um conjunto (normalmente inteiros  $\mathbb{Z}$ , ou reais  $\mathbb{R}$ ) arbitrário, o processo estocástico é uma família  $Z = \{Z_t, t \in T\}$  tal que, para cada  $t \in T, Z_t$  é uma variável aleatória (MORETTIN; TOLOI, 2006), sendo compreendido como um fenômeno estatístico que evolui ao longo do tempo com base nas leis probabilísticas (BOX; JENKINS; REINSEL, 2008). Os processos estocásticos, normalmente, são descritos com base através das seguintes

funções (MORETTIN; TOLOI, 2006) :

- Função média de  $Z_t$  :  $E[Z_t] = \mu_t$ ;
- Função de autocovariância :  $\gamma_{t,s} = \text{Cov}[Z_t, Z_s] = E [(Z_t - \mu_t)(Z_s - \mu_s)]$  , sendo t e s tempos distintos;
- Variância :  $\text{Var}[Z_t] = \sigma_t^2$ .

Nos processos estocásticos existe uma importante classe chamada de processos estacionários, onde esta se encontra num estado particular de equilíbrio estatístico. Existem dois tipos de estacionariedade que são, estritamente (fortemente) estacionária e fracamente estacionária (ou de segunda ordem). Segundo Morettin e Tolo (2006) tem-se as seguintes definições:

- Estacionariedade Forte: Um processo estocástico  $Z = \{Z_t, t \in T\}$  diz-se estritamente estacionário se todas as distribuições finito-dimensionais, ou seja,  $F(z_1, \dots, z_n; t_1, \dots, t_n) = P\{Z_{t_1} \leq z_1, \dots, Z_{t_n} \leq z_n\}$ , permanecem invariantes no tempo.

Assim, temos que sua média e variância são constantes nas translações do tempo, ou seja:

$$E(t) = \mu, \text{ e } \text{Var}(t) = \sigma^2, \forall t \in T.$$

- Estacionariedade Fraca: Um processo estocástico  $Z = \{Z_t, t \in T\}$  diz-se fracamente estacionário, se e somente se:
  1.  $E[Z_t] = \mu$ , constante  $\forall t \in T$ ;
  2.  $E^2[Z_t] < \infty$ ,  $t \in T$ ;
  3.  $\gamma_{t,s} = \text{Cov}[Z_t, Z_s]$  é uma função de t-s, chamado de defasagem.

Também temos o ruído branco ( $u_t$ ), que é outro ponto importante das séries temporais. Para Box, Jenkins e Reinsel (2008) um processo estacionário com uma sequência de variáveis aleatórias iid, denotados  $u_1, \dots, u_t$ , que assumem média 0 e variância  $\sigma_u^2$ , sendo o mesmo um processo estritamente estacionário, referimos a este como um ruído branco.

- Ruído Branco: Processo estocástico não-correlacionados puramente aleatório. Que apresenta as seguintes propriedades:

$$E(\mu_t) = 0$$

$$\text{Var}(\mu_t) = \sigma^2$$

$$\text{Cov}(\mu_t, \mu_j) = 0, \forall t \neq j.$$

Como os  $u_t$  são não-correlacionados, por conta de sua independência, então sua função de autocovariância é dada por:

$$\begin{aligned} \gamma_k &= \text{E}[u_t u_{t+k}] \\ &= \begin{cases} \sigma_u^2 & , k = 0; \\ 0 & , k \neq 0. \end{cases} \end{aligned}$$

Entretanto, se uma mesma sequência de variáveis aleatórias apresentar as mesmas características (média, variância), e estando concentradas num processo de segunda ordem, esse processo também é um ruído branco, contudo, apresenta estacionariedade fraca.

## 2.3 Funções de Autocovariância e Autocorrelação

Para uma análise estrutural em séries temporais é interessante conhecer a relação (dependência) entre observações recentes e passadas, e para esta quantificação de dependência na estrutura estocástica da série, utilizamos a função de autocovariância.

De acordo com Brockwell e Davis (2002), uma série temporal  $Z_t$  estacionária, com  $\text{E}[Z_t^2] < \infty$ , apresenta uma função de autocovariância (FACV) dada por,

$$\gamma(\tau) = \text{Cov}[Z_t, Z_{t+\tau}]$$

De acordo com Morettin e Toloí (2006), a função acima, apresenta as seguintes propriedades,

1.  $\gamma(0) > 0$ ;
2.  $\gamma(-\tau) = \gamma(\tau), \forall t \in T$ ;
3.  $|\gamma(\tau)| \leq \gamma(0), \forall t \in T$ .

A função de autocorrelação (FAC) é originada, por conta da desvantagem da dependência da unidade de medição de  $\gamma$ . A FAC é dada por,

$$\rho(\tau) = \gamma(\tau)/\gamma(0), \forall t \in T.$$

e apresenta propriedades análogas a  $\gamma(\tau)$ , menos para  $\rho(0) = 1$ .

## 2.4 Modelos de Suavização Exponencial

O propósito dos métodos é distinguir o padrão de qualquer ruído que possa estar contido nas observações e então usar esse padrão para prever valores futuros das séries (MORETTIN; TOLOI, 2006). O alisamento (suavização) é um grande método para verificar tal acontecimento nas previsões de séries temporais.

Os modelos de suavização exponencial, que são uma classe de algoritmos de previsão *ad hoc*, são métodos simples de previsões adaptativas, onde dados recentes são ajustados automaticamente para serem incluídos na série. É importante salientar que mesmo dispondo de poucas observações, este método consegue apresentar uma efetiva previsão.

Pelos erros encontrados em previsões passadas, é que são obtidas as previsões dos modelos de suavização exponencial, ou seja, considerando  $Z_1, Z_2, \dots, Z_t$  uma série temporal, suas previsões serão construídas para  $Z_{t+1}, Z_{t+2}, \dots$ .

### 2.4.1 Alisamento Exponencial Simples (AES)

O método de suavização (alisamento) exponencial simples é utilizado para séries temporais onde não existem a presença de tendência nem sazonalidade (EHLERS, 2005), ou seja, são séries localmente constantes, sendo compostas apenas por seu nível e ruído aleatório. Sendo a série decomposta como

$$Z_t = N_t + \epsilon_t, \text{ com } t = 1, 2, \dots, n. \quad (2.1)$$

Em que  $\epsilon_t \sim (0, \sigma_\epsilon^2)$  e  $N_t$  representam o nível da série no instante  $t$ . O parâmetro  $N_t$  é estimado pela média de observações passadas, onde os pesos relativamente maiores são atribuídos as observações mais recentes (pesos crescem exponencialmente ao longo do tempo), como é visto abaixo (CHATFIELD, 2004):

$$N_t = \alpha Z_t + \alpha(1 - \alpha)Z_{t-1} + (1 - \alpha)^2 Z_{t-2} + \alpha(1 - \alpha)^3 Z_{t-3} + \dots \quad (2.2)$$

onde  $Z_t$  são as  $n$  observações da série e  $\alpha$  é a constante de suavização com  $0 < \alpha < 1$ .

Normalmente é usado a forma recursiva do algoritmo, por conta do uso da observação no instante atual e do uso do nível no instante anterior dada por:

$$N_t = \alpha Z_t + (1 - \alpha)N_{t-1} \quad (2.3)$$

Assim, temos que a estimativa do nível é uma medida ponderada entre  $N_{t-1}$  (estimativa anterior do nível) e  $Z_t$  (observação atual da série).

A previsão de valores futuros com  $h$  passos a frente para qualquer  $h = 1, 2, \dots$  é dada pelo último valor do nível no momento  $t$ , isto é:

$$\widehat{Z}_t(h) = N_t, \text{ onde } h = 1, 2, \dots \quad (2.4)$$

onde, geralmente o algoritmo se inicializa pela primeira observação: ( $N_1 = Z_1$ ).

Segundo Morettin e Tolo (2006), deve-se escolher a constante de suavização  $\alpha$ , para que tal valor forneça uma melhor previsão, de modo que minimize a soma dos quadrados dos erros de previsão.

Para a interpretação de tal modelo ainda existe uma forma de correção dos erros do algoritmo, a cada passo, com base no último erro que foi observado. Ou seja, se foi positivo o último erro de previsão, conseqüentemente a próxima previsão também aumentará por conta de tal erro, porém se o erro foi negativo, conseqüentemente sua previsão será reduzida. É apresentado da seguinte maneira:

$$N_t = N_{t-1} + \alpha \epsilon_t. \quad (2.5)$$

## 2.4.2 Alisamento Exponencial de Holt

O método de Holt utiliza dois coeficientes diferentes de alisamento, são eles o nível (também presente no AES) e a tendência linear (crescente ou decrescente) (MONTGOMERY; JENNINGS; KULAHCI, 2015). Quando uma série apresenta tendência é utilizado o método de AEH, pois previsões futuras pelo AES induziriam a um erro sistemático de previsão devido ao atraso no acompanhamento da tendência.

Esses coeficientes são estimados da seguinte maneira:

$$N_t = \alpha Z_t + (1 - \alpha)(N_{t-1} + T_{t-1}) \quad (2.6)$$

$$T_t = \beta(N_t - N_{t-1}) + (1 - \beta)T_{t-1}, \quad (2.7)$$

com  $0 < \alpha < 1$  e  $0 < \beta < 1$ , onde  $\alpha$  e  $\beta$  são as constantes de suavização .

Para determinar os coeficientes de suavização  $\alpha$  e  $\beta$ , de tal forma que minimize a soma dos quadrados dos erros de previsão, utiliza-se um processo idêntico ao do AES, porém o algoritmo se inicia pela segunda observação, ou seja,  $N_2 = Z_2$  e  $T_2 = Z_2 - Z_1$  (MORETTIN; TOLOI, 2006).

Assim, a previsão é dada por:

$$\widehat{Z}_t(h) = N_t + hT_t, \text{ onde } h = 1, 2, \dots \quad (2.8)$$

A forma de correção dos erros do algoritmo de Holt é feito de maneira análoga ao do AES, que resulta em:

$$N_t = N_{t-1} + T_{t-1} + \alpha\epsilon_t \quad (2.9)$$

$$T_t = T_{t-1} + \beta\epsilon_t. \quad (2.10)$$

### 2.4.3 Alisamento Exponencial de Holt-Winters

Este tipo de alisamento é usado em previsões de séries temporais onde existe um padrão de comportamento mais complexo, ou seja, séries que apresentam tendência linear e sazonalidade (BROCKWELL; DAVIS, 2002). Este algoritmo é dividido em duas partes: os modelos aditivos e multiplicativos. A utilização desses algoritmos é baseada nos seguintes componentes: nível, tendências e sazonalidade (MORETTIN; TOLOI, 2006).

- *Série sazonal multiplicativa*: É considerado uma série sazonal com tal período  $s$ , onde ele pode ser dia(s), mês(es) ou ano(s). Temos  $F_t$  como a variante do fator sazonal que é multiplicativa, enquanto a variante tendência é aditiva (MORETTIN; TOLOI, 2006). A forma de recorrência deste algoritmo, é dada pelas três seguintes equações:

$$N_t = \frac{\alpha Z_t}{F_{ts}} + (1 - \alpha)(N_{t-1} + T_{t-1}) \quad (2.11)$$

$$T_t = \beta(N_t + N_{t-1}) + (1 - \beta)T_{t-1} \quad (2.12)$$

$$F_t = \frac{\gamma Z_t}{N_t} + (1 - \alpha)F_{t-s}, \quad (2.13)$$

com  $0 < \alpha < 1$ ,  $0 < \beta < 1$ ,  $0 < \gamma < 1$ , onde  $\alpha$ ,  $\beta$  e  $\gamma$  são as constantes de suavização do algoritmo.

Para previsões dos valores futuros é considerado que em cada equação o fator sazonal contabiliza o período correspondente. Abaixo temos as seguintes expressões destas previsões (MORETTIN; TOLOI, 2006):

$$\hat{Z}_t(h) = (N_t + hT_t)F_{t+h-s}, \text{ onde } h = 1, 2, \dots, s \quad (2.14)$$

$$\hat{Z}_t(h) = (N_t + hT_t)F_{t+h-2s}, \text{ onde } h = s+1, s+2, \dots, 2s. \quad (2.15)$$

Com alguns procedimentos matemáticos, chegamos a forma de correção dos erros:

$$N_t = N_{t-1} + T_{t-1} + \frac{\alpha\epsilon_t}{F_{t-s}} \quad (2.16)$$

$$T_t = T_{t-1} + \alpha\beta\frac{\epsilon_t}{F_{t-s}} \quad (2.17)$$

$$F_t = F_{t-s} + \gamma(1 - \alpha)\frac{\epsilon_t}{F_{t-s}} \quad (2.18)$$

- *Série sazonal aditiva*: Modificando o processo anterior em situações em que o fator sazonal passa a ser aditivo, suas atualizações dos dados são dadas pelas seguintes expressões (MORETTIN; TOLOI, 2006):

$$N_t = \alpha(Z_t - F_{ts}) + (1 - \alpha)(N_{t-1} + T_{t-1}) \quad (2.19)$$

$$T_t = \beta(N_t + N_{t-1}) + (1 - \beta)T_{t-1} \quad (2.20)$$

$$F_t = \gamma(Z_t - N_t) + (1 - \gamma)F_{t-s} \quad (2.21)$$

com  $0 < \alpha < 1$ ,  $0 < \beta < 1$ ,  $0 < \gamma < 1$ , onde  $\alpha$ ,  $\beta$  e  $\gamma$  são as constantes de suavização do algoritmo.

As previsões futuras, com  $h$  passos à frente, são apresentadas da seguinte forma (MORETTIN; TOLOI, 2006):

$$\hat{Z}_t(h) = N_t + hT_t + F_{t+h-s}, \text{ onde } h = 1, 2, \dots, s \quad (2.22)$$

$$\hat{Z}_t(h) = N_t + hT_t + F_{t+h-2s}, \text{ onde } h = s+1, s+2, \dots, 2s. \quad (2.23)$$

Com alguns procedimentos matemáticos, chegamos a forma de correção dos erros:

$$N_t = N_{t-1} + T_{t-1} + \alpha\epsilon_t \quad (2.24)$$

$$T_t = T_{t-1} + \alpha\beta\epsilon_t \quad (2.25)$$

$$F_t = F_{t-s} + \gamma(1 - \alpha)\epsilon_t. \quad (2.26)$$

## 2.5 Modelagem Box-Jenkins

Para escolha de um determinado modelo, seria interessante que compreendêssemos totalmente o funcionamento de um certo fenômeno. Desta forma, seria possível escrever uma expressão matemática que conseguiria explicar concretamente tal fenômeno. Porém, é notório que isso é praticamente impossível por diversos motivos. Não obstante, para construção de um determinado modelo, é usado uma base de conhecimento empírico, teórico e matemático para estimação de parâmetros a partir dos dados experimentais (BOX; JENKINS; REINSEL, 2008).

De acordo com Morettin e Toloi (2006), a metodologia Box e Jenkins, baseia-se em um determinado ciclo interativo para estruturar um modelo de previsão, baseado nos próprios dados da série.

Tal ciclo interativo, é dividido em quatro etapas:

1. Uma classe útil de modelos é considerada (especificação);

2. Como tal classe de modelos é extensa, são desenvolvidas subclasses deste modelo, com intuito de identificar um modelo, baseado na análise de autocorrelações e autocorrelações parciais (identificação);
3. Agora, são estimados os parâmetros do modelo identificado (estimação);
4. Pela análise de resíduos, verificamos se o modelo ajustado está adequado para fazer as previsões (verificação ou diagnóstico).

Caso o modelo escolhido não seja adequado para representar os dados, o ciclo deve ser repetido novamente, voltando a fase de identificação. Se o propósito é previsão, então devemos analisar diversos modelos, afim de escolher o melhor modelo ajustado com o menor erro quadrático médio de previsão. Porém, é possível que em estudos entre diferentes pessoas, com o mesmo dado, sejam identificados diferentes modelos. Evidentemente os modelos são postulados, pois são simples e com poucos parâmetros, contudo as previsões são bem precisas.

### 2.5.1 Modelo Autorregressivo (AR)

No processo autorregressivo, a série é modelada tendo como base uma média ponderada de seus valores prévios e também por um erro aleatório (ruído branco) (BOX; JENKINS; REINSEL, 2008). Sendo denotado AR (p), um modelo autorregressivo de ordem p, temos:

$$Z_t = \alpha + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + u_t, \quad (2.27)$$

onde  $\phi_1, \phi_2, \dots, \phi_p$  são os parâmetros do modelo e  $u_t$  é o ruído branco no tempo t.

Mas esta não é a única maneira de expressar esta modelagem. Ela tem uma notação alternativa, em que definimos um operador autorregressivo estacionário de ordem p (MORETTIN; TOLOI, 2006), dado por:

$$Z_t = \alpha + \phi_1 B Z_t + \phi_2 B^2 Z_t + \dots + \phi_p B^p Z_t + u_t. \quad (2.28)$$

Pondo em evidencia  $Z_t$ , a expressão se apresenta desta forma,

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) Z_t = \alpha + u_t \quad (2.29)$$

e resumindo tal expressão, temos

$$\phi(B) Z_t = \alpha + u_t, \quad (2.30)$$



onde  $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$  é o polinômio operador de defasagens.

No modelo autorregressivo, a variância é dada por,

$$\gamma_0 = \phi_1 \gamma_1 - \phi_2 \gamma_2 - \dots - \phi_p \gamma_p + \sigma_u^2 \quad (2.31)$$

sendo que  $\sigma_u^2$  é a variância de  $u_t$  (ruído branco).

Já a autocovariância é definida da seguinte forma:

$$\gamma_k = \phi_1 \gamma_{k-1} - \phi_2 \gamma_{k-2} - \dots - \phi_p \gamma_{k-p} + \sigma_u^2, k \geq 1. \quad (2.32)$$

Porém, fazendo  $k = 1, 2, \dots, p$  e utilizarmos  $\gamma_0 = 1$  e  $\gamma_{-k} = \gamma_k$ , na equação de autocovariância, conseguimos gerar a equação de Yule-Walker (BOX; JENKINS; REINSEL, 2008).

## 2.5.2 Modelo de Médias Móveis (MA)

O nome média móvel se dá porque a série  $Z_t$  é construída por meio de uma soma ponderada finita dos  $q + 1$  valores mais recentes do ruído branco, ou seja, a combinação linear dos erros aleatórios passados gera o valor mais recente da série (BOX; JENKINS; REINSEL, 2008). Porém a nomenclatura, média móvel, não se encaixa corretamente a tal modelagem.

Um modelo de médias móveis de ordem  $q$ , é denotado por MA( $q$ ), sendo representado por (MORETTIN; TOLOI, 2006),

$$Z_t = \mu + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \dots - \theta_q u_{t-q}. \quad (2.33)$$

Podendo reestruturá-la como

$$Z_t = \mu + \theta(B)u_t, \quad (2.34)$$

onde  $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$  é o polinômio operador de médias móveis de ordem  $q$ .

A variância do modelo MA( $q$ ), é dada pela seguinte equação,

$$\gamma_0 = (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) \sigma_u^2. \quad (2.35)$$

Sua autocovariância é dada por,

$$\left\{ \begin{array}{ll} 0 & , \text{se } k > q; \\ \gamma_k = (-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \dots + \theta_q\theta_{q-k})\sigma_u^2 & , \text{onde } k = 0, 1, \dots, q = 0; \\ \gamma(-k) & , \text{se } k < 0. \end{array} \right. \quad (2.36)$$

### 2.5.3 Modelo autorregressivo de médias móveis (ARMA)

Ao combinar  $p$  termos do modelo autorregressivo com  $q$  termos do modelo médias móveis, obtemos o que é chamado de modelo misto autorregressivo de médias móveis, denotado por ARMA ( $p, q$ ). O processo ARMA procura aproximar o conjunto de dados de maneira mais parcimoniosa, ou seja, utilizar de uma quantidade menor de parâmetros para representar a série (MORETTIN; TOLOI, 2006). Tal modelagem é uma combinação linear da média ponderada de seus  $p$  valores prévios de sua série juntamente com seu ruído branco no tempo presente e nos  $q$  tempos passados (BOX; JENKINS; REINSEL, 2008).

O modelo ARMA ( $p, q$ ), é definido da seguinte forma:

$$Z_t = \alpha + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \dots - \theta_q u_{t-q}, \quad (2.37)$$

sendo os  $\phi$ 's são parâmetros autorregressivos e os  $\theta$ 's parâmetros das médias móveis, respectivamente.

Ainda assim, podemos reescrever tal equação de maneira mais compacta:

$$\phi(B)Z_t = \alpha + \theta(B)u_t \quad (2.38)$$

onde  $\phi(B)$  e  $\theta(B)$  são os polinômios AR e MA, respectivamente.

A variância é dada por,

$$\gamma_k = \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p} - \theta_1 \gamma_{yu}(k-1) - \dots - \theta_q \gamma_{yu}(k-q), k \leq q, \quad (2.39)$$

onde  $\gamma_{yu}(k) = E[u_t Z_{t-k}]$ .

Já a autocovariância do modelo ARMA, é expressada pela seguinte equação:

$$\gamma_k = \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p}, k > q. \quad (2.40)$$

### 2.5.4 Modelo autorregressivo integrado de médias móveis (ARIMA)

Tais modelos apresentados anteriormente, são utilizados nos casos em que a série seja estacionária (média constante ao longo tempo), contudo sabemos que isso raramente

ocorre, e que geralmente as séries são não estacionárias (MORETTIN; TOLOI, 2006). Para este problema de séries não estacionárias, usamos o modelo autorregressivo integrado de médias móveis (ARIMA), que é uma generalização do modelo ARMA. Se uma série for não estacionária, a série pode se tornar estacionária ao aplicarmos  $d$  diferenças aos dados, onde no máximo 2 diferenças podem ser aplicadas (BOX; JENKINS; REINSEL, 2008). Então, para diferença de 1ª ordem é aplicada o seguinte procedimento:

$$w_t = Z_t - Z_{t-1} = (1 - B)Z_t = \Delta Z_t. \quad (2.41)$$

Já para diferenças de ordem superior ( $d > 1$ ), temos o seguinte procedimento:

$$w_t = \Delta^d Z_t = \Delta^{d-1}(\Delta^1 Z_t), \quad (2.42)$$

onde  $\Delta^d$  é o operador de defasagem.

A modelagem é denotada por ARIMA(p,d,q), onde  $p$  é o operador autorregressivo,  $d$  é o operador diferença e  $q$  é o operador de médias móveis. Essa modelagem é dada por,

$$w_t = \phi_1 w_{t-1} + \dots + \phi_p w_{t-p} + u_t - \theta_1 u_{t-1} - \dots - \theta_q u_{t-q} \quad (2.43)$$

onde  $w_t = \Delta^d Z_t$ .

Podendo ainda ser escrita como:

$$\phi(B)((1 - B)^d Z_t - \alpha) = \theta(B)u_t. \quad (2.44)$$

A variância e autocovariância da modelagem ARIMA são dadas por:

$$\begin{cases} \gamma_k = \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p} - \theta_1 \gamma_{wu}(k-1) - \dots - \theta_q \gamma_{wu}(k-q) & , \text{se } k \leq q; \\ \gamma_k = \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p} & , \text{se } k > q. \end{cases} \quad (2.45)$$

sendo que  $\gamma_0 = E[w_t^2]$ ,  $\gamma_k = E[w_t w_{t-s}]$  e  $\gamma_{wu} = E[u_t w_{t-k}]$ .

### 2.5.5 Modelo autorregressivo de médias móveis sazonal (SARIMA)

Esta modelagem é uma extensão da modelagem ARIMA(p,q,d), onde nela levamos em consideração a sazonalidade estocástica dos dados (MORETTIN; TOLOI, 2006), pois muitas das séries temporais que trabalhamos, apresentam um período de sazonalidade  $s$ , sendo  $s > 1$ .

Denotado por SARIMA (p, d, q) x (P, D, Q), onde P é o número de coeficientes sazonais autorregressivos referentes a estacionariedade, D o número de diferenças sazonais

-  $(1 - B^s)Z_t = Z_t - Z_{t-s}$  - para tornar a série estacionária de acordo com o número de períodos  $s$ , e  $Q$  sendo o número de coeficientes sazonais de médias móveis referentes a invertibilidade (PELLEGRINI, 2000).

Tal modelagem é dada por,

$$\phi(B)\Phi(B)[(1 - B)^d(1 - B)^D - \alpha]Z_t = \theta(B)\Theta(B)u_t. \quad (2.46)$$

### 2.5.6 Modelo autoregressivo de médias móveis sazonal com variáveis exógenas (SARIMAX)

O modelo SARIMAX é um modelo de regressão linear que utiliza um processo do modelo SARIMA, onde é denotado por SARIMAX (p, d, q, s) x (P, D, Q).

Tal modelagem é dada por,

$$\phi(B)\Phi(B)(1 - B)^d(1 - B)^D Z_t = \alpha + X\beta + \theta(B)\Theta(B^S)u_t, \quad (2.47)$$

sendo  $X$  uma matriz  $n \times k$  de regressores não aleatórios e  $\beta$  um vetor de parâmetros  $k \times 1$ .

### 2.5.7 Função de autocorrelação (FAC)

Segundo Montgomery, Jennings e Kulahci (2015), se uma série temporal é estacionária, tal que  $w_t$  e  $w_{t+k}$  sejam duas observações, então sua distribuição de probabilidade conjunta é a mesma para qualquer período de tempo  $t$  e  $t + k$  que são separados por um mesmo intervalo(lag)  $k$ .

A função de autocorrelação, ou correlação serial de ordem  $k$ , é uma ferramenta que é bastante utilizada na modelagem Box-Jenkins (ENDERS, 2004), pois fornece uma descrição parcial do processo ao qual se está querendo modelar, e também mede a correlação entre valores de duas variáveis,  $w_t$  e  $w_{t-k}$ , em períodos de tempos distintos (PELLEGRINI, 2000). Se temos coeficiente de autocorrelação  $\rho_1$ , então verificamos a correlação de tal variável no período  $t$  em relação ao período defasado  $t - 1$ , na qual é chamado de coeficiente de autocorrelação de lag 1, assim como um coeficiente de autocorrelação  $\rho_k$  é dito de coeficiente de autocorrelação de lag  $k$  (PELLEGRINI, 2000).

Tal função é definida como:

$$\rho_k = \frac{Cov[w_t, w_{t-k}]}{Var[w_t]} = \frac{\gamma_k}{\gamma_0} \quad (2.48)$$

A função expressada acima, para uma estimativa populacional, contudo sabemos que na prática não se trabalha com estimativas populacionais, mas, com estimativas amostrais. Dessa forma, a função de autocorrelação amostral é dada por,

$$r_k = \frac{\sum_{t=k+1}^n (w_t - \bar{w})(w_{t-k} - \bar{w})}{\sum_{t=1}^n (w_t - \bar{w})^2}, \text{ onde } n \text{ é o número de observações.} \quad (2.49)$$

Para que seja feita análise estatística acerca da função de autocorrelação amostral, devemos saber qual a distribuição de probabilidade de  $r_k$ . Então se  $n$  for grande suficientemente, e se  $\rho_k = 0$  para  $k \neq 0$ , a variância de  $r_k$  é aproximadamente igual a  $\frac{1}{n}$ , sua distribuição é  $r_k \sim N(0, \frac{1}{n})$ . Encontrada a distribuição de  $r_k$ , podemos então construir seus intervalos de confiança e usar os testes de hipóteses com o intuito de verificar a nulidade dos coeficientes de autocorrelação. Por meio do teste de *Ljung-Box* ver Vasconcellos e Alves (2000), é possível testar as hipóteses de que os  $k$  primeiros coeficientes de autocorrelação são conjuntamente iguais a zero, usando a seguinte estatística:

$$Q(K) = n(n+2) \sum_{k=1}^K \frac{r_k^2}{n-k}, \quad (2.50)$$

,onde tal estatística segue uma distribuição  $\chi^2$  com  $k$  graus de liberdade.

### 2.5.8 Função de autocorrelação parcial (FACP)

Também utilizada para identificar os modelos para uma determinada série temporal, assim como auxilia na identificação de estacionariedade da série, a função de autocorrelação parcial, denotada por  $\phi_{kk}$ , mensura a correlação entre pontos da série  $(w_t e w_{t-k})$ , descontando a influência entre os pontos intermediários, ou seja, influência de  $w_{t-1}$ ,  $w_{t-2}$ , ...,  $w_{t-k+1}$  sobre  $w_t$ . Mostrando assim, que de cada modelo AR estes são seus últimos coeficientes.

$$\left\{ \begin{array}{l} w_t = \phi_{11}w_{t-1} + u_t \\ w_t = \phi_{11}w_{t-1} + \phi_{22}w_{t-2} + u_t \\ \vdots \\ w_t = \phi_{k1}w_{t-1} + \phi_{k2}w_{t-2} + \dots + \phi_{kk}w_{t-k} + u_t \end{array} \right. \quad (2.51)$$

Outra maneira de estimar os valores de  $\phi_{kk}$ , é por meio do sistemas de equações de *Yule-Walker*, que é dado por

$$\left\{ \begin{array}{l} r_1 = \phi_{11} + \phi_{22}r_1 + \dots + \phi_{kk}r_{k-1} \\ r_2 = \phi_{11}r_1 + \phi_{22} + \dots + \phi_{kk}r_{k-2} \\ \vdots \\ r_k = \phi_{11}r_{k-1} + \phi_{22}r_{k-2} + \dots + \phi_{kk} \end{array} \right. \quad (2.52)$$

Da mesma maneira que na função de autocorrelação, para construção dos intervalos de confiança e realização dos testes de hipótese de  $\hat{\phi}_{kk}$ , devemos ter conhecimento de sua

distribuição de probabilidade. Para um  $n$  grande suficientemente,  $\phi_{kk}$  apresenta distribuição  $N(0, \frac{1}{n})$ , para  $k > p$ .

### 2.5.9 Etapas do modelo SARIMA

Tomada a decisão de utilizar o modelo SARIMA e seus possíveis derivados para aplicação da metodologia de Box e Jenkins, devemos passar por seu três estágios: identificação, estimação e diagnóstico. Na fase de identificação, são estabelecidos os parâmetros do modelo, isto é,  $(p, d, q) \times (P, D, Q)$ , que é uma fase crucial, pois requer um certo nível de conhecimento teórico. O segundo passo é a estimação, onde são fornecidas as estimativas dos parâmetros. O último passo é o diagnóstico, onde é verificado se os resultados encontrados para o modelo são satisfatórios, se estes resultados não são adequados (satisfatórios), devemos retornar a fase de identificação, com o intuito de encontrar um modelo que realmente seja adequado.

Ao passar por todas essas etapas citadas acima com sucesso, podemos utilizar o modelo para fazer previsões para a série  $Z_t$ .

#### Identificação

A fase de identificação compreende um conjunto de procedimentos a fim de proporcionar uma ideia aproximada da estrutura do modelo. Esta fase é bem complexa e crucial, devido a dificuldade no momento de determinar os parâmetros  $p$ ,  $d$  e  $q$ , pois se os mesmos forem mal determinados, o modelo que se adotará será inadequado, gerando assim resultados desastrosos.

Na determinação dos parâmetros, inicia-se determinando o parâmetro de integração (parâmetro  $d$ ), por ser o mais simples, pois podemos aplicar algum teste de estacionariedade com intuito de encontrar a ordem de integração. Tal parâmetro pode assumir os valores 0, 1 e 2, onde normalmente  $d = 0$  ou  $d = 1$  e raramente  $d = 2$ .

Quando a série é estacionária,  $d = 0$ , porém quando  $d = 1$  a série não é estacionária, ou seja, possui uma raiz unitária. Esses são os casos mais simples encontrados. Caso a série não seja estacionária, com  $d = 1$  podemos testar se  $Z_t$  tem raiz unitária utilizando ver Fuller (2009), mas quando a ordem de integração  $d > 1$  utilizamos ver Dickey e Pantula (1987).

#### Teste de Estacionariedade de Dickey-Fuller

O teste de raiz unitária de Dickey-Fuller, basicamente, estima a seguinte auto-regressão:

$$\nabla Z_t = (\rho - 1)Z_{t-1} + \varepsilon_t \quad (2.53)$$

ou então,

$$\nabla Z_t = \gamma Z_{t-1} + \varepsilon_t \quad (2.54)$$

onde  $\nabla = (Z_t - Z_{t-1})$  é o operador de diferença e  $\gamma = \rho - 1$ .

Para este caso, temos as seguintes hipóteses:

- $H_0: \gamma = 0$ , existe pelo menos uma raiz unitária, e sua variável não é estacionária;
- $H_1: \gamma < 0$ , não a raiz unitária, e sua variável seja fracamente estacionária.

Mas para o nosso estudo, vamos utilizar uma expressão alternativa do teste de Dickey-Fuller, que expressa da seguinte maneira,

$$Z_t = \rho Z_{t-1} + \varepsilon_t, \text{ onde } \varepsilon_t \text{ é o ruído branco.} \quad (2.55)$$

Para esse caso, se  $\rho = 1$  a série possui uma raiz unitária e não é estacionária, porém se  $|\rho| < 1$  a série não possui raiz unitária e é estacionária. Porém, para testar se a série apresenta ou não raiz unitária ver Dickey e Fuller (1979) propõe um processo gerador da série que é dado por um dos três modelos dados abaixo:

1.  $\Delta Z_t = \alpha + \beta_t + \lambda_3 Z_{t-1} + \varepsilon_t$
2.  $\Delta Z_t = \alpha + \lambda_2 Z_{t-1} + \varepsilon_t$
3.  $\Delta Z_t = \lambda_1 Z_{t-1} + \varepsilon_t$

onde  $\lambda_i = \rho - 1$ ,  $\forall_i = 1, 2, 3$  e  $\alpha$  e  $\beta$  sendo as constantes que se quer estimar.

As hipóteses a serem testadas são apresentadas abaixo, onde  $H_0 : \rho = 1$  equivale a  $H_0 : \lambda = 0$ .

Logo, se ao menos umas das hipóteses apresentadas na Tabela 1 não for rejeitada, a série possui pelo menos uma raiz unitária assim sendo não estacionária.

As seguintes estatísticas  $\tau_1, \tau_2, \tau_3, \delta_1, \delta_2$  e  $\delta_3$  presentes na Tabela 1, são encontradas das seguintes formas:

$$\tau_3 = \frac{\lambda_3}{\sigma \lambda_3}$$

$$\delta_3 = \frac{SQR(1) - SQR(1)}{\frac{3SQR(1)}{n}}$$

Tabela 1 – Testes da raiz unitária de Dickey-Fuller

Modelos	$H_0$	Regras de decisão
1	$\lambda_3 = 0$	$\tau_3 > \text{valor crítico} \Rightarrow H_0$ não é rejeitada
	$(\alpha, \beta, \lambda_3) = (0,0,0)$	$\delta_2 < \text{valor crítico} \Rightarrow H_0$ não é rejeitada
	$(\alpha, \beta, \lambda_3) = (\alpha,0,0)$	$\delta_3 < \text{valor crítico} \Rightarrow H_0$ não é rejeitada
2	$\lambda_2 = 0$	$\tau_2 > \text{valor crítico} \Rightarrow H_0$ não é rejeitada
	$(\alpha, \lambda_2) = (0,0)$	$\delta_1 < \text{valor crítico} \Rightarrow H_0$ não é rejeitada
3	$\lambda_1 = 0$	$\tau_1 > \text{valor crítico} \Rightarrow H_0$ não é rejeitada

Fonte: Adaptada de Dickey e Fuller (1979)

$$\tau_2 = \frac{\lambda_2}{\sigma\lambda_2}$$

$$\delta_2 = \frac{SQR(2) - SQR(2)}{\frac{3SQR(2)}{n}}$$

$$\tau_1 = \frac{\lambda_1}{\sigma\lambda_1}$$

$$\delta_1 = \frac{SQR(3) - SQR(3)}{\frac{3SQR(3)}{n}}$$

onde  $SQR(1)$  é a soma de quadrado do resíduo do modelo  $\Delta Z_t = \alpha + \beta_t + \lambda_3 Z_{t-1} + \varepsilon_t$ ,  $SQR(2)$  é a soma de quadrado do resíduo do modelo  $\Delta Z_t = \alpha + \lambda_2 Z_{t-1} + \varepsilon_t$ ,  $SQR(3)$  é a soma de quadrado do resíduo do modelo  $\Delta Z_t = \lambda_1 Z_{t-1} + \varepsilon_t$ , já  $\sigma\lambda_1$  é a variância de  $\lambda_1$ ,  $\sigma\lambda_2$  é a variância de  $\lambda_2$  e  $\sigma\lambda_3$  é a variância de  $\lambda_3$ .

Ver Fuller (2009) para verificar os valores críticos destes testes, que também estão implementados em alguns softwares estatísticos.

### Análise Gráfica

Feita a série estacionária, ou seja, encontrado o valor do parâmetro  $d$ , o passo seguinte é encontrar os parâmetros  $p$  e  $q$ , para que a etapa de identificação seja concluída. De acordo com Morettin e Toloi (2006) esta escolha é feita com base na análise gráfica das funções de autocorrelação (FAC) e autocorrelação parcial (FACP).

Um processo  $AR(p)$  tem FAC infinita com decaimentos exponenciais e/ou senoides amortecidas, ou seja, tende a zero à medida que  $k$  aumenta, já a FACP é diferente de zero quando  $k \leq p$ , e igual a zero quando  $k > p$ . Um processo  $MA(q)$  tem FAC finita, no sentido que ela apresenta um corte após a defasagem  $q$ , a FACP se comporta de maneira similar à FAC de um processo  $AR(p)$ , isto é, composta por exponenciais e/ou senoides amortecidas. Ao combinarmos as componentes dos modelos  $AR$  e  $MA$ , formamos um processo  $ARMA(p, q)$ , que tem uma FAC infinita que decai de acordo com as exponencias



e/ou senoides amortecidas após a defasagem ( $q - p$ ), e uma FACP que se comporta como a FACP de um processo MA puro.

## Estimação

Identificado um modelo provisório para a série, o passo seguinte é estimar seus parâmetros, onde se pode usar o procedimento iterativo de Mínimos Quadrados Condicionais ou o Método de Máxima Verossimilhança (MORETTIN; TOLOI, 2006).

Neste estudo será utilizado o método de Máxima Verossimilhança pelo fato dele maximizar a verossimilhança de um conjunto de observações, e também devido a aplicabilidade do software usado apresentar cálculos bem implementados das estimativas dos parâmetros.

Considerando um modelo ARIMA( $p, d, q$ ) e colocando seus  $p + q + 1$  parâmetros no vetor  $\xi = (\phi_i, \theta_j, \sigma_\xi^2)$ , onde  $\rho_i = (\phi_1, \dots, \phi_p)$ ,  $\theta_j = (\theta_1, \dots, \theta_q)$ . Para estimar  $\xi$ , será usado o método de Máxima Verossimilhança: dadas as  $N$  observações  $Z_1, \dots, Z_N$ , considera-se a função de verossimilhança  $L = (\xi | Z_1, \dots, Z_N)$  como função de  $\xi$ . Além disso, deve-se supor que  $\xi$  (ruído branco) tenha distribuição normal (MORETTIN; TOLOI, 2006). Porém, sabemos que nossa modelagem é do tipo SARIMAX, então consideramos que o ruído branco tem distribuição normal, e  $\mathbf{w} = (w_1, w_2, \dots, w_n)$  apresentará distribuição normal conjunta com o vetor de médias nulo e matriz de covariância  $\Sigma$ .

A função de máxima verossimilhança é dada por Brockwell e Davis (2002) (adaptada)

$$L(\phi_i, \theta_j, \sigma_\xi^2/w) = (2\pi\sigma_\xi^2)^{-\frac{n}{2}} \left| \Sigma \right|^{-\frac{1}{2}} e^{-\frac{1}{2} \mathbf{x}' \Sigma^{-1} \mathbf{w}} \quad (2.56)$$

Ao maximizar esta função obtemos as estimativas dos parâmetros  $\phi_i$ ,  $\theta_j$  e  $\sigma_\xi^2$ .

## Validação

Tendo o modelo sido identificado e os parâmetros estimados, as verificações de diagnóstico são então aplicadas ao modelo ajustado. Um método útil de verificar um modelo é superestimá-lo, ou seja, estimar os parâmetros num modelo um pouco mais geral do qual acreditamos ser verdadeiro. Este método pressupõe que podemos adivinhar a direção em que o modelo é provável que seja inadequado. Portanto, é necessário complementar essa abordagem por meio de verificações menos específicas aplicadas aos resíduos do modelo ajustado (BOX; JENKINS; REINSEL, 2008), ou seja, a etapa de validação é dividida em análise residual e avaliação de quantos parâmetros são adequados (ordem do modelo) (FAVA, 2000).

Na fase da análise residual é verificado se os resíduos do modelo "adotado" apresentam comportamento de um ruído branco. Com isto, os coeficientes de autocorrelação residual devem ser estatisticamente iguais a zero. Para testar esta condição, usamos do teste de *Ljung-Box* (FAVA, 2000).

A fase de avaliação da quantidade de parâmetros adequados, refere-se ao atendimento do critério de parcimônia do modelo (FAVA, 2000), onde os modelos podem ser parcimoniosos ou não. Quando um modelo apresenta poucos parâmetros que possivelmente não são significativamente informativos, este modelo é parcimonioso, já um modelo não parcimonioso apresenta mais coeficientes, porém alguns desses coeficientes podem ser não-significativos ou correlacionados. Por não ter um teste ou metodologia que nos diga quantos parâmetros são necessários para tal modelo, recorreremos ao auxílio de softwares estatísticos que realizam análises extensas, onde o modelo é estimado com diversas combinações de  $p$  e  $q$ . Então para escolha do melhor modelo dentre os apresentados, e descarte daqueles que são não significativos, existem alguns critérios que podem auxiliar, ver Brockwell e Davis (2002):

- Final Prediction Error (FPE)

$$FPE = \tilde{\sigma}^2 x \frac{T+p}{T-p} \quad (2.57)$$

- Akaike Information Criterion (AIC)

$$AIC = 2\log\hat{L} + 2(p+q) \quad (2.58)$$

onde  $\hat{L}$  é o valor maximizado da verossimilhança.

- Corrected Akaike Information Criterion (AICC)

$$AICC = -2\log\hat{L} + \frac{2(p+q)T}{T-p-q-1} \quad (2.59)$$

- Bayesian Information Criterion

$$BIC = -2\log\hat{L} + (p+q)\log T \quad (2.60)$$

Devemos levar em consideração que tanto o AIC quanto o AICC não são consistentes (não apresentam valores bem concretos), já que os modelos apresentados pelos mesmos são superparametrizados assintoticamente. O BIC em contrapartida é consistente e rigoroso na inclusão de parâmetros do modelo.

## 2.6 Avaliação de Previsão

Em séries temporais estamos interessados em prever valores, ou seja,  $Z_{t+h}$ , com  $h \geq 1$ , supondo que temos observações, ...,  $Z_{t-2}$ ,  $Z_{t-1}$ ,  $Z_t$ , até o instante  $t$ , que chamamos de origem de previsão, podemos prever os valores futuros da série  $Z$  no instante  $t + h$ , com base nas  $T$  observações anteriores. A previsão para os instantes futuros da série, ou seja,  $t + h$ , é definida pela esperança condicional de  $Z_{t+h}$ , com base em todos os valores passados ...,  $Z_{t-2}$ ,  $Z_{t-1}$ ,  $Z_t$  (MORETTIN; TOLOI, 2006), veja abaixo:

$$\hat{Z}_t(h) = E[Z_{t+h}] = E[Z_{t+h}|Z_t, Z_{t-1}, \dots] \quad (2.61)$$

O erro de previsão com  $h$  passos-a-frente, é definido como:

$$e_t(h) = Z_{t+h} - \hat{Z}_t(h) \quad (2.62)$$

onde  $\hat{Z}_t(h)$  é a previsão de  $Z_{t+h}$ , com  $h$  passos-a-frente.

Uma notação alternativa pode ser feita para o erro de previsão. Se uma série temporal apresentar  $T$  observações e dela são retiradas as  $n$  últimas observações, podemos gerar  $n$  previsões para essa série. Então seu erro de previsão será,

$$e_j = Z_j - \hat{Z}_j \quad (2.63)$$

onde  $\hat{Z}_j$  é a previsão de  $Z_j$ . Esta notação apresentada, serve para melhor compreensão das expressões que serão ditas a seguir.

É possível avaliar a capacidade de previsão do modelo estimado por algumas medidas apresentadas a seguir, onde  $k = T - n$ .

### 1. Erro Percentual Médio (MPE)

$$MPE = \left( \frac{1}{n} \sum_{j=k+1}^T \frac{e_j}{Z_j} \right) \times 100 \quad (2.64)$$

### 2. Erro Quadrático Médio (MSE)

$$MSE = \frac{1}{n} \sum_{j=k+1}^T (e_j)^2 \quad (2.65)$$

### 3. Erro Absoluto Médio (MAE)

$$MAE = \frac{1}{n} \sum_{j=k+1}^T |e_j| \quad (2.66)$$

4. Erro Total (TE)

$$TE = \sum_{j=k+1}^T e_j \quad (2.67)$$

5. Erro Absoluto Total (TAE)

$$TAE = \sum_{j=k+1}^T |e_j| \quad (2.68)$$

6. Erro Percentual Total (TPE)

$$TPE = \left( \frac{\sum_{j=k+1}^T e_j}{\sum_{j=k+1}^T Z_j} \right) \times 100 \quad (2.69)$$

Onde para seu estudo, qualquer erro pode ser utilizado, porém em séries temporais o melhor erro de se verificar é o TPE.

---

### 3 Aplicação

Este estudo tem por finalidade fazer previsões avaliando metodologias utilizadas no processo de séries temporais, tendo o propósito de prever o valor total de arrecadação do ICMS da Paraíba até Dezembro de 2016 tendo como base dados no período de janeiro de 2003 até agosto de 2016. Para análise dos dados utilizaremos do software R (R Core Team, 2016).

Inicialmente na análise devemos fazer a leitura dos dados da arrecadação do ICMS da Paraíba e também do IGP-DI, usado para o deflacionamento. Após leitura dos dados, a fim de evitar uma série com dinâmica inflacionária, dividimos os valores do ICMS pelos valores do IGP-DI.

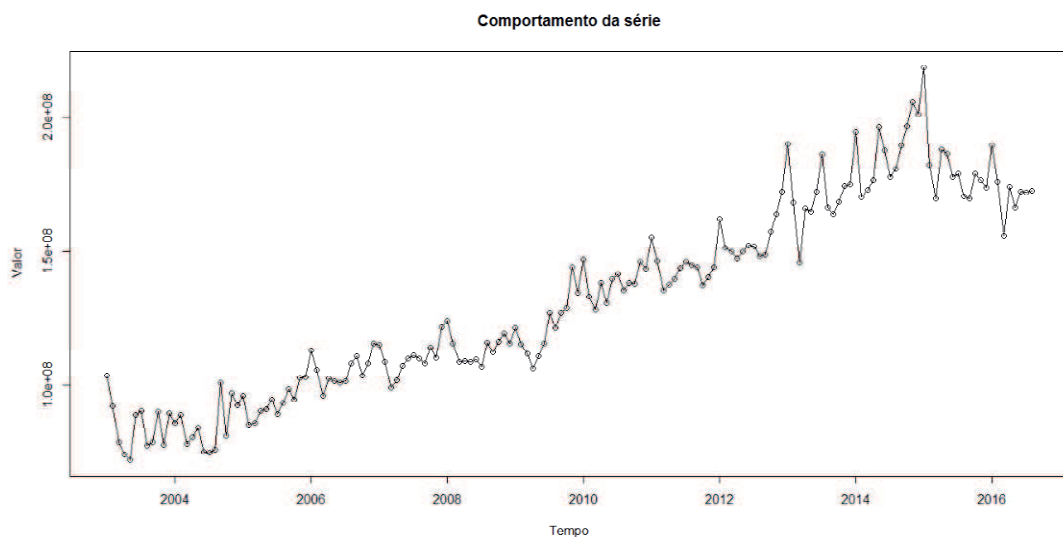


Figura 1 – ICMS na Paraíba entre janeiro de 2003 até agosto de 2016.

Na Figura 1 observa-se que a série apresenta indícios de sazonalidade, tendência crescente e não-estacionariedade.

Ao aplicarmos o comando *summary* nos dados, percebemos que o valor médio arrecadado foi de 214,812 milhões de reais, e sua arrecadação mediana de 186,713 milhões de reais, já seus valores extremos foram de 73,622 milhões (valor mínimo) e 436,477 milhões de reais (valor máximo). Referente a taxa de crescimento média mensal, observamos que essa foi de aproximadamente 0.31%, e seus taxas extremas foram de -21.72% (valor mín.) e 28.16%(valor máx.).

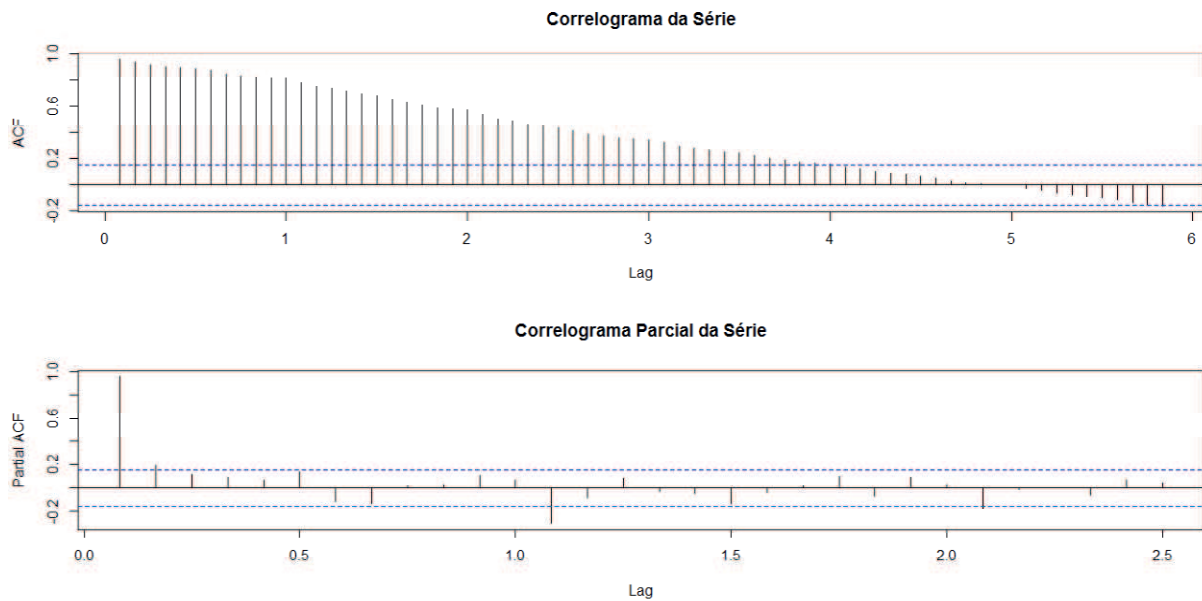


Figura 2 – Gráficos do Correlograma e Correlograma Parcial dos dados reais.

Na Figura 2, vemos que o ACF apresenta um decaimento lento para zero evidenciando a não-estacionariedade da série. Já no PACF temos picos que ultrapassam o limite estipulado, picos estes presentes praticamente nos lags 0.0 e 1.0, o que indica correlação entre os meses indicados.

Aplicou-se o logaritmo nos dados e novamente, para verificar a estacionariedade da série, os devidos correlogramas porém aplicando-se a primeira diferença ( $d = 1$ ). Podemos observar na Figura 3, que o gráfico ACF e PACF apresentam alguns picos que ultrapassam o limite de -0.2 e 0.2. Entretanto, no ACF da primeira diferença os picos vão se repetindo nos lags 1.0, 2.0, 3.0 e assim sucessivamente, sugerindo uma forte sazonalidade na série.

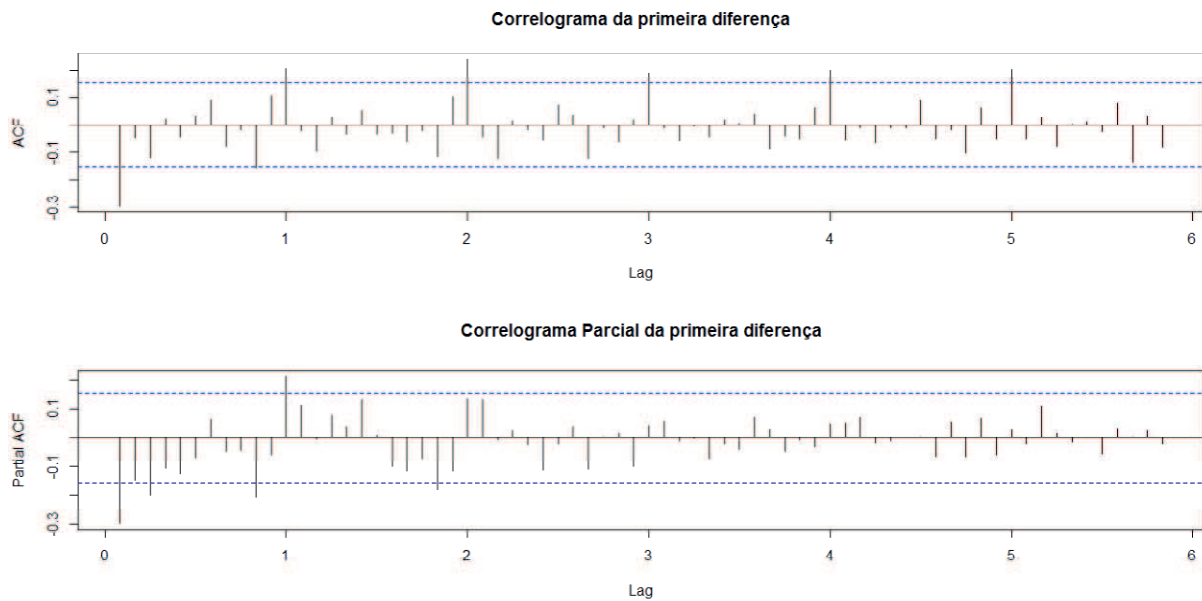


Figura 3 – Gráficos do Correlograma e Correlograma Parcial da primeira diferença dos dados logaritmados.

Como objetivo principal é a previsão, então retirou-se o último valor da série e o armazenando-o em outro objeto, com o intuito de prever tal valor que foi retirado. Assim, com a retirada desse último dado, formatou-se os dados pela função `ts`, a fim de aplicar as devidas metodologias para a previsão dos dados.

### 3.1 Alisamento Exponencial

A metodologia inicial usada foi a classe de algoritmos de alisamento exponencial. Observado uma presença de sazonalidade nos dados, então deve ser utilizado nessa etapa o algoritmo de Holt-Winters. Neste momento pode-se utilizar funções `HoltWinters` ou `hw`. A fim de manter a escala original da série, é preciso modelar a série logaritmada e suas devidas previsões serão exponenciadas.

Um ponto importante para a previsão deste método é a escolha de seus parâmetros de suavização, pois esses parâmetros ajudam a minimizar a soma dos quadrados dos erros de previsão a um passo a frente.

Tabela 2 – Valores otimizados das constantes de suavização para Holt-Winters aditivo e Multiplicativo

Método	$\alpha$	$\beta$	$\gamma$
HW aditivo	0,2996351	0,03042671	0,5151928
HW multiplicativo	0,2995106	0,03003409	0,5157786

Por meio da predição dos algoritmos aditivo e multiplicativo de Holt-Winters

encontrou-se um valor de previsão de um pouco mais que 385,6 milhões de reais, e seus erros foram de 3.87%. A essa previsão trouxe valores abaixo do valor observado, ou seja, uma subprevisão. Verificou-se que os fatores sazonais estimados indicam que o mês de junho é que apresenta a maior arrecadação, diferente do mês de agosto que arrecadou menos.

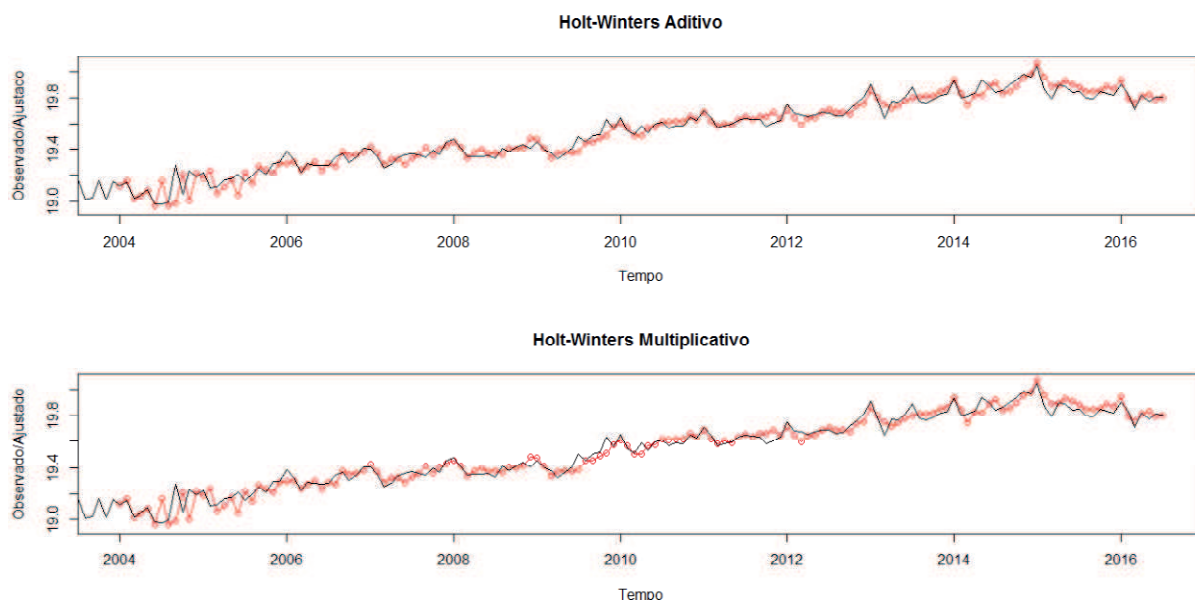


Figura 4 – Gráficos dos dados da série e seu ajuste de Holt-Winters.

Nota-se acima os dados da série pela linha preta, e seu ajuste por Holt-Winters de vermelho, e percebemos que de certa forma os dados previstos ficaram bem ajustados aos dados observados.

Utilizando os mesmos métodos e modelos, realizou-se uma previsão da série com 5 passos a frente, ou seja, até dezembro de 2016.

Tabela 3 – Previsões adquiridas pelos algoritmos Holt-Winters aditivo e Multiplicativo

Mês/2016	HW aditivo	HW multiplicativo
Agosto	385.629.826	385.635.921
Setembro	389.993.766	389.974.898
Outubro	404.844.598	404.812.535
Novembro	411.758.616	411.745.886
Dezembro	410.564.657	410.614.120

Na Tabela 3 destaca-se as previsões com 5 passos a frente, dos modelos Holt-Winters aditivo e multiplicativo descritos acima, e ao compararmos os dois modelos vemos que praticamente as previsões são as mesmas. Observando os resultados das previsões, nota-se que até setembro houve uma subprevisão, já nos meses seguintes há uma superprevisão.



O passo seguinte, utilizaremos a função `ts` presente no pacote (`forecast`), que disponibiliza ferramentas importantes para modelagem e previsão de séries temporais. Onde utilizamos um fator de atenuação de tendência no segundo procedimento do algoritmo.

A previsão obtida aqui, usando atenuação de tendência, foi de aproximadamente 388,3 milhões de reais, com um erro relativo de 3.21%, ou seja, uma subprevisão dos dados com um erro relativo um pouco menor. Na Figura 5, temos a série dos dados suas previsões com o devido intervalo de confiança, onde o primeiro gráfico (sem atenuação) mostra uma pequena tendência decrescente no intervalo de predição, enquanto o segundo gráfico (com atenuação) apresenta uma previsão mais contínua dos dados, porém em ambos os casos percebemos sazonalidade no início da previsão de 2016 e 2017.

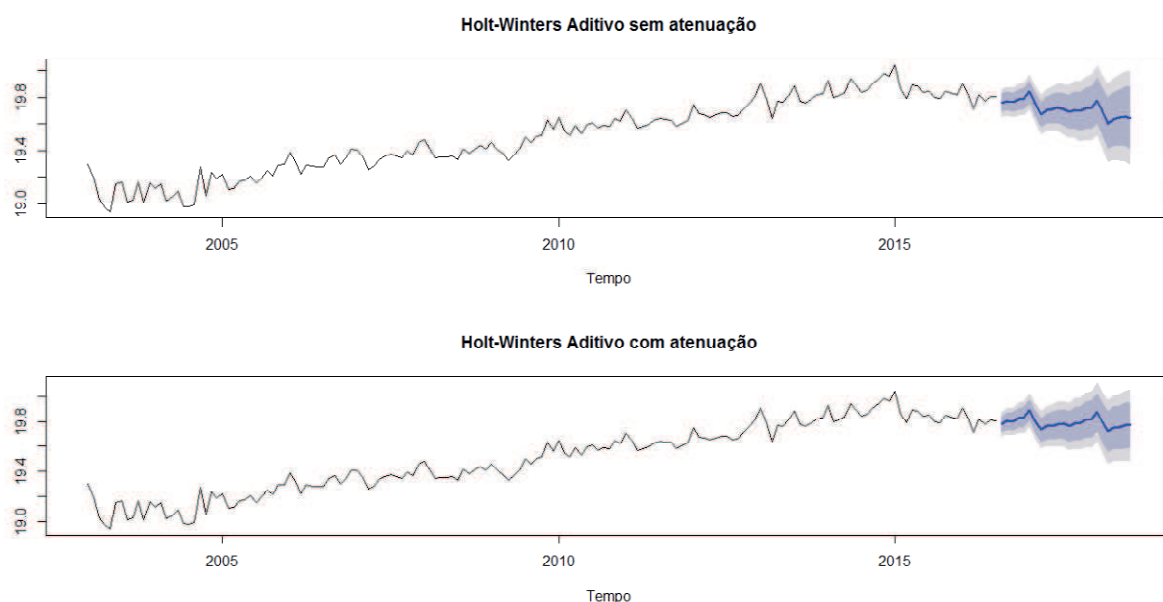


Figura 5 – Gráficos da previsão dos dados sem e com atenuação de tendência.

Existe uma vertente do algoritmo de alisamento exponencial implementada na função `ets` do pacote `forecast`; verificar Hyndman, Khandakar et al. (2007). Esta vertente faz uso de uma abordagem de modelagem inovacional de espaços de estados. Nota-se que tal aplicação de forma automática gera um erro de 3.22%; sua previsão de arrecadação para o mês de agosto de 2016 é de 388,25 milhões de reais.

Tabela 4 – Erros de previsão

ME	RMSE	MAE	MPE	MAPE	MASE
0,004535497	0,05123058	0,03884906	0,022975	0,1997951	0,4442726

## 3.2 Modelagem SARIMA

Nesta etapa a classe de modelagem SARIMA será utilizada com intuito de gerar uma previsão para arrecadação dos dados de ICMS no período de dezembro de 2016. Para encontrar o modelo mais adequado, utilizou-se a função `auto.arima` do pacote `forecast` que podem ser vistos bem a fundo em Hyndman, Khandakar et al. (2007), que apresentará o modelo com melhor capacidade preditiva.

Com uso da função `auto.arima` o modelo SARIMA(0,1,1)x(0,0,2) foi utilizado, e tal modelagem obteve uma previsão da arrecadação do ICMS de dezembro de 2016 foi pouco mais que 394,1 milhões de reais, com erro relativo de aproximadamente 1.74%. (Mais uma vez o valor predito foi abaixo do observado.)

Com base em pesquisas anteriores, também ajustaremos um modelo SARIMA (0,1,1)x(0,1,1) que é conhecido como "*modelo airline*". E essa modelagem forneceu uma previsão de arrecadação de pouco mais de 394,8 milhões de reais, com um erro relativo relativo de 1.57%.

Um importante ponto para a análise é lembrar da crise que ocorreu em âmbito mundial, que afetou a economia de todos os países. A crise dos *commodities*, veio afetar o Brasil no início de 2009, pois sempre foi um grande meio de arrecadação econômica no país, vemos tal declínio ao compararmos 2008 com 2009, onde houve uma queda de 14% nos preços do *commodities*. Sabendo de tal acontecimento, utilizou-se uma variável *dummy* com valores 0 até dezembro de 2008 e valor 1 a partir de janeiro 2009, com intuito de diferenciar esses acontecimentos bem marcantes na economia mundial.

Avaliando novamente os modelos SARIMA(0,1,1)x(0,0,2), e o "*modelo airline*", e ajustando a variável dicotômica como variável regressora, transformando na modelagem SARIMAX, e novamente as previsões são calculadas.

Obtivemos uma previsão de aproximadamente 394,2 milhões de reais, com um erro relativo de 1.73% para o modelo SARIMAX(0,1,1)x(0,0,2), e uma previsão de mais de 394,8 milhões de reais, com erro relativo de 1.58% para o "*modelo airline*". Como vemos, há uma subprevisão dos dados e uma tendência que possivelmente se dá por conta desta crise financeira. Observando tal comportamento será criada uma variável tendência para multiplicarmos com a *dummy*, para fazer um novo ajuste aos modelos já mostrados anteriormente.

O novo ajuste do "*modelo airline*" forneceu uma previsão de aproximadamente 416 milhões de reais, com erro relativo de -3.69%, já o outro modelo teve previsão de 383,8 milhões de reais, com erro relativo de 4.32%, e percebeu-se uma pequena subprevisão.

O próximo passo é juntar as duas variáveis regressoras, para confirmar que as mesmas apresentaram alterações com a crise.

O modelo SARIMAX(0,1,1)x(0,0,2) apresentou uma previsão de aproximadamente 397 milhões de reais, com erro relativo de 1.05%, que ao comparar com o valor observado foi um pouco menor, ou seja obteve uma subprevisão. Já o *modelo airline* apresentou uma subprevisão, pouco mais de 395 milhões de reais com erro relativo de 1.52%, ou seja, algo próximo do esperado.

Feitas todas as análises com a modelagem de SARIMA e SARIMAX, com 1 passo a frente, iremos refazer novamente todas essas análises com 5 passos a frente, ou seja, até dezembro de 2016.

Alguns dos critérios de escolha dos modelos são o AIC (Critério de Informação de Akaike) e BIC (Critério de Informação Bayesiano), que tem como base a soma dos quadrados dos erros (SQE) para melhor balancear o ajuste das variáveis do modelo. Sabendo disso, apresentaremos uma tabela com os valores de AIC e BIC, dos modelos selecionados.

Tabela 5 – Valores referentes ao critérios de escolha dos modelos da série

	Modelos	AIC	BIC
I	SARIMA (0,1,1)x(0,0,2)	-429.58	-417.23
II	<i>modelo airline</i>	-419.36	-408.32
III	SARIMAX (0,1,1)x(0,0,2)	-429.58	-412.14
IV	SARIMAX <i>airline</i>	-417.73	-403.68
V	SARIMAX (0,1,1)x(0,0,2) c tendência	-429.64	-412.20
VI	SARIMAX <i>airline</i> c tendência	-417.65	-403.61
VII	SARIMAX (0,1,1)x(0,0,2) com regressores	-429.15	-408.63
VIII	SARIMAX <i>airline</i> com regressores	-415.81	-398.76

Observando esse conjunto de modelos, vemos de maneira geral todos são bem semelhantes pelo fato de serem significativos e com intervalos bem próximos, porém o primeiro modelo expressado na tabela é o que apresenta menores valores dos critérios de avaliação, e que conseqüentemente tem a menor SQE. Em contrapartida o último modelo apresentado na tabela, é que tem piores valores, assim tendo um alto SQE.

Mostrados os critérios de escolha das modelagens, as devidas previsões, com 5 passos a frente, dos modelos foram calculados como pode ser verificado na Tabela 5.

Tabela 6 – Previsões com 5 passos a frente dos modelos SARIMA e SAMIRAX, Agosto - Outubro

	Modelos	Agosto/16	Setembro/16	Outubro/16
I	SARIMA (0,1,1)x(0,0,2)	394.182.878	398.437.217	404.111.438
II	<i>modelo airline</i>	394.856.782	405.565.663	408.800.322
III	SARIMAX (0,1,1)x(0,0,2)	397.594.036	401.899.224	407.605.804
IV	SARIMAX <i>airline</i>	394.794.006	405.351.355	408.743.685
V	SARIMAX (0,1,1)x(0,0,2) c tendência	383.818.487	388.008.670	393.422.421
VI	SARIMAX <i>airline</i> c tendência	415.983.976	427.256.986	430.852.820
VII	SARIMAX (0,1,1)x(0,0,2) com regressores	396.947.709	400.538.903	405.624.675
VIII	SARIMAX <i>airline</i> com regressores	395.048.138	405.134.705	408.497.995

Tabela 7 – Previsões com 5 passos a frente dos modelos SARIMA e SAMIRAX, Novembro e Dezembro

	Modelos	Novembro/16	Dezembro/16
I	SARIMA (0,1,1)x(0,0,2)	404.008.728	400.376.013
II	<i>modelo airline</i>	422.714.575	427.254.741
III	SARIMAX (0,1,1)x(0,0,2)	407.509.652	403.838.864
IV	SARIMAX <i>airline</i>	423.356.995	427.798.063
V	SARIMAX (0,1,1)x(0,0,2) c tendência	393.341.430	389.747.659
VI	SARIMAX <i>airline</i> c tendência	446.278.994	451.105.927
VII	SARIMAX (0,1,1)x(0,0,2) com regressores	404.794.790	400.493.711
VIII	SARIMAX <i>airline</i> com regressores	422.751.299	426.729.760

De maneira geral tem-se que o modelo VI é o que apresentou as maiores valores de previsão, ou seja, que apresentou um melhor arrecadação. Já o modelo V, foi aquele teve os menores valores de previsão, ou seja, o modelo que apresenta a pior arrecadação de ICMS no estado da Paraíba. Por via de regra, os modelos apresentados obtiveram previsões bem próximas, ou seja, o intervalo entre os modelos são bem pequenos.

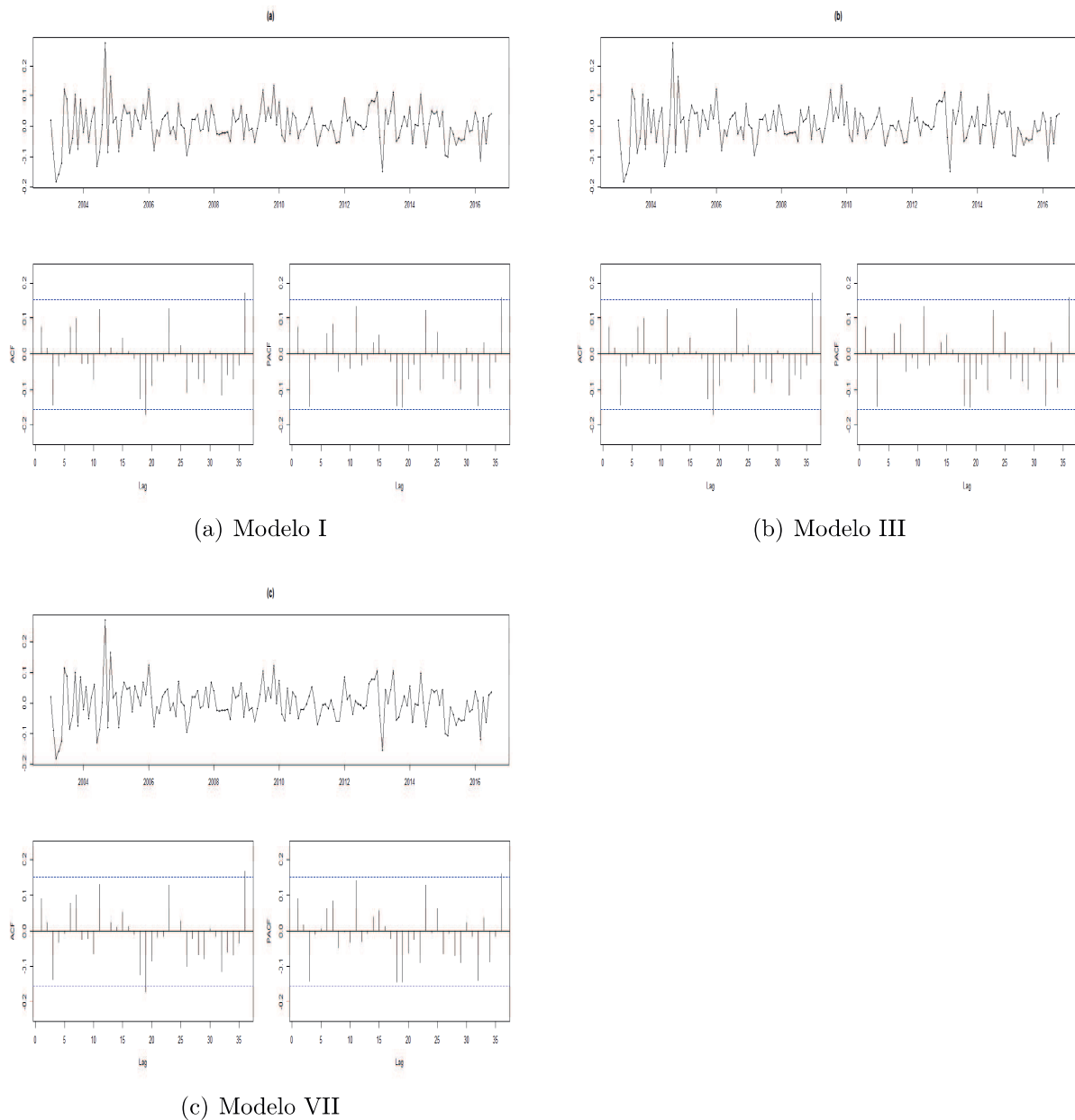
Para finalizar a análise, a seguir os erros relativos de cada modelo, para suas devidas previsões.

Tabela 8 – Erros relativos de previsão dos modelos

Modelos	Agosto/16	Setembro/16	Outubro/16	Novembro/16	Dezembro/16
I	1.74	0.68	-0.73	-0.71	0.19
II	1.57	-1.09	-1.90	-5.37	-6.50
III	0.89	-0.18	-1.61	-1.58	-0.67
IV	1.59	-1.04	-1.89	-5.53	-6.64
V	4.32	3.28	1.93	1.95	2.85
VI	-3.69	-6.50	-7.39	-11.25	-12.45
VII	1.05	0.16	-1.11	-0.90	0.17
VIII	1.52	-0.99	-1.83	-5.38	-6.37

Ao observar os erros na Tabela 8, vemos que o modelo V é o que apresenta maiores erros (positivos), o que pode aumentar os valores de sua previsão. Já o modelo VI é o que apresenta os maiores erros negativamente, o que leva suas previsões a terem com valores abaixo do que se deseja. Mas escolheu-se o modelo I como melhor modelo de previsão, pois ele apresentou o melhor critério BIC (pois é o critério mais consistente), os melhores erros relativos de previsão, e como os demais modelos forneceu boas previsões.

Figura 6 – Gráficos de diagnóstico dos resíduos dos modelos I,III e VII para a série de arrecadação de ICMS na Paraíba



Na Figura 6 é visto pelos gráficos de resíduos que eles não apresentam valores altos, assim evidenciando que os modelos são bem especificados para representar os dados.

Também verificou-se o teste adf, ou seja, o teste de estacionaridade de Dickey-Fuller, onde seu p-valor foi maior que 0.05 assim não rejeitando a hipótese nula, isto é, a série possui ao menos uma raiz unitária e é não estacionária. Porém aplicado a primeira diferença, reverteu-se essa situação, tornando nossa série estacionária.

## 4 Conclusão

Neste trabalho avaliou-se algumas diferentes modelagens para a arrecadação de ICMS na Paraíba focando-se nas suas respectivas previsões. Os algoritmos de Holt-Winters aditivo e multiplicativo juntamente com a modelagem SARIMA e SARIMAX foram utilizados. Na modelagem SARIMAX foi aplicada uma variável explicativa *dummy* assumindo o valor 0 (zero) até dezembro de 2008 antes da explosão da crise e valor 1 (um) aos dados posteriores a este período.

Previsões com 1 e 5 passos a frente foram calculadas, sendo avaliados por suas somas dos quadrados dos erros. Para a modelagem da série utilizou-se o método de Box-Jenkins, obtendo assim o modelo SARIMA (0,1,1) x (0,0,2), mas também ajustamos o "*modelo airline*" que foi obtido e usado em estudos anteriores.

Ao proceder as análises e verificar os resultados, as previsões do algoritmo de Holt-Winters e da modelagem SARIMA foram bem similares; entretanto, a modelagem SARIMA é mais satisfatória devido ao fato de terem apresentado erros menores. A modelagem SARIMAX é uma adaptação do modelo SARIMA. A utilização da modelagem SARIMAX apresentou previsões e erros similares aos do modelo SARIMA. Tal fato indica a influência da crise na arrecadação do ICMS, ou seja, uma certa presença de tendência. Para estender e verificar melhor tal situação, a variável tendência foi adicionada ao modelo SARIMAX, e em outro a interação da *dummy* e da tendência (regressores).

De maneira geral todas as modelagens apresentaram previsões bem próximas umas das outras, mas o modelo com o melhor ajuste foi o SARIMA (0,1,1) x (0,0,2) por conta de critérios como BIC e erros relativos bem pequenos. Assim observou-se que a crise afetou sim a arrecadação o ICMS no Estado, pois os valores de arrecadação poderiam ser bem maiores do que se foi arrecadado.

# Referências

BOX, G.; JENKINS, G.; REINSEL, G. *Time Series Analysis: Forecasting and Control 4ª ed.* [S.l.]: NJ: John Wiley & Sons, Inc., 2008. Citado 8 vezes nas páginas 15, 16, 21, 22, 23, 24, 25 e 31.

BRASIL. Lei complementar nº 101, de 4 de maio de 2000. estabelece normas de finanças públicas voltadas para a responsabilidade na gestão fiscal e dá outras providências. *Diário Oficial da União*, 2000. Citado na página 13.

BROCKWELL, P.; DAVIS, R. *Introduction to time series and forecasting*. Springer, New York., Sections 3, 3 and 8, 2002. Citado 4 vezes nas páginas 17, 20, 31 e 32.

CHATFIELD, C. *The analysis of time series: An introduction*. [S.l.]: Chapman & Hall, Londyn, 2004. Citado na página 18.

CRUZ, C. Ferreira da; SENHORAS, E. M. Avaliação da responsabilidade na gestão fiscal com base nos pilares da lei de responsabilidade fiscal: um estudo com municípios de santa catarina. *Cadernos de Finanças Públicas*, n. 13, 2013. Citado na página 13.

DICKEY, D. A.; FULLER, W. A. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, Taylor & Francis, v. 74, n. 366a, p. 427–431, 1979. Citado 2 vezes nas páginas 29 e 30.

DICKEY, D. A.; PANTULA, S. G. Determining the order of differencing in autoregressive processes. *Journal of Business & Economic Statistics*, Taylor & Francis Group, v. 5, n. 4, p. 455–461, 1987. Citado na página 28.

EHLERS, R. Análise de séries temporais. *Departamento de Estatística, Universidade Federal do Paraná, Curitiba.*, 2005. Disponível em: <<http://www.each.usp.br/rvicente/AnaliseDeSeriesTemporais.pdf>>. Acesso em: 07 de Agosto de 2016. Citado na página 18.

ENDERS, W. Applied econometric time series, by walter. *Technometrics*, v. 46, n. 2, p. 264, 2004. Citado na página 26.

FAVA, V. L. Análise de séries de tempo. *Vasconcellos, MAS; Alves, D. Manual de econometria: nível intermediário. São Paulo: Ed. Atlas*, 2000. Citado 2 vezes nas páginas 31 e 32.

FULLER, W. A. *Introduction to statistical time series*. [S.l.]: John Wiley & Sons, 2009. v. 428. Citado 2 vezes nas páginas 28 e 30.

HYNDMAN, R. J.; KHANDAKAR, Y. et al. *Automatic time series for forecasting: the forecast package for R*. [S.l.]: Monash University, Department of Econometrics and Business Statistics, 2007. Citado 2 vezes nas páginas 39 e 40.

IPEADATA. *IGP-DI Mensal*. Disponível em: <<http://ipeadata.gov.br>>. Acesso em: 27 de Dezembro de 2016. Citado na página 14.



- IPEADATA. *Imposto sobre a circulação de mercadorias (ICMS) - Paraíba (PB)*. Disponível em: <<http://ipeadata.gov.br>>. Acesso em: 26 de Dezembro de 2016. Citado na página 13.
- MENDONÇA, M. J. C. de; MEDRANO, L. A. *Um modelo de combinação de previsões para arrecadação de receita tributária no Brasil*. [S.l.], 2016. Citado na página 13.
- MONTGOMERY, D. C.; JENNINGS, C. L.; KULAHCI, M. *Introduction to time series analysis and forecasting*. [S.l.]: John Wiley & Sons, 2015. Citado 2 vezes nas páginas 19 e 26.
- MORETTIN, P. A.; TOLOI, C. *Análise de séries temporais*. [S.l.]: São Paulo: Editora Edgar Blucher, 2006. Citado 14 vezes nas páginas 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 31 e 33.
- PELLEGRINI, F. R. Metodologia para implementação de sistemas de previsão de demanda. *Porto Alegre: UFRGS*, 2000. Citado na página 26.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2016. Disponível em: <<https://www.R-project.org/>>. Citado na página 35.
- REZENDE, F. Icms: Como era, o que mudou ao longo do tempo, perspectivas e novas mudanças. *Cadernos Fórum Fiscal*, v. 10, p. 1–50, 2009. Citado na página 13.
- RUBIN, I. *The Politics of Public Budgeting*. [S.l.]: Washington: Ed. CQPress, 2010. Citado na página 13.
- 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. Citado na página 13.
- VASCONCELLOS, M. A. S.; ALVES, D. Manual de econometria. *São Paulo: Atlas*, 2000. Citado na página 27.
- WORDBANK. Pefa - public expenditure and financial accountability. 2000. Disponível em: <<http://www.pefa.org/en/content/pefa-framework-material-1>>. Acesso em: 15 de Novembro de 2016. Citado na página 13.

# Apêndices

# APÊNDICE A – Conjunto de Dados

Tabela 9 – Conjunto de Dado (2003 - 2009)

Mês	Ano						
	2003	2004	2005	2006	2007	2008	2009
Jan	103,697	91,239	113,262	135,168	142,683	166,970	175,841
Fev	93,940	95,364	101,333	126,127	135,087	156,500	166,273
Mar	81,132	84,503	102,860	114,336	123,491	148,249	160,712
Abr	76,426	88,479	108,422	122,308	126,805	151,022	153,035
Mai	73,622	93,547	108,762	121,891	133,872	153,610	159,265
Jun	90,200	84,602	112,537	121,705	137,732	157,258	165,586
Jul	91,887	85,433	105,598	122,473	140,556	153,882	181,051
Ago	79,674	87,600	110,038	130,799	140,869	166,816	173,441
Set	81,442	116,682	116,386	134,919	139,748	162,633	181,888
Out	93,704	94,535	112,253	127,033	148,912	169,267	184,186
Nove	81,182	113,674	122,134	133,197	145,717	173,333	206,296
Dez	94,048	108,889	122,979	142,830	162,441	167,814	193,335

Obs: Valores por milhão

Tabela 10 – Conjunto de Dado (2010 - 2016)

Mês	Ano						
	2010	2011	2012	2013	2014	2015	2016
Jan	213,516	250,930	272,083	344,652	373,911	436,477	423,011
Fev	194,584	238,269	254,632	305,780	331,353	367,506	395,467
Mar	189,241	221,736	254,442	265,502	339,589	345,780	351,966
Abr	205,935	225,461	252,115	303,060	346,514	385,136	395,893
Mai	196,472	228,618	259,076	302,501	383,408	384,408	383,122
Jun	210,812	235,638	265,979	317,798	364,374	368,959	399,775
Jul	214,806	239,816	268,875	343,837	344,151	372,812	398,944
Ago	207,785	239,664	264,921	310,442	350,459	359,070	401,167
Set	214,308	239,694	266,930	309,045	367,946	362,583	
Out	216,763	229,196	282,202	318,464	385,225	388,020	
Nove	232,197	234,452	295,415	331,789	405,870	385,673	
Dez	229,339	241,150	312,076	334,627	399,130	383,866	

Obs: Valores por milhão

# APÊNDICE B – Previsões Intervalares de Holt-Winters

Tabela 11 – Previsões intervalares do algoritmo Holt-Winters sem atenuação de tendência

Mês/Ano	Ponto previsão	Inferior 80%	Superior 80%	Inferior 95%	Superior 95%
Agosto/16	19.75749	19.68877	19.82620	19.65239	19.86258
Setembro/16	19.77306	19.70188	19.84425	19.66420	19.88193
Outubro/16	19.76173	19.68755	19.83591	19.64828	19.87518
Novembro/16	19.78444	19.70672	19.86217	19.66558	19.90331
Dezembro/16	19.79121	19.70941	19.87301	19.66611	19.91631
Janeiro/17	19.84516	19.75876	19.93156	19.71303	19.97729
Fevereiro/17	19.75768	19.66618	19.84918	19.61774	19.89762
Março/17	19.66990	19.57282	19.76698	19.52143	19.81837
Abril/17	19.70437	19.60126	19.80748	19.54667	19.86207
Mai/17	19.71390	19.60432	19.82347	19.54632	19.88147
Junho/17	19.72341	19.60698	19.83984	19.54534	19.90147
Julho/17	19.71159	19.58791	19.83526	19.52245	19.90073
Agosto/17	19.69052	19.55926	19.82179	19.48977	19.89128
Setembro/17	19.70610	19.56690	19.84530	19.49322	19.91899
Outubro/17	19.69477	19.54732	19.84221	19.46927	19.92027
Novembro/17	19.71748	19.56149	19.87348	19.47891	19.95606
Dezembro/17	19.72425	19.55941	19.88909	19.47216	19.97635
Janeiro/18	19.77820	19.60425	19.95215	19.51216	20.04423
Fevereiro/18	19.69072	19.50739	19.87404	19.41035	19.97109
Março/18	19.60294	19.40999	19.79589	19.30785	19.89803
Abril/18	19.63741	19.43459	19.84023	19.32722	19.94760
Mai/18	19.64693	19.43401	19.85986	19.32129	19.97258
Junho/18	19.65645	19.43319	19.87970	19.31500	19.99789
Julho/18	19.64463	19.41082	19.87843	19.28705	20.00220

Tabela 12 – Previsões intervalares do algoritmo Holt-Winters com atenuação de tendência

Mês/Ano	Ponto previsão	Inferior 80%	Superior 80%	Inferior 95%	Superior 95%
Agosto/16	19.77716	19.71150	19.84281	19.67675	19.87757
Setembro/16	19.80146	19.73166	19.87125	19.69472	19.90820
Outubro/16	19.80017	19.72598	19.87436	19.68671	19.91363
Novembro/16	19.82668	19.74787	19.90548	19.70616	19.94720
Dezembro/16	19.83459	19.75098	19.91820	19.70672	19.96246
Janeiro/17	19.89236	19.80378	19.98094	19.75689	20.02783
Fevereiro/17	19.80252	19.70884	19.89620	19.65924	19.94579
Março/17	19.72961	19.63071	19.82852	19.57835	19.88087
Abril/17	19.75884	19.65462	19.86306	19.59944	19.91824
Mai/17	19.76451	19.65489	19.87414	19.59686	19.93217
Junho/17	19.77902	19.66393	19.89412	19.60300	19.95505
Julho/17	19.78145	19.66083	19.90208	19.59697	19.96593
Agosto/17	19.75860	19.63240	19.88479	19.56559	19.95160
Setembro/17	19.78362	19.65182	19.91543	19.58204	19.98520
Outubro/17	19.78303	19.64559	19.92047	19.57283	19.99323
Novembro/17	19.81021	19.66711	19.95330	19.59136	20.02905
Dezembro/17	19.81876	19.67000	19.96753	19.59125	20.04628
Janeiro/18	19.87715	19.72270	20.03159	19.64095	20.11335
Fevereiro/18	19.78790	19.62778	19.94802	19.54301	20.03279
Março/18	19.71557	19.54976	19.88137	19.46199	19.96914
Abril/18	19.74534	19.57387	19.91681	19.48309	20.00759
Mai/18	19.75154	19.57440	19.92868	19.48063	20.02245
Junho/18	19.76656	19.58377	19.94934	19.48700	20.04611
Julho/18	19.76947	19.58105	19.95790	19.48130	20.05765

# APÊNDICE C – Rotina do R

```

require(TSA)      #carregando pacotes#
require(tseries)

icmspb = read.table("novoicmspb.txt", header = T)  #leitura dos dados#
lengths(icmspb)

summary(icmspb)

igp.di = read.table("igpdi.txt", header = T)  #leitura dos dados#
length(igp.di[,2])

igp.di = igp.di[,2]/igp.di[,2][length(igp.di[,2])];igp.di

icmspbr = icmspb/igp.di  #defalcionamento#

#foratação dos dados para uma série temporal#
icmspbr = ts(icmspbr, start = c(2003,1), frequency = 12)

#gráfico dos dados#
plot(icmspbr,xlab = "Tempo", ylab = "Valor", main = "Comportamento da série", type = 'o')
acf(icmspbr, main = "Correlograma da Série", lag.max = 70) #correlograma#
#correlograma parcial dos dados#
pacf(icmspbr, main = "Correlograma Parcial da Série", lag.max = 30)
summary(icmspbr)
sqrt(var(icmspbr))
summary(diff(log(icmspbr))*100)

icmspbrl = log(icmspbr) #logaritmo dos dados#
plot(icmspbrl)
acf(icmspbrl)
pacf(icmspbrl)
acf(diff(icmspbrl), main = "Correlograma da primeira diferença", lag.max = 70)
pacf(diff(icmspbrl), main = "Correlograma Parcial da primeira diferença", lag.max = 70)

icmspbr[length(icmspbr)] #última observação#
icmspbrl[length(icmspbrl)] #última observação logaritmada#
observado = icmspbr[length(icmspbr)] # dado a ser predito#
observadolog = icmspbrl[length(icmspbrl)]
icmspbrlt = icmspbrl[-length(icmspbrl)] #série truncada
length(icmspbrlt)
icmspbrlt = ts(icmspbrlt, start = c(2003,1), frequency = 12)

##Alisamneto Exponencial##

```

```
hw1 = HoltWinters(icmspbrrlt, seasonal = "additive") #HW aditivo#
predict(hw1,1)
prev.hw1 = predict(hw1,1)[1] #previsão a 1 passo a frente#
exp(prev.hw1) #previsão HW aditivo#
(observado - exp(prev.hw1))/observado*100 #erro relativo de previsão#

hw2 = HoltWinters(icmspbrrlt, seasonal = "multiplicative") #HW multiplicativo
predict(hw2,1)
prev.hw2 = predict(hw2,1)[1]
exp(prev.hw2) #previsão HW multiplicativo#
(observado - exp(prev.hw2))/observado*100

plot(hw1, main = "Holt-Winters Aditivo",xlab = "Tempo",
     ylab = "Observado/Ajustado", type = 'o')
plot(hw2, main = "Holt-Winters Multiplicativo",xlab = "Tempo",
     ylab = "Observado/Ajustado", type = 'o')

require(forecast) #carregando pacote#
hw3 = hw(icmspbrrlt);hw3
hw4 = hw(icmspbrrlt, damped = T);hw4 #atenuação de tendência#
exp(19.77716) #previsão com atenuação de tendência#
observado
(observado - exp(19.77716))/observado*100
plot(hw3, main = "Holt-Winters Aditivo sem atenuação",xlab = "Tempo")
plot(hw4, main = "Holt-Winters Aditivo com atenuação",xlab = "Tempo")

##Modelagem SARIMA##

auto.arima(icmspbrrlt)
arimafit1 = arima(icmspbrrlt, order = c(0,1,1),
seasonal = list(order=c(0,0,2))) #modelo selecionado#
predict(arimafit1,1)$pred
exp(predict(arimafit1,1)$pred)
observado
(observado - exp(predict(arimafit1,1)$pred))/observado*100

arimafit2 = arima(icmspbrrlt, order = c(0,1,1),
seasonal = list(order = c(0,1,1)))
summary(arimafit2)
(observado - exp(predict(arimafit2,1)$pred))/observado*100
exp(predict(arimafit2,1)$pred)

##Modelagem SARIMAX##

dummy = rep(0, length(icmspbrrlt))
dummy[(length(icmspbrrlt)-92): length(icmspbrrlt)] = 1
```

```
arimafit1d = arima(icmspbrrlt, order = c(0,1,1),
seasonal = list(order = c(0,0,2)),xreg = dummy)
summary(arimafit1d)
exp(predict(arimafit1d,1,newxreg = 1)$pred)
(observado - exp(predict(arimafit1d,1, newxreg = 1)$pred))/observado*100

arimafit2d = arima(icmspbrrlt, order = c(0,1,1),
seasonal = list(order = c(0,1,1)),xreg = dummy)
summary(arimafit2d)
exp(predict(arimafit2d,1,newxreg = 1)$pred)
(observado - exp(predict(arimafit2d,1,newxreg = 1)$pred))/observado*100

tendencia = 1:length(icmspbrrlt)
dtendencia = dummy*tendencia #dummy de interação#

arimafit1dt = arima(icmspbrrlt, order = c(0,1,1),
seasonal = list(order = c(0,0,2)),xreg = dtendencia)
summary(arimafit1dt)
exp(predict(arimafit1dt,1,newxreg=1)$pred)
(observado - exp(predict(arimafit1dt,1,newxreg = 1)$pred))/observado*100

arimafit2dt = arima(icmspbrrlt, order = c(0,1,1),
seasonal = list(order = c(0,1,1)),xreg = dtendencia)
summary(arimafit2dt)
exp(predict(arimafit2dt,1,newxreg=1)$pred)
(observado - exp(predict(arimafit2dt,1,newxreg = 1)$pred))/observado*100
observado - exp(predict(arimafit2dt,1,newxreg = 1)$pred)

arimafit1dt2 = arima(icmspbrrlt, order = c(0,1,1),
seasonal = list(order=c(0,0,2)),xreg = cbind(dummy,dtendencia))
summary(arimafit1dt2)
exp(predict(arimafit1dt2,1,newxreg=cbind(1,164))$pred)
(observado - exp(predict(arimafit1dt2,1,newxreg = cbind(1,164))$pred))/observado*100
observado - exp(predict(arimafit1dt2,1,newxreg = cbind(1,164))$pred)

arimafit2dt2 = arima(icmspbrrlt, order = c(0,1,1),
seasonal = list(order=c(0,1,1)),xreg = cbind(dummy,dtendencia))
summary(arimafit2dt2)
exp(predict(arimafit2dt2,1,newxreg=cbind(1,164))$pred)
(observado - exp(predict(arimafit2dt2,1,newxreg = cbind(1,164))$pred))/observado*100
observado - exp(predict(arimafit2dt2,1,newxreg = cbind(1,164))$pred)

##Previsões com 5 passos a frente##

hw1p = HoltWinters(icmspbrrlt, seasonal = "additive")
predict(hw1p,5)
exp(predict(hw1p,5))
```



```

(observado - exp(predict(hw1,5)))/observado*100

hw2p = HoltWinters(icmspbrlt, seasonal = "multiplicative")
predict(hw2,5)
exp(predict(hw2,5))
(observado - exp(predict(hw2,5)))/observado*100
observado

#SARIMA#

auto.arima(icmspbrlt)
arimafit1 = arima(icmspbrlt, order = c(0,1,1), seasonal = list(order=c(0,0,2)))
predict(arimafit1,5)$pred
exp(predict(arimafit1,5)$pred)
observado
(observado - exp(predict(arimafit1,5)$pred))/observado*100

arimafit2 = arima(icmspbrlt, order = c(0,1,1),seasonal = list(order = c(0,1,1)))
summary(arimafit2)
arimafit2
(observado - exp(predict(arimafit2,5)$pred))/observado*100
exp(predict(arimafit2,5)$pred)

##SARIMAX##
summary(arimafit1d)
exp(predict(arimafit1d,5,newxreg = 5)$pred)
(observado - exp(predict(arimafit1d,5, newxreg = 5)$pred))/observado*100

summary(arimafit2d)
exp(predict(arimafit2d,5,newxreg = 1)$pred)
(observado - exp(predict(arimafit2d,5,newxreg = 1)$pred))/observado*100

summary(arimafit1dt)
exp(predict(arimafit1dt,5,newxreg=1)$pred)
(observado - exp(predict(arimafit1dt,5,newxreg = 1)$pred))/observado*100

summary(arimafit2dt)
exp(predict(arimafit2dt,5,newxreg=1)$pred)
(observado - exp(predict(arimafit2dt,5,newxreg = 1)$pred))/observado*100
observado - exp(predict(arimafit2dt,5,newxreg = 1)$pred)

exp(predict(arimafit1dt2,5,newxreg=cbind(1,164))$pred)
(observado - exp(predict(arimafit1dt2,5,newxreg = cbind(1,164))$pred))/observado*100
observado - exp(predict(arimafit1dt2,5,newxreg = cbind(1,164))$pred)

exp(predict(arimafit2dt2,5,newxreg=cbind(1,164))$pred)
(observado - exp(predict(arimafit2dt2,5,newxreg = cbind(1,164))$pred))/observado*100

```

```
observado - exp(predict(arimafit2dt2,5,newxreg = cbind(1,164))$pred)
```