

UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE ESTATÍSTICA

LENILSON ALMEIDA DE ARAÚJO RODRIGUES

AJUSTE DE UM MODELO DE SÉRIES TEMPORAIS A SÉRIE
DO FATURAMENTO DO COMÉRCIO VAREJISTA DA
GRANDE JOÃO PESSOA-PB

João Pessoa - PB
Julho / 2010

LENILSON ALMEIDA DE ARAÚJO RODRIGUES

**AJUSTE DE UM MODELO DE SÉRIES TEMPORAIS A SÉRIE
DO FATURAMENTO DO COMÉRCIO VAREJISTA DA
GRANDE JOÃO PESSOA-PB**

Monografia apresentado por Lenilson Almeida de Araújo Rodrigues ao Curso de Graduação em Estatística da Universidade Federal da Paraíba, como requisito parcial para obtenção do título de Bacharel. Orientador: Prof. Doutor Eufrásio de Andrade Lima Neto. Área de concentração: Séries Temporais

João Pessoa – PB
Julho / 2010

LENILSON ALMEIDA DE ARAÚJO RODRIGUES

**AJUSTE DE UM MODELO DE SÉRIES TEMPORAIS A SÉRIE
DO FATURAMENTO DO COMÉRCIO VAREJISTA DA
GRANDE JOÃO PESSOA-PB**

Aprovada em ____/____/____.

BANCA EXAMINADORA

Prof. Doutor [Eufrásio de Andrade Lima Neto](#)
Universidade Federal da Paraíba

Prof. Doutor [Ulisses Umbelino dos Anjos](#)
Universidade Federal da Paraíba

Prof^a. Doutora [Renata Patrícia Lima Jerônimo](#)
Universidade Federal da Paraíba

Dedico a meu pai Lucas Rodrigues da Silva por se
dedicar à família e pelo exemplo de vida...

Agradecimentos

Em primeiro lugar agradeço a Deus, por ter guiado o meu caminho até este momento.

Aos meus pais, Lucas e M^a Fátima Rodrigues, por estarem presentes nos momento importante da minha vida me dando apoio e incentivo a sempre prosseguir com os meus objetivos.

A minha irmã, Lenilza Rodrigues, pelo apoio e incentivo.

As minhas tias M^a do Socorro Almeida e M^a Adriana Almeida pelo apoio e incentivo.

Ao professor Eufrásio de Andrade Lima neto, pela orientação neste trabalho e, também, pelos seus ensinamentos, fundamentais para a minha formação profissional.

Aos Professores [Ulisses Umbelino dos Anjos](#) e [Renata Patrícia Lima Jerônimo](#) que aceitaram o convite para participar da banca examinadora.

Aos demais professores do Departamento de Estatística pelos ensinamentos e conhecimento que me proporcionaram durante toda graduação.

À Ivonice Marques, coordenadora do Departamento Econômico da FECOMÉRCIO-PB e a Eliaquim Moraes Melo Técnico IFEP pelas sugestões e conselhos, que contribuíram para o aperfeiçoamento do trabalho e pelos esclarecimentos e sugestões durante o estágio na FECOMERCIO.

Aos meus colegas de curso, Danillo Fagner, Anna Paola Bezerra, Rodrigo Cabral, Josemir Almeida, Marcilio Regis, Pedro Rafael, Julice Soares, Tiê Dias, Everlane Suane, Abner, Telmo Cristiano, Adeilda Fernandes, Lerivam Ferreira, Issac Cabral por terem sido grandes parceiros durante a graduação.

A meus colegas de estágio, Anderson Mendes, Luciano Batista, Brunno Falcão, Talita Pereira, Caroline Silva, Denis Henrique, Cybelle Nóbrega pelo apoio e companheirismo.

As demais pessoas que contribuíram direta ou indiretamente na conclusão deste curso.

RESUMO

O presente trabalho tem por objetivo apresentar um modelo de séries temporais para a previsão dos valores futuros do faturamento do comércio varejista da Grande João Pessoa, composta pelas cidades de João Pessoa, Cabedelo, Santa Rita e Bayeux, no Estado da Paraíba. O modelo foi determinado utilizando-se a fundamentação teórica de séries temporais. A estratégia de formação do modelo foi aplicada a um conjunto de dados referentes ao faturamento do comércio varejista local no período de janeiro de 1999 a janeiro de 2010, o que representa uma amostra de 133 observações. Os modelos obtidos apresentaram-se de fácil interpretação e utilização.

Palavras-chave: Comércio Varejista, Séries temporais, Previsão.

ABSTRACT

This paper aims to present a time series model to predict the future values of sales of retail trade in the main cities (João Pessoa, Cabedelo, Santa Rita and Bayeux) of metropolitan region of João Pessoa, in the State of Paraíba. The model was determined using the time series theory. The model was fitted to a dataset representing the retail sales between January 1999 and January 2010, representing a sample of 133 observations. The final models presented an easy interpretation and use.

Keywords: Retail business, Time series, Forecasting.

LISTA DE FIGURAS

1- Série do faturamento do comércio varejista em geral no período de janeiro de 1999 a janeiro de 2010.....	33
2- Faturamento do comércio varejista em geral – Decomposição por sazonalidade, tendência e resíduo.....	34
3- Logaritmo da série faturamento do comércio varejista em geral no período de janeiro de 1999 a janeiro de 2010.....	35
4-Logaritmo da série faturamento do comércio varejista em geral – Decomposição por sazonalidade, tendência e resíduo.....	35
5 Correlograma do logaritmo da série faturamento do comércio varejista em geral.....	36
6 Logaritmo da série faturamento do comércio varejista em geral – Série diferenciada.....	37
7 Correlograma do logaritmo da série faturamento do comércio varejista em geral – Série diferenciada.....	38
8 Logaritmo da série faturamento do comércio varejista em geral – Série Diferenciada na Tendência e Sazonalidade.....	38
9- Correlograma do Logaritmo da série faturamento do comércio varejista em geral - Série Diferenciada na Tendência e Sazonalidade.....	39
10- Correlograma parcial do logaritmo da série faturamento do comércio varejista em geral - Série Diferenciada na Tendência e Sazonalidade.....	40
11-Análise de diagnóstico do modelo SARIMA (1,1,1)x(1,0,1) ₁₂	42

LISTA DE TABELAS

1-Estrutura do Comércio Varejista da Região da Grande João Pessoa-2009.....	13
2-Critérios AIC, AICc e BIC para os modelos SARIMA.....	40
3-Valores dos erros de previsão EPM, REQM e EAM, para a série Faturamento do comércio varejista em geral transformada a 3, 6 e 12 passos à frente.....	43
4-Previsões três passos à frente - modelo SARIMA(2,1,1)x(1,0,1) ₁₂	43
5-Previsões seis passos à frente - modelo SARIMA (2,1,0)x(2,0,2) ₁₂	44
6-Previsões doze passos à frente – modelo SARIMA(1,1,1)x(1,0,1) ₁₂	44
7-Valores observados e os valores estimados dos modelos SARIMA-Fev-Abr 2010.....	45
8-Erro percentual médio (EPM) para os valores observados e os valores estimados dos modelos SARIMA-Fev-Abr 2010.....	45

SUMÁRIO

LISTA DE FIGURAS

LISTA DE TABELAS

RESUMO

ABSTRACT

1. INTRODUÇÃO	11
2. PROBLEMATIZAÇÃO	12
2.1 Objetivos.....	13
2.1.1 Objetivo Geral.....	13
2.1.2 Objetivos Específicos.....	14
3. METODOLOGIA	15
3.1 Introdução.....	15
3.1.1 Tendência.....	15
3.1.2 Sazonalidade.....	16
3.2 Transformações nos dados.....	16
3.3 Suavização Exponencial.....	17
3.3.1 Modelos para série que apresentam tendência.....	17
3.3.2 Suavização exponencial de Holt (SEH).....	17
3.3.3 Modelos para séries que apresentam sazonalidade.....	18
3.3.4 Suavização exponencial sazonal de Holt-Winters (HW).....	18
3.3.4.1 Série Sazonal multiplicativa.....	19
3.3.4.2 Série Sazonal Aditiva.....	20
3.4 Modelos de Box e Jenkins.....	21
3.4.1 Modelo Auto-Regressivo (AR(p)).....	21
3.4.2 Modelo de Médias Móveis (MA(q)).....	22
3.4.3 Modelo Auto-Regressivo e de Médias Móveis (ARMA(p,q)).....	22
3.4.4 Modelo Auto-Regressivo Integrado e de Médias Móveis (ARIMA (p, d, q)).....	23
3.4.5 Modelo Sazonal Auto-Regressivo Integrado e de Médias Móveis (SARIMA (p, d, q) x (P, D, Q)s).....	23
3.5 Identificação do modelo.....	24
3.5.1 Função de autocorrelação (FAC).....	25
3.5.2 Função de autocorrelação parcial (FACP).....	27

3.6 Métodos baseados em uma função penalizadora.....	29
3.6.1 Critério de Informação de Akaike (AIC).....	29
3.6.2 Critério de Informação Bayesiano (BIC).....	30
3.7 Análise residual.....	30
3.7.1 Teste de Ljung-Box.....	31
3.8 Erros de Previsão.....	31
4. ANÁLISE E DISCUSSÕES.....	33
4.1 Série do Faturamento Comércio Varejista em Geral.....	33
4.2 Suavizamento Exponencial.....	36
4.3 Modelos de Box-Jenkins.....	36
4.4 Comparação dos modelos.....	42
4.5 Previsões Valores Futuros.....	43
5. CONSIDERAÇÕES FINAIS.....	46
REFERENCIAL BIBLIOGRÁFICO.....	47

1. INTRODUÇÃO

Este trabalho é caracterizado como requisito para obtenção do título de Bacharel em estatística apresentado no curso de Graduação em Estatística pela Universidade Federal da Paraíba.

Para a obtenção das informações contida neste trabalho, se fez necessário um vínculo temporário com a Federação do Comércio de Bens, Serviços e Turismo do Estado da Paraíba Fecomercio-PB, no Departamento Econômico, tal que o período de tempo estabelecido se fizesse necessário para a realização desta pesquisa. Dentre os diversos trabalhos realizados pelo departamento econômico havia o interesse em realizar previsões para os faturamentos do comércio varejista para Grande João Pessoa-PB.

Em vista dos fatos mencionados, este trabalho está voltado a analisar a série temporal composta pelo faturamento do comércio varejista nas cidades de João Pessoa, Bayeux, Santa Rita e Cabedelo que compõe a Grande João Pessoa-PB. As informações obtidas percorrem o período de tempo de janeiro de 1999 a janeiro de 2010, num total de 133 observações mensais.

O capítulo 2 apresenta algumas informações sobre o comércio varejista na Grande João Pessoa e os ramos em que estão divididos.. No capítulo 3, serão abordadas algumas técnicas descritivas para análise de uma série temporal, definição, assim como os modelos utilizados no estudo e a metodologia abordada para a identificação e escolha do modelo mais adequado. O Capítulo 4 se refere à análise dos dados. Finalmente, no Capítulo 5, apresentamos as considerações finais.

2. PROBLEMATIZAÇÃO

Para a realização do trabalho de monografia foi escolhido os dados referentes ao comércio varejista da Grande João Pessoa-PB. A escolha deste setor dar-se à importância que ocupa no cenário econômico da Paraíba e do Brasil, desempenhando importante papel na atividade econômica, principalmente, no que diz respeito à geração de emprego e renda. De acordo com o IBGE, no ano de 2006, o comércio apresentou a segunda maior participação na formação do PIB da Paraíba (12,17%), ficando abaixo apenas do setor de administração, saúde e educação pública (30,63%). Considera-se, para fins de estudo, o comércio varejista como a atividade comercial onde 51% ou mais das vendas são voltadas para os consumidores finais e com 51% ou mais destes constituídos de pessoas físicas.

O comércio varejista, para fins deste estudo, está estruturado em cinco grandes grupos: bens de consumo duráveis, semiduráveis, não-duráveis, comércio automotivo e materiais de construção que, por sua vez, foram subdivididos em doze ramos mercantis conforme se vê na Tabela 1. As atividades escolhidas obedeceram ao critério de importância relativa no total do universo para o Estado, pois esses ramos comerciais, de acordo com a Federação do Comércio de Bens, Serviços e Turismo do Estado da Paraíba – FECOMÉRCIO/PB representam, aproximadamente, 80% do volume das vendas do comércio varejista local.

As informações contida neste trabalho sobre o comércio varejista da Grande João Pessoa-PB foram fornecidas pelo Departamento da Pesquisas da Fecomercio-PB, local onde realizo o estágio obrigatório para obtenção do título de Bacharelado em estatística.

A Federação do Comércio de Bens, Serviços e Turismo do Estado da Paraíba foi fundada em 18 de julho de 1949, mas só a partir do ano 2000 vem desempenhando importante papel na área de pesquisas econômicas e sociais. Reconhecida pelo Ministério do Trabalho, Indústria e Comércio através da Carta Sindical de Reconhecimento nº 777.162 de 21 de Dezembro de 1949, apresenta sua sede e foro à Rua Desembargador Souto Maior, 291 – Centro – João Pessoa - PB e base territorial em todo o Estado da Paraíba.

A Federação é constituída pelos Sindicatos filiados, representantes das categorias econômicas do comércio de bens e de serviços, enquadradas nos grupos do plano da Confederação Nacional do Comércio – CNC, integrantes do sistema confederativo de representação sindical (Sicomercio), instituído com fundamento no art. 8º, inciso IV da Constituição Federal.

Os dados constam de 133 observações do comércio varejista em geral da Região da Grande João Pessoa-PB, que percorrem o período de tempo de janeiro de 1999 a janeiro de 2010. A Tabela 1 ilustra os grupos e subgrupos, número de estabelecimentos constantes em cada subgrupo e o faturamento real (em R\$), deflacionado pelo IPCA do IBGE, obtido em cada ramo no ano de 2009.

Tabela 1: Estrutura do Comércio Varejista da Região da Grande João Pessoa-2009

Grupo /Subgrupo	Nº Estabe-cimentos	Faturamento (em R\$)
COMÉRCIO EM GERAL	-	4.761.506.357
1. BENS DE CONSUMO DURÁVEIS	-	304.972.473
1.1 Móveis e Decorações	26	71.597.624
1.2 Lojas de Utilidades Domésticas	11	195.158.517
1.3 Cine-Foto-Som e Óticas	30	38.216.331
2. BENS DE CONSUMO SEMI-DURÁVEIS	-	564.099.748
2.1 Vestuário	29	160.939.356
2.2 Tecidos	8	52.720.232
2.3 Calçados	20	350.440.160
3. BENS DE CONSUMO NÃO DURÁVEIS	-	2.246.142.693
3.1 Supermercados	19	1.115.289.480
3.2 Farmácias, Drogarias e Perfumarias	40	384.499.826
3.3 Combustíveis	33	746.353.387
4. COMÉRCIO AUTOMOTIVO	-	1.204.526.858
4.1 Concessionárias de Veículos	15	1.106.247.734
4.2 Autopeças e Acessórios	35	98.279.124
5. MATERIAIS DE CONSTRUÇÃO	61	441.764.586

2.1 Objetivos

2.1.1 Objetivo Geral

Ajustar um modelo de séries temporais para a série do faturamento do comércio varejista em geral da Região da Grande João Pessoa-PB e realizar previsões futuras a curto, médio e longo prazo.

2.1.2 Objetivos Específicos

- i) Ajustar modelos de Holt-Winters e de Box-Jenkins;

- ii) Realizar uma análise residual
- iii) Obter previsões para 3, 6 e 12 períodos à frente;
- iv) Calcular os erros de previsão para os modelos escolhidos;
- v) Verificar o modelo com melhor poder de previsão;
- vi) Realizar previsões futuras.

3. METODOLOGIA

3.1 Introdução

Considere um conjunto de observações $\{Z_t, t = 1, \dots, N\}$ de uma série temporal, onde z_t será um modelo de decomposição escrito como uma soma de três componentes não-observáveis,

$$z_t = T_t + S_t + a_t \quad (1.1)$$

onde T_t e S_t representam a tendência e sazonalidade, respectivamente, e a_t é uma componente aleatória, de média zero e variância constante σ_a^2 (MORETTIN, 2005).

Segundo Morettin (2005) a importância do modelo (1.1) é o de estimar S_t e construir a série livre de sazonalidade ou sazonalidade ajustada. Se \hat{S}_t for uma estimativa de S_t ,

$$z_t^{SA} = z_t - \hat{S}_t \quad (1.2)$$

será a série sazonalmente ajustada. As componentes T_t e S_t são, em geral, bastante relacionadas e a influencia da tendência sobre a componente sazonal pode ser muito forte (Pierce, 1979).

Segundo Morettin (2005), ao estimarmos T_t e S_t e subtrairmos de z_t , obteremos uma estimativa da componente aleatória a_t .

3.1.1 Tendência

Segundo Morettin (2005), suponha que a componente sazonal não esteja presente. O modelo considerado será o seguinte

$$z_t = T_t + a_t \quad (1.3)$$

onde a_t é o ruído branco, com variância σ_a^2 .

Estimando-se a tendência através de \hat{T}_t , podemos obter a série ajustada para tendência ou livre de tendência, denotada por

$$y_t = z_t - \hat{T}_t.$$

Um procedimento para eliminar a tendência de uma série é tomar diferenças da série original.

3.1.2 Sazonalidade

A série temporal pode apresentar um comportamento periódico, estando associado as variações climáticas e datas comemorativas. O período de sazonalidade é representado por s .

O que podemos observar em séries sazonais é a ocorrência de relações entre:

- (i) observações para meses sucessivos em um ano particular;
- (ii) as observações para o mesmo mês em ano sucessivos.

Assim, z_t é relacionada com z_{t-1}, z_{t-2}, \dots , mas também com z_{t-s}, z_{t-2s}, \dots . Isto implica que séries sazonais são caracterizadas por apresentarem alta correlação em “lags sazonais”, isto é, lags que são múltiplos do período s . O objetivo será ajustar uma série para a componente sazonal, ou seja, estimar S_t e subtrair a série estimada de z_t no modelo (1.1).

Se o modelo for multiplicativo, da forma

$$z_t = T_t S_t a_t. \quad (1.4)$$

Então, a série sazonalmente ajustada será

$$z_t^{SA} = z_t / \hat{S}_t. \quad (1.5)$$

3.2 Transformações nos dados

Há basicamente, duas razões para se transformar os dados originais: estabilizar a variância e tornar o efeito sazonal aditivo. Entretanto, Nelson (1976) conclui que transformações não melhoram a qualidade da previsão. No caso de séries econômicas e financeiras poderá ser necessário aplicar antes, à série original, alguma transformação não-linear, como a logarítmica ou, em geral, uma transformação da forma

$$Z_t^{(\lambda)} = \begin{cases} \frac{Z_t^\lambda - c}{\lambda}, & \text{se } \lambda \neq 0 \\ \log Z_t, & \text{se } \lambda = 0 \end{cases},$$

chamada *transformação de Box-Cox* (1964). Aqui λ e c são parâmetros a serem estimados. Outra razão para efetuar transformações é obter uma distribuição para os dados mais simétrica e próxima da normal (MORETTIN, 2005).

3.3 Suavização Exponencial

3.3.1 Modelos para série que apresentam tendência

Segundo Morettin (2005), vamos considerar o caso de uma série temporal que não apresente sazonalidade, ou seja, que não apresente um ciclo repetitivo em subperíodos de um período de tempo determinado, que apresenta soma de nível, tendência e ruído branco com média zero e variância constante (σ_a^2), isto é,

$$z_t = \mu_t + T_t + a_t, \quad t = 1, \dots, N$$

3.3.2 Suavização exponencial de Holt (SEH)

Para séries temporais que apresentam certo comportamento linear como, por exemplo, crescente ou decrescente ao longo do tempo, utilizaremos o método de Holt.

Segundo Morettin (2005) para evitar que uma série apresente tendência linear positiva (ou negativa) e que apresente previsões que subestimam (ou superestimam) continuamente os verdadeiros valores da série, um dos métodos a ser aplicado é a suavização exponencial de Holt. Este método suaviza o nível e utiliza uma nova constante de suavização para modelar a tendência da série.

A modelagem de suavização exponencial de Holt para o nível e a tendência da série, no instante t, serão estimadas por

$$\hat{N}_t = Az_t + (1-A)(\hat{N}_{t-1} + \hat{T}_{t-1}), \quad 0 < A < 1 \text{ e } t = 2, \dots, N, \quad (1.6)$$

$$\hat{T}_t = C(\hat{N}_t - \hat{N}_{t-1}) + (1-C)\hat{T}_{t-1}, \quad 0 < C < 1 \text{ e } t = 2, \dots, N, \quad (1.7)$$

respectivamente. Onde A e C são constantes de suavização. As fórmulas acima modificam as estimativas prévias quando uma nova observação é obtida.

A previsão para o valor z_{t+h} , com origem em t, é dada por

$$\hat{z}_t(h) = \hat{N}_t + h\hat{T}_t, \quad \forall h > 0, \quad (1.8)$$

ou seja, a previsão é feita adicionando-se ao valor básico (\hat{N}_t) a tendência multiplicada pelo número de passos à frente (h) que se deseja prever (MORETTIN 2005).

As equações (1.6) e (1.7) podem ser utilizadas para atualização da previsão, tomando-se uma nova observação z_{t+1} . Assim,

$$\hat{N}_{t+1} = Az_{t+1} + (1-A)(\hat{N}_t + \hat{T}_t)$$

$$\hat{T}_{t+1} = C(\hat{N}_{t+1} - \hat{N}_t) + (1-C)\hat{T}_t$$

e a nova previsão para o valor z_{t+h} será

$$\hat{z}_{t+1}(h-1) = \hat{N}_{t+1} + (h-1)\hat{T}_{t+1}. \quad (1.9)$$

Para que as equações acima possam ser utilizadas, temos que fazer hipóteses sobre os valores iniciais. O procedimento mais simples é colocar $\hat{T}_2 = z_2 - z_1$ e $\hat{N}_2 = z_2$ (MORETTIN 2005).

Um critério de escolha para as constantes de suavização é a minimização da soma dos erros quadrados de previsão um passo a frente. Ou seja, escolhemos A e C de forma a minimizar (PEREIRA 2006).

$$S = \sum_{t=4}^n (z_t - \hat{z}_t(h))^2, \quad t = 4, \dots, n.$$

3.3.3 Modelos para séries que apresentam sazonalidade

Para séries temporais que apresentam um comportamento repetitivo, utilizaremos o método de Holt-Winters.

3.3.4 Suavização exponencial sazonal de Holt-Winters (HW)

Segundo Morettin (2005), existem dois tipos de procedimentos cuja utilização depende das características da série considerada. Tais procedimentos são baseados em três equações \hat{N}_t , \hat{T}_t e \hat{F}_t com constantes de suavização diferentes, que são associadas a cada uma das componentes da série: nível, tendência e sazonalidade, respectivamente.

3.3.4.1 Série Sazonal multiplicativa

Considere uma série sazonal com períodos s . A variante mais usual do método Holt-Winters considera o fator sazonal F_t como sendo multiplicativo, enquanto a tendência permanece aditiva, isto é,

$$z_t = \mu_t F_t + T_t + a_t, \quad t = 1, \dots, N. \quad (1.10)$$

com a sazonalidade multiplicativa, as três equações de suavização são dadas por

O procedimento anterior pode ser modificado para tratar situações onde o fator sazonal é aditivo, ou seja,

$$z_t = \mu_t + T_t + F_t + a_t. \quad (1.16)$$

As estimativas do fator sazonal, nível e tendência da série são dadas por

$$\hat{F}_t = D(z_t - \hat{N}_t) + (1 - D)\hat{F}_{t-s}, \quad 0 < D < 1, \quad (1.17)$$

$$\hat{N}_t = A(z_t - \hat{F}_{t-s}) + (1 - A)(\hat{z}_{t-1} + \hat{T}_{t-1}), \quad 0 < A < 1, \quad (1.18)$$

$$\hat{T}_t = C(\hat{N}_t - \hat{N}_{t-1}) + (1 - C)\hat{T}_{t-1}, \quad 0 < C < 1, \quad (1.19)$$

respectivamente; Novamente, A, C e D representam as constantes de suavização.

As previsões futuras da série são dadas por

$$\begin{aligned} \hat{z}_t(h) &= \hat{N}_t + h\hat{T}_t + \hat{F}_{t+h-s}, \quad h = 1, 2, \dots, s, \\ \hat{z}_t(h) &= \hat{N}_t + h\hat{T}_t + \hat{F}_{t+h-2s}, \quad h = s + 1, \dots, 2s, \end{aligned} \quad (1.20)$$

onde \hat{N}_t , \hat{T}_t e \hat{F}_t são dados por (1.18), (1.19) e (1.17), respectivamente.

As atualizações são feitas utilizando (1.17), (1.18) e (1.19):

$$\begin{aligned} \hat{F}_{t+1} &= D(z_{t+1} - \hat{N}_{t+1}) + (1 - D)\hat{F}_{t+1-s}, \\ \hat{N}_{t+1} &= A(z_{t+1} - \hat{F}_{t+1-s}) + (1 - A)(\hat{N}_t + \hat{T}_t) \\ \hat{T}_{t+1} &= C(\hat{N}_{t+1} - \hat{N}_t) + (1 - C)\hat{T}_t \end{aligned}$$

e a nova previsão para o valor z_{t+h} será

$$\begin{aligned} \hat{z}_{t+1}(h-1) &= \hat{N}_{t+1} + (h-1)\hat{T}_{t+1} + \hat{F}_{t+1+h-s}, \quad h = 1, \dots, s+1, \\ \hat{z}_{t+1}(h-1) &= \hat{N}_{t+1} + (h-1)\hat{T}_{t+1} + \hat{F}_{t+1+h-2s}, \quad h = s+2, \dots, 2s+1. \end{aligned} \quad (1.21)$$

3.4 Modelos de Box e Jenkins

Um método de previsão bastante utilizado na análise de modelos paramétricos são conhecidos por modelos de Box-Jenkins (1970), representados por um conjunto de processos estocásticos conhecidos por modelos ARIMA. Os modelos ARIMA, por analisarem apenas uma série temporal de cada vez, são conhecidos como modelos univariados. Podemos descrever o modelo ARIMA como auto-regressivo (AR(p)), Integrado (I(d)) e de médias móveis (MA(q)), sendo as três componentes denominadas filtros. Uma série temporal pode conter os três filtros como também apenas um subconjunto deles.

Para uma mesma série temporal podemos ter mais de um modelo diferente, dependendo da interpretação dos pesquisadores.

Segundo Morettin (2005), em geral, os modelos postulados são parcimoniosos, pois contêm um número pequeno de parâmetros e as previsões obtidas são bastante precisas, comparando-se aos demais métodos de previsão.

A metodologia Box-Jenkins é utilizada aos processos estocásticos que sejam estacionários. Um processo estocástico é dito estacionário de segunda ordem quando as condições a seguir forem satisfeitas, para qualquer instante de tempo t ,

$$\begin{aligned} E[z_t] &= E[z_{t+k}] = \mu \\ \text{Var}[z_t] &= E[(z_t - \mu)^2] = \sigma^2 \\ \text{Cov}[z_t, z_{t+k}] &= E[(z_t - \mu)(z_{t+k} - \mu)] \end{aligned}$$

As duas primeiras condições indicam que a média e a variância de z_t não variam com o tempo, e a terceira indica que as autocovariâncias não dependem do tempo e sim da distância k que separa as observações.

Se o processo não se apresentar estacionário, podemos torná-la estacionário por meio de sucessivas diferenciações na série temporal (SILVA 2005).

3.4.1 Modelo Auto-Regressivo (AR(p))

Obtemos um modelo auto-regressivo de ordem p (AR(p)), através da soma ponderada de valores passados $\tilde{z}_{t-1}, \tilde{z}_{t-2}, \dots$, mais um ruído branco a_t , sendo definido por

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + \dots + \phi_p \tilde{z}_{t-p} + a_t, \quad (1.22)$$

onde, $\tilde{z}_t = z_t - \mu$, ϕ_i são parâmetros da estrutura AR(p), $i=1, \dots, p$, e a_t é ruído branco com média zero e variância σ_a^2 .

Utilizando um operador defasagem auto-regressivo estacionário de ordem p .

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (1.23)$$

tem-se,

$$\phi(B)\tilde{z}_t = a_t \quad (1.24)$$

3.4.2 Modelo de Médias Móveis (MA(q))

Os modelos de médias móveis são formados por combinações linear do ruído branco a_t , ocorridos no período corrente e nos períodos passados. Apresentam a seguinte notação

$$z_t = \mu + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}. \quad (1.25)$$

Sendo $\tilde{z}_t = z_t - \mu$, teremos

$$\tilde{z}_t = (1 - \theta_1 B - \dots - \theta_q B^q) a_t = \theta(B) a_t, \quad (1.26)$$

onde,

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q, \quad (1.27)$$

é o operador de média móveis de ordem q , θ_i são parâmetros da estrutura MA(q), $i=1, \dots, q$, e a_t é ruído branco com média zero e variância σ_a^2 .

3.4.3 Modelo Auto-Regressivo e de Médias Móveis (ARMA(p,q))

Esse modelo é uma combinação dos dois modelos vistos anteriormente, onde \tilde{z}_t é descrito por seus valores passados e pelo ruído branco corrente e passados. Os modelos ARMA (p, q) são descrito como

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \dots + \phi_p \tilde{z}_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}, \quad (1.28)$$

onde, ϕ_i são parâmetros da estrutura AR(p), $i=1, \dots, p$, θ_i são parâmetros da estrutura MA(q), $i=1, \dots, q$, e a_t é ruído branco com média zero e variância σ_a^2 .

Se $\phi(B)$ e $\theta(B)$ são os operadores auto-regressivos e de médias móveis, respectivamente, introduzidos anteriormente. Assim, podemos escrever (1.28) na forma

$$\phi(B) \tilde{z}_t = \theta(B) a_t. \quad (1.29)$$

3.4.4 Modelo Auto-Regressivo Integrado e de Médias Móveis (ARIMA (p, d, q))

Quando temos uma série temporal cujo processo estocástico apresenta um comportamento não estacionário, onde se precisa aplicar diferenças na série original para torna-lá estacionária, o número total de diferenças é chamada ordem de integração. Um processo é dito ser integrado de ordem d se z_t é não-estacionária, mas a série resultante após a aplicação de d diferenças é estacionária. Daí dizermos que Z_t segue um modelo de previsão auto-regressivo, integrado, de médias móveis ou modelo ARIMA(p, d, q), definido por

$$\phi(B)\Delta^d z_t = \theta(B)a_t \quad (1.30)$$

onde: $\phi(B)$ representa o operador auto-regressivo de ordem p, $\theta(B)$ representa o operador de médias móveis de ordem q, a_t representa o ruído branco com média zero e variância σ_a^2 , d representa o número de diferenças e $\Delta = 1 - B$ representa o operador diferença.

Este operador diferença é definido como

$$\begin{aligned} z_t - z_{t-1} &= z_t - Bz_t = (1 - B)z_t = \Delta z_t \\ \Delta^d &= (1 - B)^d \end{aligned}$$

3.4.5 Modelo Sazonal Auto-Regressivo Integrado e de Médias Móveis (SARIMA (p, d, q) x (P, D, Q)s)

Quando a série temporal apresentar sazonalidade, nem sempre é possível transformar z_t de forma a retirar esta sazonalidade, ou ainda, a própria sazonalidade pode apresentar um padrão dinâmico. Isto significa que há necessidade de se considerar uma sazonalidade estocástica, ou seja, ajustar à série original um modelo ARIMA sazonal (SARIMA). Seja z_t a série de interesse observada s períodos por ano. A idéia é relacionar uma observação z_t , correspondente a um determinado mês, janeiro por exemplo, com observações correspondentes a janeiros anteriores, através de um modelo ARIMA sazonal da forma

$$\Phi(B^s)\Delta_s^D z_t = \Theta(B^s)\alpha_t \quad (1.31)$$

onde $s=12$, $\Delta_s = 1 - B^s$ e $\Phi(B^s)$ e $\Theta(B^s)$ são polinômios em B^s de grau P e Q, respectivamente, e satisfazem as condições de estacionaridade e invertibilidade.

Analogamente, teríamos um modelo análogo ao (1.31), que relacionaria os meses de dezembro

$$\Phi(B^s)\Delta_s^D z_{t-1} = \Theta(B^s)\alpha_{t-1}, \quad (1.32)$$

onde as componentes do erro $\alpha_t, \alpha_{t-1}, \dots$, neste modelo, não são ruídos brancos, isto é, existe uma correlação entre eles. Pode-se esperar que α_t esteja relacionada com $\alpha_{t-1}, \alpha_{t-2}, \dots$. Para descrever esta relação introduzimos, para os α_t , o modelo ARIMA usual

$$\phi(B)\Delta^d \alpha_t = \theta(B)a_t \quad (1.33)$$

onde a_t é o ruído branco com média zero e variância σ_a^2 , $\Delta = \Delta_1 = 1 - B$ e $\phi(B)$ e $\theta(B)$ são polinômios em B de grau p e q , respectivamente, e satisfazem as condições de estacionaridade e invertibilidade.

Substituindo (1.32) em (1.33) tem-se o modelo sazonal multiplicativo ARIMA, também conhecido como SARIMA, de ordem $(p, d, q) \times (P, D, Q)_s$, representado por

$$\phi(B)\Phi(B^s)\Delta^d\Delta_s^D z_t = \theta(B)\Theta(B^s)a_t. \quad (1.34)$$

O modelo (1.34) implica que devemos tomar d diferenças sazonais da série z_t , de modo que o processo

$$w_t = \Delta^d \Delta_s^D z_t$$

seja estacionário.

3.5 Identificação do modelo

A metodologia Box-Jenkins têm como principal objetivo determinar as três componentes que configuram qualquer estrutura que compõe o processo gerador da série, isto é, os p parâmetros auto-regressivos, processo de diferenciação d e os q parâmetros de médias móveis.

O processo utilizado na previsão através da metodologia Box-Jenkins consiste em realizar a diferenciação da série original tantas vezes quanto necessário para torná-la estacionária, além de identificar os valores p e q através da análise das funções de autocorrelação (FAC) e de autocorrelação parcial (FACP) estimadas, vista na seção a seguir.

A análise destas funções permite a tomada da decisão acerca do número p de parâmetros auto-regressivos e do número q de parâmetros de médias móveis que devem ser escolhidos, de maneira a se obter uma estrutura que tenha o menor número de parâmetros dentre todas as estruturas que se ajustem aos dados.

Para verificar a condição de estacionaridade podemos utilizar o gráfico da série original para obtermos uma primeira indicação da infração ou não desta condição. A identificação visa encontrar uma classe de estruturas a serem ajustadas e validadas.

3.5.1 Função de autocorrelação (FAC)

Vamos multiplicar ambos os membros da equação (1.22) por \tilde{z}_{t-j} e tomando-se a esperança obteremos

$$E(\tilde{z}_t \tilde{z}_{t-j}) = \phi_1 E(\tilde{z}_{t-1} \tilde{z}_{t-j}) + \phi_2 E(\tilde{z}_{t-2} \tilde{z}_{t-j}) + \dots + \phi_p E(\tilde{z}_{t-p} \tilde{z}_{t-j}) + E(a_t \tilde{z}_{t-j}).$$

Devido (1.22), \tilde{z}_{t-j} só envolve ruídos até a_{t-j} , não correlacionados com a_t , $E(a_t \tilde{z}_{t-j}) = 0$, $j > 0$, do onde resulta

$$\gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} + \dots + \phi_p \gamma_{j-p}, \quad j > 0. \quad (1.35)$$

Dividindo-se por $\gamma_0 = \text{Var}(z_t)$, obteremos

$$\rho_j = \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2} + \dots + \phi_p \rho_{j-p}, \quad j > 0. \quad (1.36)$$

Que também pode ser escrita como

$$\phi(B) \rho_j = 0,$$

Onde o operador B agora age em j : $B\rho_j = \rho_{j-1}$ etc. se

$$\phi(B) = \prod_{i=1}^p (1 - G_i B),$$

então pode-se demonstrar que a solução geral de (1.36) é, no caso de raízes distintas,

$$\rho_j = A_1 G_1^j + A_2 G_2^j + \dots + A_p G_p^j.$$

Como $|G_i| < 1$, duas situações podem ocorrer:

- (a) se G_i for real, o termo $A_i G_i^j$ decai geometricamente para zero (amortecimento exponencial);
- (b) um par de raízes complexas conjugadas contribui com um termo da forma $A d^j \text{sen}(2\pi f j + \varphi)$ (senóide amortecida).

Genericamente, a função de autocorrelação de um processo auto-regressivo é construída de uma mistura de polinômios, exponenciais e senóides amortecidas.

Para $j=0$ na expressão de $E(\tilde{z}_t \tilde{z}_{t-j})$ obtemos

$$\text{Var}(\tilde{z}_t) = \text{Var}(z_t) = \gamma_0 = \phi_1 \gamma_{-1} + \dots + \phi_p \gamma_{-p} + \sigma_a^2$$

e com $\gamma_{-j} = \gamma_j$, obtemos

$$1 = \phi_1 \rho_1 + \dots + \phi_p \rho_p + \frac{\sigma_a^2}{\gamma_0},$$

ou seja,

$$\text{Var}(z_t) = \sigma_z^2 = \frac{\sigma_a^2}{1 - \phi_1 \rho_1 - \dots - \phi_p \rho_p}.$$

Se fizermos $j=1,2,\dots,p$ em (1.36) obtemos

$$\begin{aligned} \rho_1 &= \phi_1 + \phi_2 \rho_1 + \dots + \phi_p \rho_{p-1} \\ \rho_2 &= \phi_1 \rho_1 + \phi_2 + \dots + \phi_p \rho_{p-2} \\ &\dots \\ \rho_p &= \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \dots + \phi_p, \end{aligned}$$

que são denominadas *equações de Yule-Walker*. Em forma matricial podemos escrever

$$\begin{bmatrix} 1 & \rho_1 \dots \rho_{p-1} \\ \rho_1 & 1 \dots \rho_{p-2} \\ \dots & \dots \\ \rho_{p-1} & \rho_{p-2} \dots 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \dots \\ \phi_p \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \dots \\ \rho_p \end{bmatrix}. \quad (1.37)$$

Podemos estimar os coeficientes ϕ_1, \dots, ϕ_p do modelo $\text{AR}(p)$ usando (1.37), substituindo-se as FAC ρ_j por suas estimativas r_j dadas por

$$r_j = \frac{c_j}{c_0}, \quad j = 0, 1, \dots, N-1,$$

onde c_j é a estimativa da função de auto-covariância γ_j ,

$$c_j = \frac{1}{N} \sum_{t=1}^{N-j} [(X_t - \bar{X})(X_{t+j} - \bar{X})], \quad j = 0, 1, \dots, N-1$$

sendo $\bar{X} = \frac{1}{N} \sum_{t=1}^N X_t$ a média amostral. Colocando, aqui $c_{-j} = c_j$ e $r_{-j} = r_j$ (MORETTIN 2005).

3.5.2 Função de autocorrelação parcial (FACP)

Box, Jenkins e Rinsel (1994) propuseram a utilização de um instrumento para facilitar o procedimento de identificação de um modelo, denominado função de autocorrelação parcial (FACP).

Denotaremos por ϕ_{kj} o j-ésimo coeficiente de um modelo AR(k), onde ϕ_{kk} seja o último coeficiente. Sabendo que

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \dots + \phi_{kk}\rho_{j-k}, \quad j = 1, \dots, k,$$

A partir daí podemos obter o sistema conhecido como equação de Yule-Walker

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ \vdots & & & & \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \cdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \cdots \\ \rho_k \end{bmatrix}$$

Resolvendo estas equações sucessivamente para $k=1,2,3,\dots$ obtemos

$$\begin{aligned} \phi_{11} &= \rho_1 \\ \phi_{22} &= \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \\ \phi_{33} &= \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} \end{aligned}$$

e, em geral,

$$\phi_{kk} = \frac{|P_k^*|}{|P_k|},$$

onde P_k é a matriz de autocorrelações e P_k^* é a matriz P_k com a última coluna substituída pelo vetor de autocorrelações.

Sendo a quantidade ϕ_{kk} , encarada como função de k, é chamada função de autocorrelação parcial.

Para os processos estudados, temos que

- (i) Um processo AR(p) tem FACP $\phi_{kk} \neq 0$, para $k \leq p$ e $\phi_{kk} = 0$, para $k > p$;
- (ii) Um processo MA(q) tem FACP que se comporta de maneira similar à FAC de um processo AR(p);
- (iii) Um processo ARMA(p,q) tem FACP que se comporta como a FACP de um processo MA puro.

Outra maneira para encontra a estimativa da FACP seria substituir nas equações de Yule-Walker as FAC ρ_j por suas estimativas,

$$r_j = \hat{\phi}_{k1} r_{j-1} + \dots + \hat{\phi}_{kk} r_{j-k}, \quad j = 1, \dots, k$$

e resolver estas equações para $k=1,2,\dots$

3.6 Métodos baseados em uma função penalizadora

Segundo Morettin (2005) vamos escolher as ordens p e q que minimizem a quantidade

$$P(p, q) = \ln \hat{\sigma}_{p,q}^2 + (p + q) \frac{C(N)}{N} \quad (1.38)$$

em que $\hat{\sigma}_{p,q}^2$ é uma estimativa da variância residual obtida ajustando um modelo ARMA(p,q) às N observações da série e C(N) é uma função do tamanho da série.

A quantidade $(p + q) \frac{C(N)}{N}$, denominada termo penalizador, aumenta quando o número de parâmetros aumenta, enquanto que a variância residual $\hat{\sigma}_{p,q}^2$ diminui. Assim, minimizar (1.38) corresponde a identificar as ordens p e q que equilibrem esse comportamento (MORETTIN 2005).

Veremos agora dois procedimentos de identificação que minimizam funções penalizadoras.

3.6.1 Critério de Informação de Akaike (AIC)

Neste critério objetiva escolher o modelo cujas ordens k e l minimizam

$$AIC(k, d, l) = N \ln \hat{\sigma}_a^2 + \frac{N}{N-d} 2(k+l+1 + \delta_{d0}) + N \ln 2\pi + N, \quad (1.39)$$

em que

$$\delta_{d0} = \begin{cases} 1, & d = 0 \\ 0, & d \neq 0 \end{cases}$$

e $\hat{\sigma}_a^2$ é o estimador de máxima verossimilhança de $\hat{\sigma}_a^2$ (MORETTIN 2005).

Segundo Morettin (2005), ao comparar vários modelos, com N fixo, os dois últimos termos de (1.39) podem ser desprezados. Levando-se em conta que, geralmente, identificarmos a série apropriadamente diferenciada, obtemos

$$AIC(k, l) = N \ln \hat{\sigma}_a^2 + 2(k + l + 2) \quad (1.40)$$

como critério para a determinação das ordens p e q. Vamos estipular valores limites superiores k e l e calcular (1.40) para todas as possíveis combinações (k,l) com $0 \leq k \leq K$ e $0 \leq l \leq L$. Então podemos escrever o AIC da seguinte maneira

$$AIC(k, l) = \ln \hat{\sigma}_{k,l}^2 + \frac{2(k + l)}{N}, \quad (1.41)$$

pois os valores de k e l que minimizem esta última expressão são os mesmos que minimizam (1.40). Para o caso de modelos AR(p), o critério AIC reduz-se a

$$AIC(k, l) = N \ln \hat{\sigma}_k^2 + 2k, \quad k \leq K. \quad (1.42)$$

Existem várias correções para melhorar o comportamento do AIC, no sentido de diminuir a probabilidade de selecionar uma ordem maior de que a verdadeira. Hurvich e Tsai (1989) propõe uma correção para o AIC, dada por

$$AIC_c(k) = AIC(k) + \frac{2(k + 1)(k + 2)}{N - k + 2}, \quad k \leq K.$$

Um utilizando simulações, mostram que esta correção é útil quando N é pequeno ou quando K é uma fração “moderadamente grande” de N (MORETTIN 2005).

3.6.2 Critério de Informação Bayesiano (BIC)

Baseados em uma perspectiva bayesiana, introduziram-se critérios equivalentes para seleção de modelos. O critério de informação Bayesiano é dado por

$$BIC = -2 \ln(ML) + (\text{número de parâmetros}) \ln N.$$

No caso de um processo ARMA, essa quantidade é dada por

$$BIC(k, l) = \ln \hat{\sigma}_{k,l}^2 + (k + l) \frac{\ln N}{N}, \quad (1.43)$$

onde $\hat{\sigma}_{k,l}^2$ é a estimativa de máxima verossimilhança da variância residual do modelo ARMA (k,l) (MORETTIN 2005).

3.7 Análise residual

Espera-se que um modelo bem ajustado apresente os resíduos aleatoriamente distribuídos em torno de zero, com variância aproximadamente constante e sejam não correlacionados. Se a variância dos resíduos for crescente, uma transformação logarítmica nos dados pode ser apropriada. O fenômeno de “não constância” na variância é denominado de volatilidade na literatura de séries temporais e pode ser tratado através de transformações nos dados (EHLERS, 2007). Caso os resíduos padronizados encontrem-se dentro do intervalo de -2 a 2, sinaliza que o modelo ajustado apresentou um bom ajuste aos dados. Além disso, o correlograma dos resíduos deve apresentar as autocorrelações amostrais entre os limites assintóticos de significância ao nível de 95%. O teste de Ljung-Box, descrito na próxima seção, também deve ser utilizado para verificar se os resíduos apresentam-se não-correlacionados. Logo em seguida aplica-se um teste de normalidade, para se verificar a adequação do modelo.

3.7.1 Teste de Ljung-Box

Seja a estatística Q , definida por:

$$Q = n \sum_{j=1}^h \hat{\rho}^2(j),$$

onde ρ representa a autocorrelação entre as observações z_t e z_{t+j} . Considere uma sequência de observações com variância finita Z_1, \dots, Z_n , independente e identicamente distribuída (iid), temos que Q será distribuída aproximadamente como uma soma de quadrados de uma sequência de variáveis aleatórias que seguem uma distribuição $N(0,1)$. Logo, $\sqrt{n}\hat{\rho}(j)$, $j = 1, \dots, h$, apresentará uma distribuição qui-quadrado com h graus de liberdade. Grandes valores de Q sugerem que a hipótese de independência deve ser rejeitada, ou seja, os dados apresentam alguma autocorrelação. Portanto, rejeitamos a hipótese de iid ao nível α de significância se $Q > \chi_{1-\alpha}^2(h)$, onde $\chi_{1-\alpha}^2(h)$ é $1-\alpha$ quantil de uma qui-quadrado com h graus de liberdade.

3.8 Erros de Previsão

Um dos objetivos principais na análise de séries temporais é realizar previsões de valores futuros. Para isto, utilizamos alguns critérios para verificar quanto o modelo é preciso em suas previsões futuras.

Das N observações que constitui a série temporal, vamos destinar os n primeiros valores ($n < N$) para criar um modelo que represente esta série incompleta. Segundo Pereira (2006) utilizamos os outros $N-n$ dados para verificar a qualidade de $N-n$ previsões feitas a partir do modelo estimado. Logo, temos que os erros de previsão dado por

$$\varepsilon_j = z_j - \hat{z}_j,$$

onde \hat{z}_j denota a previsão da observação z_j .

As capacidades das diferentes previsões do modelo podem ser comparadas pelas medidas a seguir, onde, em todas as medidas, o erro de previsão é utilizado e as previsões são feitas a partir dos n primeiros valores da série. A escolha do modelo será através do que apresentar os menores erros de previsão para h passos à frente. As medidas consideradas são:

(i) Erro Percentual Médio (EPM):

$$EPM = \left(\frac{\sum_{j=t+1}^{t+h} z_j - \sum_{j=t+1}^{t+h} \hat{z}_j}{\sum_{j=t+1}^{t+h} z_j} \right) \times 100.$$

(ii) Raiz Erro Quadrático Médio (REQM):

$$EQM = \sqrt{\frac{1}{h} \sum_{j=t+1}^{t+h} (z_j - \hat{z}_j)^2}.$$

(iii) Erro Absoluto Médio (EAM):

$$EAM = \frac{1}{h} \left(\sum_{j=t+1}^{t+h} |z_j - \hat{z}_j| \right).$$

4. ANÁLISE E DISCUSSÕES

A análise e discussão constam da série do Faturamento do comércio varejista em geral da Grande João Pessoa-PB. As análises serão realizadas na plataforma R (www.r-project.org), versão 2.10.1.

4.1 Série do Faturamento Comércio Varejista em Geral

A série a ser ajustada será o faturamento do comércio varejista em geral da Grande João Pessoa-PB que percorre o período de tempo de janeiro de 1999 a janeiro de 2010. A série Faturamento do comércio é composta por 133 observações mensais e a fonte de dados é o Instituto Fecomercio de Pesquisas Econômicas, Sociais da Paraíba-IFEP. Primeiramente, vamos analisar o gráfico da série, pois este ira indicar algum comportamento característico da série (tendência, sazonalidade, descontinuidade, entre outros). Observa-se na Figura 1 que a série apresenta tendência podendo ser uma indicação de não estacionariedade e também um comportamento sazonal. A Figura 2 apresenta a decomposição da série segundo a sazonalidade, à tendência e o resíduo.

Figura 1: Faturamento do comércio varejista em geral no período de janeiro de 1999 a janeiro de 2010

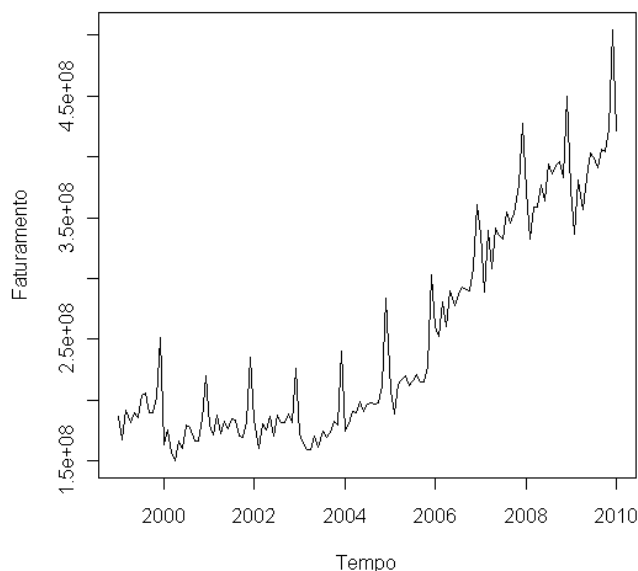
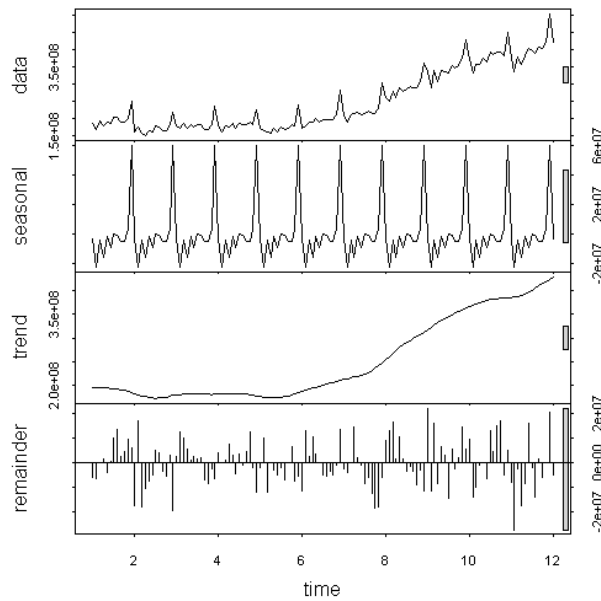
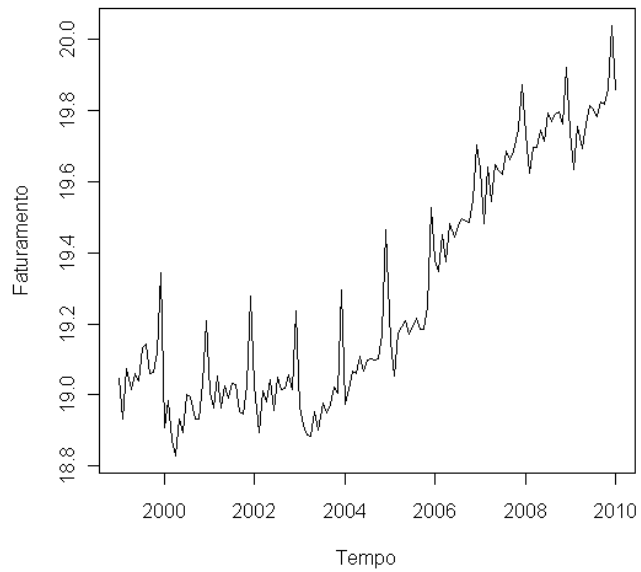


Figura 2: Faturamento do comércio varejista em geral – Decomposição por sazonalidade, tendência e resíduo



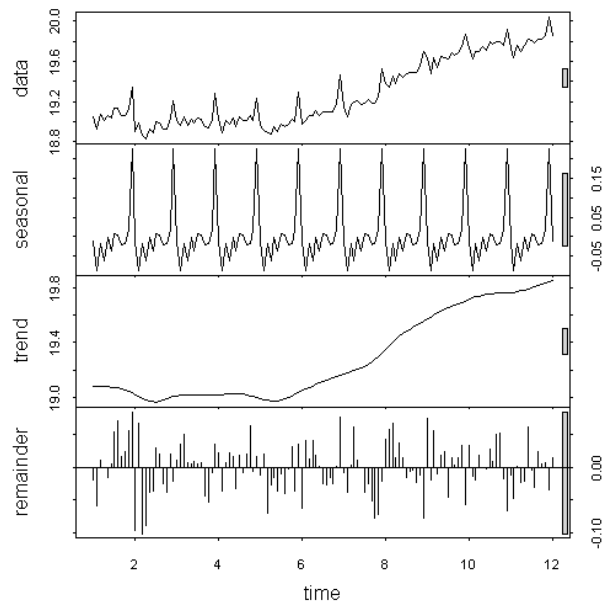
Inicialmente, ao realizamos o ajuste de modelos de séries temporais a partir dos métodos de suavizamento exponencial de Holts-Winters e através da metodologia Box-Jenkins, verificou-se que a hipótese de normalidade dos resíduos para os modelos de Box-Jenkins não foi aceita. O modelo SARIMA(1,1,0)x(1,1,1)₁₂ apresentou um p-valor=0,02153, enquanto que no modelo SARIMA(0,1,2)x(1,1,1)₁₂ obtemos um p-valor = 0,04289. Por este motivo, decidiu-se aplicar uma transformação aos dados da série original. Foi utiliza a família de transformações de Box-Cox (1964), sendo o valor escolhido $\lambda = 0$, o que implica no logaritmo das observações originais, como apresentado na Figura 3.

Figura 3: Logaritmo da série faturamento do comércio varejista em geral no período de janeiro de 1999 a janeiro de 2010



A Figura 4 apresenta a decomposição da série transformada, segundo a sazonalidade, a tendência e o resíduo.

Figura 4: Logaritmo da série faturamento do comércio varejista em geral – Decomposição por sazonalidade, tendência e resíduo



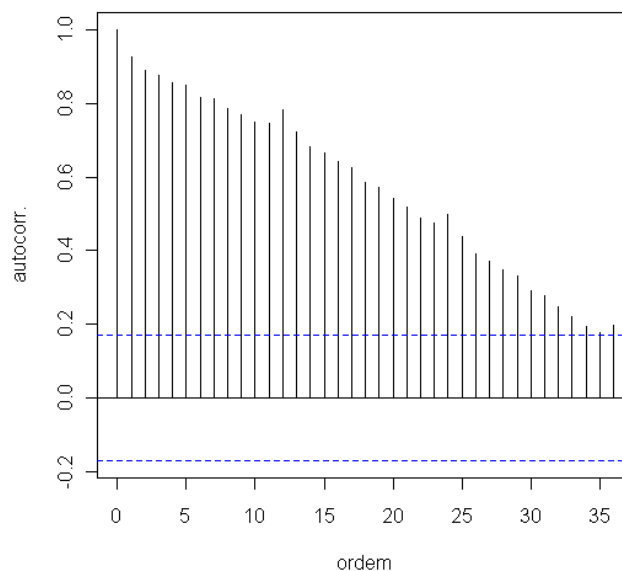
4.2 Suavizamento Exponencial

Para a modelagem e previsão do logaritmo da série através dos métodos de suavizamento exponencial, consideramos o algoritmo de tendência linear de Holt-Winters. Considerando a sazonalidade anual e aditiva, obteve-se com este procedimento $\alpha=0.4343$, $\beta=0.0420$ e $\gamma=0.9185$. Considerando a sazonalidade anual e multiplicativa obteve-se um $\alpha=0,4300$, $\beta=0,0418$ e $\gamma=0.9320$. Como o valor de γ foi alto, verifica-se uma indicação de forte sazonalidade. Por outro lado, como α , β e γ em ambos os modelos são próximos podemos concluir que tanto o modelo de suavização exponencial aditivo como multiplicativo podem ser usado para realizar previsões para a série transformada.

4.3 Modelos de Box-Jenkins

Vamos dar início aos procedimentos de ajuste dos modelos através da metodologia Box-Jenkins. Através das Figuras 3 e 4 temos indícios que a série não apresenta um comportamento estacionário. Para a confirmação desta suposição vamos observar a figura correlograma da série Faturamento do comércio varejista em geral.

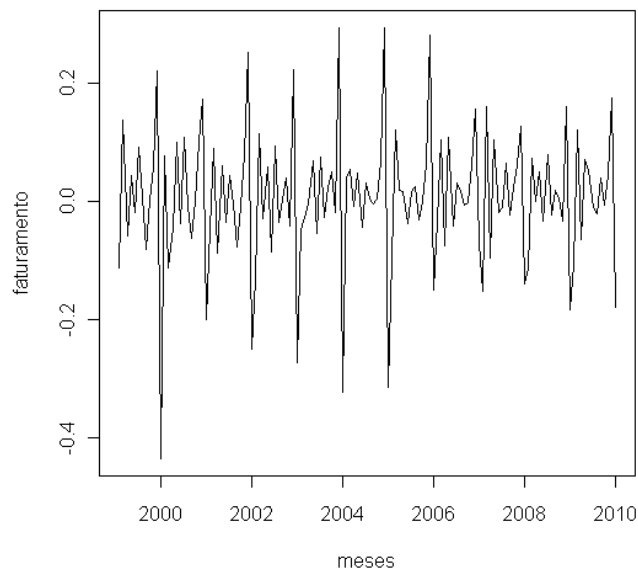
Figura 5: Correlograma do logaritmo da série faturamento do comércio varejista em geral



A Figura 5 mostra o correlograma do logaritmo da série onde as linhas tracejadas são os limites assintóticos de significância ao nível de 95% de confiança. Observamos um decaimento muito lento em direção a zero, ou seja, a série apresenta um comportamento não

estacionário e as ordens múltiplas de 12 se destacam mostrando um comportamento sazonal. Será aplicada uma diferença simples no logaritmo da série Faturamento do comércio varejista para torná-la estacionária. Após aplicar a diferença no logaritmo da série, temos que a mesma apresenta um comportamento estacionário (Figura 6).

Figura 6: Logaritmo da série faturamento do comércio varejista em geral – Série diferenciada



Através da Figura 7, podemos confirmar que o problema da estacionariedade foi resolvido, porém percebemos que o comportamento sazonal da série ainda permanece. Com isto, aplicando uma diferença simples e uma diferença sazonal de modo a torná-la estacionária e eliminando o problema da sazonalidade (Figura 8).

Figura 7: Correlograma do logaritmo da série faturamento do comércio varejista em geral – Série diferenciada

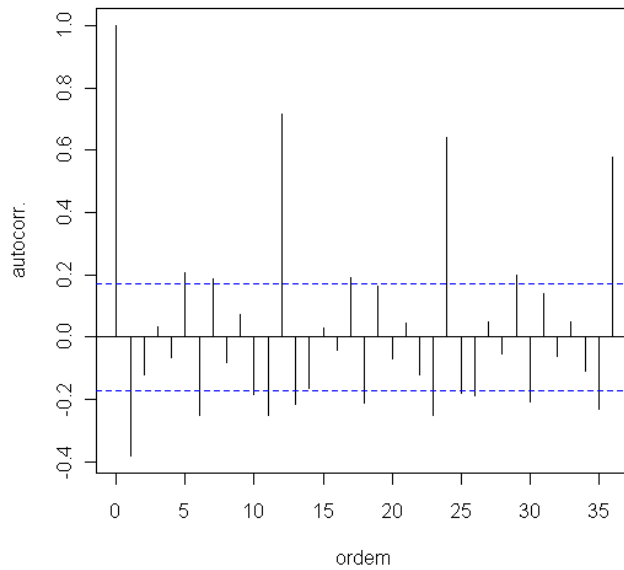
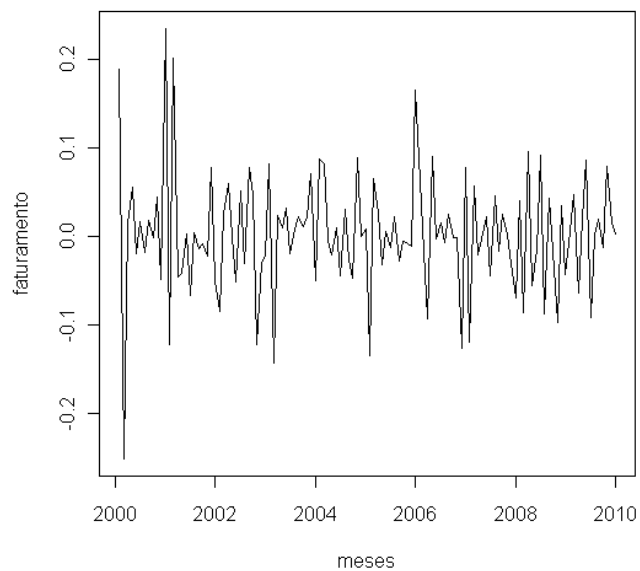


Figura 8: Logaritmo da série faturamento do comércio varejista em geral – Série Diferenciada na Tendência e Sazonalidade



As Figuras 9 e 10 mostram, respectivamente, o correlograma e o correlograma parcial do logaritmo da série com uma diferença simples e uma diferença sazonal. Pelo correlograma, identificamos a ordem média móvel igual a 1 ($MA = 1$). O mesmo ocorre no correlograma parcial, indicando uma ordem auto-regressiva igual a 1 ($AR = 1$).

Gráfico 9: Correlograma do Logaritmo da série faturamento do comércio varejista em geral - Série Diferenciada na Tendência e Sazonalidade.

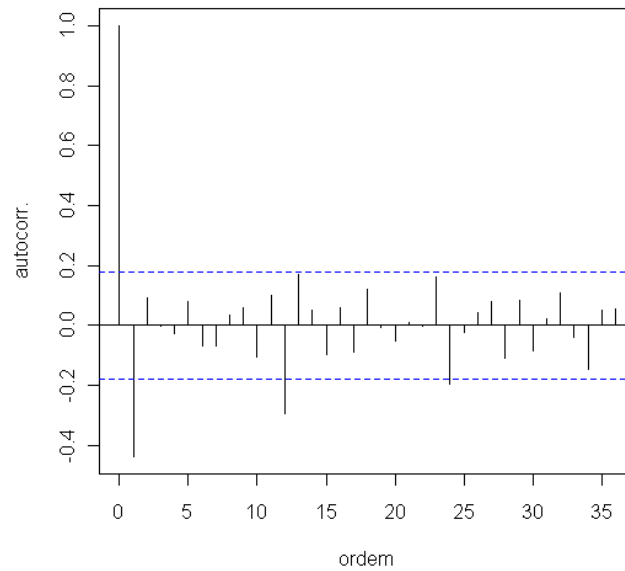
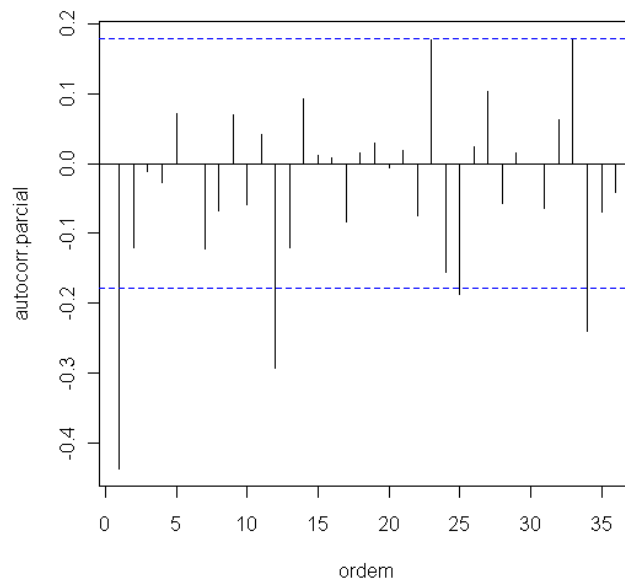


Gráfico 10: Correlograma parcial do logaritmo da série faturamento do comércio varejista em geral - Série Diferenciada na Tendência e Sazonalidade



Com base nas características apresentadas pelos correlogramas da série, usaremos um modelo SARIMA, apresentando uma diferença simples ($d=1$) e uma diferença sazonal ($D=1$).

A partir dos resultados acima apresentados, foram estimados vários modelos de séries temporais em função das ordens p , q , P e Q entre 0 e 2, e $D = 0$ ou 1. Para a identificação do melhor modelo foram utilizados os critérios AIC, AIC_c e BIC.

Tabela 2: Critérios AIC, AIC_c e BIC para os modelos SARIMA

Modelos	Critérios		
	AIC	AIC_c	BIC
SARIMA(1,1,1) \times (1,0,1) ₁₂	-379,05	-378,57	-364,63
SARIMA(2,1,0) \times (2,0,2) ₁₂	-377,53	-376,63	-357,35
SARIMA(0,1,1) \times (1,0,2) ₁₂	-376,65	-376,18	-362,24
SARIMA(2,1,1) \times (1,0,1) ₁₂	-377,38	-376,71	-360,08
SARIMA(2,1,2) \times (1,0,1) ₁₂	-375,30	-374,39	-355,12

De acordo com os resultados apresentados na Tabela 2, o modelo SARIMA(1,1,1) \times (1,0,1)₁₂ apresentou os menores AIC, AIC_c e BIC. Enquanto que o modelo SARIMA(2,1,2) \times (1,0,1)₁₂ apresentou os maiores AIC, AIC_c e BIC. Como o critério de escolha para a comparação dos modelos pelos erros de previsão se aplica aos que apresentarem os menores critérios, será dispensado o modelo SARIMA(2,1,2) \times (1,0,1)₁₂.

Através da análise de diagnóstico, podemos considerar que um modelo de séries temporais está bem ajustado quando o gráfico dos resíduos padronizados apresenta-se entre -2 e 2, o gráfico do correlograma dos resíduos encontra-se dentro dos limites assintóticos e

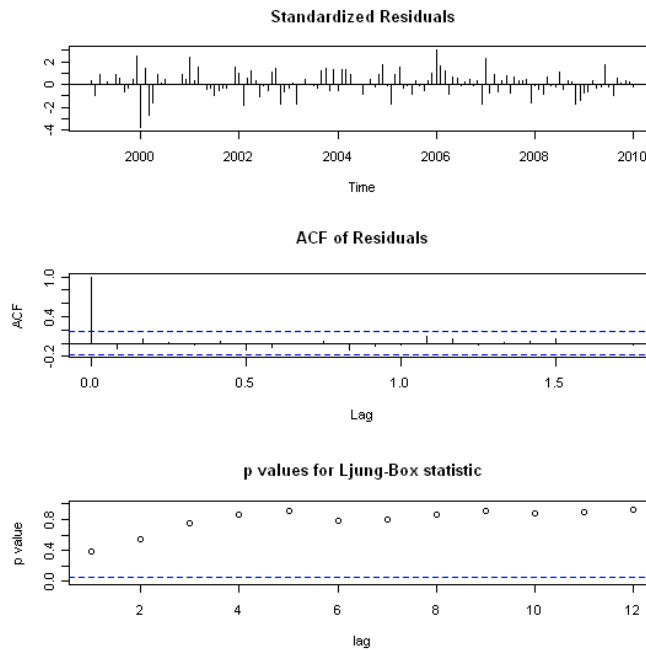
quando a estatística de Ljung-Box apresenta p-valor acima do nível de significância α para os diferentes *lags*.

Análise residual do modelo $SARIMA(1,1,1) \times (1,0,1)_{12}$ é apresentada na Figura 11. Observamos que a maior parte dos resíduos padronizados encontram-se dentro do intervalo de -2 a 2, sinalizando que o modelo apresentou um bom ajuste aos dados. O segundo gráfico apresenta o correlograma dos resíduos e também os limites assintóticos de significância ao nível de 95%. Observamos que as autocorrelações amostrais estão dentro dos limites assintóticos. Desta forma, podemos afirmar que os resíduos não estão correlacionados. Chegamos a esta mesma conclusão ao analisar os p-valores para o teste de Ljung-Box.

A análise residual para os modelos $SARIMA(2,1,0) \times (2,0,2)_{12}$, $SARIMA(2,1,1) \times (1,0,1)_{12}$ e $SARIMA(0,1,1) \times (1,0,2)_{12}$ apresentou resultados semelhantes obtidos para o modelo $SARIMA(1,1,1) \times (1,0,1)_{12}$.

Em seguida, foi realizado o teste de normalidade de Lilliefors, a um nível de significância de 5%, para os modelos acima citados. Para o modelo $SARIMA(1,1,1) \times (1,0,1)_{12}$ temos um p-valor = 0,1185. O modelo $SARIMA(2,1,0) \times (2,0,2)_{12}$ apresentou um p-valor = 0,0626, enquanto que no modelo $SARIMA(2,1,1) \times (1,0,1)_{12}$ obtemos um p-valor = 0,0960. Por último, o modelo $SARIMA(0,1,1) \times (1,0,2)_{12}$ obteve um p-valor = 0,0799. A partir destes resultados, concluímos que a hipótese de normalidade dos resíduos é aceita em todos os modelos.

Figura 11: Análise de diagnóstico do modelo $SARIMA(1,1,1) \times (1,0,1)_{12}$



4.4 Comparação dos modelos

Após a análise de diagnóstico apresentamos, para os modelos SARIMA e alisamento exponencial Holt-Winters, os seguintes erros de previsão: Erro percentual Médio (EPM), Raiz do Erro Quadrático Médio (REQM) e Erro Absoluto Médio (EAM), para três, seis e doze passos a frente, como pode ser visto na Tabela 3. Pelos valores dos erros de previsão EPM, REQM e EAM observamos que a três passos à frente o modelo SARIMA(2,1,1)x(1,0,1)₁₂ apresentou a melhor performance de previsão. O modelo SARIMA(2,1,0)x(2,0,2) mostrou a melhor previsão seis passos à frente. A doze passos à frente, temos a melhor previsão para o modelo SARIMA(1,1,1)x(1,0,1)₁₂ pelos critérios EPM, REQM e EAM.

Tabela 3: Valores dos erros de previsão EPM, REQM e EAM, para a série Faturamento do comércio varejista em geral transformada a 3, 6 e 12 passos à frente.

Previsões	Modelos	EPM	REQM	EAM
h=3	Holt-Winters Aditivo	0,2138	0,0481	0,0426

	Holt-Winters Multiplicativo	0,2277	0,0504	0,0454
	SARIMA(1,1,1)x(1,0,1)	0,0813	0,0175	0,0162
	SARIMA(2,1,1)x(1,0,1)	0,0718	0,0155	0,0143
	SARIMA(0,1,1)x(1,0,2)	0,1326	0,0265	0,0264
	SARIMA(2,1,0)x(2,0,2)	0,1364	0,0274	0,0272
h=6	Holt-Winters Aditivo	0,2939	0,0665	0,0584
	Holt-Winters Multiplicativo	0,3034	0,0686	0,0603
	SARIMA(1,1,1)x(1,0,1)	-0,0478	0,0222	0,0149
	SARIMA(2,1,1)x(1,0,1)	-0,0610	0,0235	0,0153
	SARIMA(0,1,1)x(1,0,2)	0,0334	0,0211	0,0190
	SARIMA(2,1,0)x(2,0,2)	-0,0004	0,0225	0,0185
h=12	Holt-Winters Aditivo	-0,0602	0,0507	0,0440
	Holt-Winters Multiplicativo	-0,0692	0,0515	0,0444
	SARIMA(1,1,1)x(1,0,1)	0,0327	0,0285	0,0244
	SARIMA(2,1,1)x(1,0,1)	0,0454	0,0291	0,0251
	SARIMA(0,1,1)x(1,0,2)	0,0499	0,0335	0,0285
	SARIMA(2,1,0)x(2,0,2)	0,0138	0,0326	0,0276

4.5 Previsões Valores Futuros

A Tabela 4 apresenta as previsões e os respectivos intervalos de confiança, ao nível de 95% de confiança, para a série de faturamento do comércio varejista em geral três passos à frente, segundo o modelo SARIMA(2,1,1)x(1,0,1)₁₂.

Tabela 4: Previsões três passos à frente - modelo SARIMA(2,1,1)x(1,0,1)₁₂.

Mês/Ano	Previsão	Limite inferior	Limite Superior
Fevereiro/2010	387.406.956	350.776.309	427.862.845
Março/2010	426.352.372	381.398.691	476.604.534
Abril/2010	408.884.427	358.708.095	466.079.459

A Tabela 5 mostra as previsões para a série faturamento do comércio varejista em geral seis passos à frente, segundo o modelo SARIMA(2,1,0)x(2,0,2)₁₂.

Tabela 5: Previsões seis passos à frente - modelo SARIMA(2,1,0)x(2,0,2)₁₂.

Mês/Ano	Previsão	Limite inferior	Limite Superior
Fevereiro/2010	382.043.033	346353664	421409948
Março/2010	425.895.870	381907280	474951124
Abril/2010	404.230.210	356153218	458797098

Maio/2010	432.472.921	375851164	497624712
Junho/2010	434.341.738	372727099	506141748
Julho/2010	438.649.084	372030721	517196587

Por último, a Tabela 6 apresenta as previsões para a série faturamento do comércio varejista em geral, doze passos à frente, segundo o modelo SARIMA(1,1,1)x(1,0,1)₁₂.

Tabela 6: Previsões doze passos à frente- modelo SARIMA(1,1,1)x(1,0,1)₁₂.

Mês/Ano	Previsão	Limite inferior	Limite Superior
Fevereiro/2010	386.810.890	350154401	427304824
Março/2010	426.506.710	381555944	476753086
Abril/2010	408.656.378	358995927	465186435
Maio/2010	436.607.636	378376772	503800026
Junho/2010	434.462.261	371619213	507932448
Julho/2010	444.944.387	376095931	526396302
Agosto/2010	444.892.580	371883481	532235008
Setembro/2010	448.793.154	371235522	542553939
Outubro/2010	450.926.356	369314222	550573378
Novembro/2010	464.727.191	377034424	572816030
Dezembro/2010	552.290.496	444038665	686932956
Janeiro/2011	466.589.785	371890461	585403634

Recentemente, foram disponibilizados o faturamento do comércio varejista em geral para os meses de fevereiro, março e abril de 2010. Pode-se verificar na Tabela 7, que os valores previstos pelos modelos estiveram próximos aos valores reais. Vale ressaltar, que o modelo SARIMA(2,1,1)x(1,0,1)₁₂ apresentou o menor erro percentual médio, em torno de 2,61%, enquanto que os modelos SARIMA(1,1,1)x(1,0,1)₁₂ e SARIMA(2,1,0)x(2,0,2)₁₂ apresentaram, respectivamente, erros percentuais médios em torno de 2,66% e 3,44%, conforme apresentado na Tabela 8.

Tabela 7: valores observados e os valores estimados dos modelos SARIMA-
Fev-Abr 2010

Mês/Ano	Valores mensais (em R\$)	SARIMA(2,1,1)x(1,0,1)₁₂	SARIMA(2,1,0)x(2,0,2)₁₂	SARIMA(1,1,1)x(1,0,1)₁₂
Fev/2010	390.951.990	387.406.956	382.043.033	386.810.890
Mar/2010	451.306.385	426.352.372	425.895.870	426.506.710
Abr/2010	413.151.021	408.884.427	404.230.210	408.656.378

Tabela 8: Erro percentual médio (EPM) para os valores observados e os valores estimados dos modelos SARIMA-Fev-Abr 2010

Modelos	EPM
SARIMA(2,1,1)x(1,0,1) ₁₂	2,61
SARIMA(1,1,1)x(1,0,1) ₁₂	2,66
SARIMA(2,1,0)x(2,0,2) ₁₂	3,44

5. CONSIDERAÇÕES FINAIS

De acordo com os objetivos estabelecidos no presente trabalho, os modelos de previsão aqui apresentados obedeceram a todos os pressupostos teóricos necessários.

Foram utilizadas a metodologia de Suavizamento exponencial de Holt-Winters e a metodologia Box-Jenkins, através dos modelos SARIMA, para a série do faturamento do comércio geral na Grande João Pessoa, a partir das informações obtidas no período de janeiro de 1999 a janeiro de 2010.

Observamos que a metodologia de suavizamento exponencial, através do algoritmo de Holt-Winters, apresentou uma menor precisão em suas estimativas, enquanto que para os modelos obtidos através da metodologia de Box-Jenkins apresentaram os melhores resultados. Para tanto, foram utilizados os critérios de comparação AIC, AIC_c e BIC para a seleção dos modelos. A comparação dos erros de previsão (EPM, REQM e EAM) à três, seis e doze passos à frente, respectivamente, apontou como os melhores modelos: SARIMA(2,1,1)x(1,0,1)₁₂, SARIMA(2,1,0)x(2,0,2)₁₂, SARIMA(1,1,1)x(1,0,1)₁₂, respectivamente.

Após a conclusão deste estudo, estavam disponíveis os dados sobre o faturamento do comércio varejista em geral nos meses de fevereiro/2010, março/2010 e abril/2010. As previsões fornecidas pelos modelos SARIMA estiveram próximos aos valores reais. Como já era esperado, o modelo SARIMA(2,1,1)x(1,0,1)₁₂, que realizou as previsões três passos à frente, apresentou os melhores resultados com um erro percentual médio de 2,61%.

REFERENCIAL BIBLIOGRÁFICO

- BOX, G. E. P. AND COX, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B*, 26, 211-243.
- BOX, G. E. P., JENKINS, G. M. AND REINSEL, G. C. *Time Series Analysis, Forecasting and Control*, 3rd ed., Prentice Hall, Englewood Cliffs, 1994.
- BOX, G.E.P. & JENKINS, G.M., *Time Series Analysis, Forecasting and Control*. Holden-Day, San Francisco, California, 1970.
- EHLERS, R.S., *Análise de séries temporais*, 2007.
- HURVICH, C.M.; TSAI, C.L., *Regression and time series model selection in small samples*. *Biometrika*, 76, 297–307, 1989.
- MORRETIN, P. A.; TOLOI, C. M. C., *Análise de séries temporais*. 2. ed. São Paulo: Atual Editora, 2005.
- NELSON, H. L. (1976). *The Use of Box-Cox Transformations in Economic Time Series Analysis: An Empirical Study*. Doctoral Dissertation, Univ. of California, San Diego.
- PEREIRA, T. L. ; SOUZA, T. C. E. ; CASSIANO, K. M. . Comparação de Métodos de Previsão: Uma Aplicação ao IGP-DI. In: 17o Simpósio nacional de Probabilidade e Estatística, 2006, Caxambu - MG. Anais do 170 SINAPE.
- PIERCE, D. A. (1979). Seasonal adjustment when both deterministic and stochastic seasonality are present. *In Seasonal Analysis of Economic Time Series* (Arnold Zellner, Ed.), 242-269. Washington, D. C., U. S. Dept. of Commerce, Bureau of the Census.
- SILVA, LUZIA MARIA O.; SOUZA, REINALDO CASTRO (orientador). *Uma Aplicação de Árvores Neurais e KNN para a Identificação de Modelos ARMA Não-sazonais e Sazonais*. Rio de Janeiro, 2005, 15-40. Tese de Doutorado-Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.