

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Exatas
Especialização em Estatística: Ênfase em Estatística

Felipe Barbosa de Sá

Uma Comparação dos Modelos SARIMA e Holt-Winters para a previsão do Desvio do Custo Relativo de Manutenção dos Veículos da Frota de uma Empresa de Distribuição de Energia Elétrica Brasileira.

Belo Horizonte

2021

Felipe Barbosa de Sá

Uma Comparação dos Modelos SARIMA e Holt-Winters para a previsão do Desvio do Custo Relativo de Manutenção dos Veículos da Frota de uma Empresa de Distribuição de Energia Elétrica Brasileira.

Monografia apresentada ao Programa de Especialização em Estatística: Ênfase em Estatística da Universidade Federal de Minas Gerais.

Orientador: Thiago Rezende dos Santos

Belo Horizonte

2021

Sá, Felipe Barbosa de.

S111u Uma comparação dos modelos SARIMA e Holt-Winters para a previsão do desvio do custo relativo de manutenção dos veículos da frota de uma empresa de distribuição de energia elétrica brasileira.[manuscrito]: / Felipe Barbosa de Sá. — 2021.
v,23.f. il.

Orientador: Thiago Rezende dos Santos.

Monografia (especialização) - Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Estatística.

Referências 22-23.

1. Estatística. 2. Análise de séries temporais. 3. Energia elétrica, Distribuição de - Equipamentos, acessórios, etc - Manutenção e reparos – Custos. 4; Previsão. I. Santos, Thiago Rezende dos.II. Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Estatística .III.Título.

CDU 519.2 (043)



Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Estatística

E-mail: pgest@ufmg.br

Tel: 3409-5923 – FAX: 3409-5924

Programa de Pós-Graduação / Especialização

Av. Pres. Antônio Carlos, 6627 - Pampulha

31270-901 – Belo Horizonte – MG

ATA DO 221ª. TRABALHO DE FIM DE CURSO DE ESPECIALIZAÇÃO EM ESTATÍSTICA DE FELIPE BARBOSA DE SÁ.

Aos oito dias do mês de março de 2021, às 17:00 horas, com utilização de recursos de videoconferência a distância, reuniram-se os professores abaixo relacionados, formando a Comissão Examinadora homologada pela Comissão do Curso de Especialização em Estatística, para julgar a apresentação do trabalho de fim de curso do aluno **Felipe Barbosa de Sá**, intitulado: “*Uma Comparação dos Modelos SARIMA e Holt-Winters para a previsão do Desvio do Custo Relativo de Manutenção dos Veículos da Frota de uma Empresa de Distribuição de Energia Elétrica Brasileira*”, como requisito para obtenção do Grau de Especialista em Estatística. Abrindo a sessão, o Presidente da Comissão, Professor Thiago Rezende dos Santos – Orientador, após dar conhecimento aos presentes do teor das normas regulamentares, passou a palavra ao candidato para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa do candidato. Após a defesa, os membros da banca examinadora reuniram-se sem a presença do candidato e do público, para julgamento e expedição do resultado final. Foi atribuída a seguinte indicação: o candidato foi considerado. Aprovado condicional às modificações sugeridas pela banca examinadora no prazo de 30 dias a partir da data de hoje por unanimidade. O resultado final foi comunicado publicamente ao candidato pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente Ata, que será assinada por todos os membros participantes da banca examinadora. Belo Horizonte, 08 de março de 2021.

Prof. Thiago Rezende dos Santos (Orientador)
Departamento de Estatística / UFMG

Jussiane Nader
Gonçalves

Assinado de forma digital por
Jussiane Nader Gonçalves
Dados: 2021.03.08 19:50:59 -03'00'

**Profa. Jussiane Nader
Gonçalves**
**Departamento de
Estatística/UFMG**

Prof. Arthur Tarso Rego
PUC Minas

Resumo

Em tempos de muita incerteza econômica, realizar boas previsões pode ajudar bastante no planejamento financeiro das empresas daqui em diante. Em face disto, o presente trabalho comparou os modelos *Seasonal Autoregressive Integrated Moving Average* SARIMA e Holt-Winters para prever o desvio do custo de manutenção relativo, mensurado em R\$ por Km, da frota de uma empresa do setor de distribuição de energia elétrica brasileira. As previsões via Holt-Winters, para a série temporal proposta, tiveram um desempenho melhor, segundo os critérios das estatísticas de erro Erro Médio Absoluto (EMA), Erro Médio Percentual Absoluto (EMPA) e Erro Quadrático Médio (EQM). Outros modelos poderão ser incorporados e comparados em trabalhos futuros.

Palavras Chave: Série Temporal, Desvio do Custo de Manutenção Relativo, Previsão.

Abstract

In times of great economic uncertainty, performing good predictions can help the financial planning of companies going forward. Therefore, the present study compared the Seasonal Autoregressive Integrated Moving Average SARIMA and Holt-Winters models to predict the deviation of the relative maintenance cost, measured in R\$ per Km, of the fleet of a Brazilian company in the electricity distribution sector. The Holt-Winters results, for a proposed time series, performed better, according to the criteria of Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE). Other model models will be incorporated and compared in future works.

Key Words: Time Series, Deviation of the Relative Maintenance Cost, Forecast.

Lista de Figuras:

Figura 1: Histograma do desvio do custo relativo de manutenção.	10
Figura 2: Teste de normalidade de Shapiro-Wilk do desvio do custo relativo de manutenção.	11
Figura 3: Gráfico de probabilidade normal do desvio do custo relativo de manutenção.	11
Figura 4: Teste de Dickey-Fuller da série temporal diferenciada.	12
Figura 5: Teste de Dickey-Fuller aumentado com tendência da série temporal diferenciada.	12
Figura 6: Teste de Phillips-Perron da série temporal diferenciada.	12
Figura 7: Gráficos da série temporal, FAC e FACP dos dados originais.	13
Figura 8: Gráficos da série temporal, FAC e FACP dos dados transformados.	14
Figura 9: Teste de normalidade dos resíduos do modelo SARIMA(1,1,0)x(1,0,0) ₁₂	15
Figura 10: Série temporal e estimativas pontuais e intervalares (previsão) do modelo SARIMA(1,1,0)x(1,0,0) ₁₂ no período de 9 meses.	17
Figura 11: Série temporal e estimativas pontuais e intervalares (previsão) do modelo de Holt-Winters no período de 9 meses.	18
Figura 12: Comparativo dos modelos de previsão para o período de validação de 9 meses.	19

Lista de Tabelas:

Tabela 1: Estatísticas descritivas do desvio do custo relativo de manutenção.	10
Tabela 2: Verificação dos modelos ARIMA através do comparativo das estatísticas AIC e BIC.	15
Tabela 3: Parâmetros estimados do modelo SARIMA(1,1,0)x(1,0,0) ₁₂ dos dados referente ao desvio do custo relativo de manutenção.	15
Tabela 4: Teste de autocorrelação dos resíduos do modelo SARIMA(1,1,0)x(1,0,0) ₁₂	16
Tabela 5: Acurácia do modelo SARIMA(1,1,0)x(1,0,0) ₁₂	16
Tabela 6: Estimativas pontuais e intervalares do modelo SARIMA(1,1,0)x(1,0,0) ₁₂ no período de 9 meses.	16
Tabela 7: Parâmetros estimados do modelo de Holt-Winters dos dados referente ao desvio do custo relativo de manutenção.	17
Tabela 8: Acurácia do modelo de Holt-Winters.	18
Tabela 9: Estimativas pontuais e intervalares do modelo de Holt-Winters no período de 9 meses.	18
Tabela 10: Comparativo entre as estatísticas de acurácia dos modelos SARIMA(1,1,0)x(1,0,0) ₁₂ e de Holt-Winters no período de 9 meses.	19

Sumário

1. Introdução	1
2. Métodos	4
2.1 Modelos ARMA.....	4
2.1.1 Modelos Autorregressivos (AR).....	5
2.1.2 Médias Móveis (MA).....	5
2.1.3 Modelos Autorregressivos e de Médias Móveis (ARMA).....	6
2.2.1 Modelos Autorregressivos Integrados e de Médias Móveis (ARIMA).....	6
2.2.2 Modelos Sazonais Autorregressivos Integrados e de Médias Móveis (SARIMA).....	7
2.2.3 Identificação da Ordem do Modelo ARMA.....	7
2.3 Modelo de Holt-Winters	7
2.4 Estatísticas de Erro de Previsão – Acurácia do Modelo.....	8
2.5 Base de Dados	8
3 Resultados	10
3.1 Análise Descritiva	10
3.2 Modelagem Box e Jenkins.....	14
3.3 Modelo de Holt-Winters	17
3.4 Comparativo entre os modelos	19
4 Conclusão	21
5 Referências Bibliográficas.....	22

1. Introdução

No cenário econômico atual, o grau de incerteza cresceu significativamente no tocante às projeções e no planejamento financeiro de todas as empresas ao redor do mundo. Os impactos desta incerteza afetam variáveis como renda, emprego, tributação, dentre outros.

Quando avaliamos o cenário de incerteza na visão das empresas, precisamos refletir todas as variáveis que impactam suas operações corriqueiras, sejam elas direta ou indiretamente ligadas ao processo produtivo do empreendimento. É necessário, portanto, realizar projeções financeiras assertivas a fim de dar continuidade no negócio.

Existem diversas metodologias que podem ser utilizadas na construção de uma previsão, sendo necessário avaliar qual modelo é o ideal para a série temporal de interesse. Desta forma, são muitos os trabalhos que estudam a idiosincrasia de uma série temporal buscando prever seu comportamento para os próximos períodos, comparando os diversos métodos mais largamente utilizados. Doganis *et al* (2006) utilizou métodos combinados de aprendizado de máquinas avançadas, mais precisamente o *Radial Basis Function* (RBF) e *Genetic Algorithms* (GA), para construir modelos de séries temporais não lineares para previsão de vendas de produtos altamente perecíveis. A combinação trouxe resultados bastante precisos, podendo ser aplicados não só em uma melhoria no planejamento das indústrias alimentícias, bem como na redução dos impactos ambientais.

Cranage e Andrew (1992) compararam modelos econométricos convencionais com os de séries temporais, mais precisamente $ARIMA(p,d,q)$ e Suavização Exponencial, e identificaram que estes últimos foram tão precisos quanto o primeiro, ou até melhores, tendo em vista que possuem um custo de aplicação menor (coleta de dados, tempo e etc.). Dentre os dois modelos de séries temporais, o de $ARIMA(p,d,q)$ mostrou melhores resultados, apesar de um custo de estimação maior.

Ao comparar o modelo SARIMA com os Modelos Estruturais para prever casos de incidência da Dengue em Belo Horizonte, Santos (2017) observou que para o período estudado no trabalho o segundo apresentou melhores resultados que o primeiro, quando analisados os critérios de Erro Quadrático Médio (EQM) e Erro Absoluto Médio (EAM).

Para prever casos de Malária e Hepatite A nos EUA, Nobre *et al* (2001) comparou o modelo SARIMA e o Modelo Linear Dinâmico obtendo resultados parecidos para ambos quando uma série temporal longa está disponível. Porém, o Modelo Linear Dinâmico é preferível para diferentes tipos de séries temporais, em

função de não ser necessário novos ciclos de identificação e modelagem quando uma nova base de dados entrar no modelo.

Harvey e Todd (1983) compararam os resultados de um estudo de previsões que utilizou o modelo SARIMA com o Modelo Estrutural por eles aplicado, em uma série temporal de seis variáveis macroeconômicas do Reino Unido. Apesar de que para algumas séries o modelo SARIMA se ajustou melhor, e para outras o modelo estrutural apresentou menores estatísticas de erro, eles concluíram que o segundo é mais atrativo em função de vantagens conceituais, quando comparado ao primeiro.

Com o intuito de mensurar o *Value at Risk* (*VaR*) do preço de curto prazo da eletricidade para cinco grandes mercados mundiais: Victoria (Austrália), NordPool (Escandinávia), Alberta (Canadá), Hayward (Nova Zelândia) e PJM (EUA), Chan e Gray (2006) utilizaram a teoria de valor extremo e o modelo EGARCH obtendo bons resultados, tendo em vista as características complexas que os retornos deste mercado apresentam.

Buscando identificar modelos para prever as vendas agregadas do setor de varejo dos Estados Unidos, Alon *et al* (2001) constatou que os modelos de Redes Neurais Artificiais são mais eficientes que os mais tradicionalmente utilizados, sendo eles $ARIMA(p,d,q)$, Suavização Exponencial e Regressão Multivariada. No entanto, ressaltou que o modelo de suavização é altamente viável quando a situação econômica do país está relativamente estável e quando a série temporal disponível é pequena.

Já Sfetsos (1999) identificou que o modelo de Redes Neurais Lógicas produziu bons resultados na previsão da velocidade média do vento por hora.

Lippi *et al* (2013) constatou que o modelo SARIMA, combinado com o filtro de Kalman, trouxe melhores previsões das condições de tráfego de curto prazo em uma rede de transporte do que outros modelos testados, tendo em vista que a sazonalidade se mostrou como uma característica fundamental da série estudada.

O presente trabalho tem o intuito de construir projeções para o desvio percentual no custo relativo de manutenção da frota de uma empresa brasileira do setor de distribuição de energia elétrica (realizado *versus* orçado), buscando prever os gastos que serão realizados nos próximos meses.

O setor de distribuição de energia elétrica, possui influência significativa na economia brasileira, tendo em vista a relação direta com os demais setores, além do consumo das famílias. Um ativo vital no processo de distribuição são os veículos utilizados na manutenção e conservação das redes e subestações de energia elétrica, tendo este uma significativa relevância na estrutura de custos das empresas deste setor. Portanto, faz-se importante avaliar os impactos do atual cenário econômico no

custo de manutenção da frota, tendo em vista que ele é altamente influenciado por mudanças na conjuntura econômica em geral.

Foram comparados os modelos de SARIMA e suavização exponencial de Holt-Winters a fim de identificar o que melhor se enquadra na série temporal proposta pelo presente trabalho.

Na Seção 2 estão apresentadas as metodologias $ARIMA(p,d,q)$ e Suavização Exponencial de Holt-Winters, além de uma breve descrição dos dados utilizados. Na Seção 3 estão os resultados do trabalho e na Seção 4 a conclusão.

2. Métodos

Os métodos utilizados e comparados no presente trabalho são o modelo SARIMA e o modelo de Suavização Exponencial de Holt-Winters.

2.1 Modelos ARMA

BOX e JENKINS (1976), propuseram um modelo que ficou largamente conhecido na literatura e nos mercados que trabalham com séries temporais, o modelo ARMA. Esta metodologia consiste, basicamente, em ajustar um modelo autorregressivo e de médias móveis a um dado conjunto de dados. Segundo Morettin (2006), a construção do modelo será baseada em um ciclo iterativo, em que o próprio conjunto de dados mostrará qual o modelo ideal a ser ajustado. De acordo com ele, os estágios deste ciclo são:

Especificação: classe de modelos que é considerada para análise;

Identificação: A identificação do modelo é feita com base na autocorrelação, autocorrelação parcial e demais critérios;

Estimação: Estimam-se os parâmetros do modelo, a partir da identificação;

Diagnóstico: É feita uma análise residual para averiguar se o modelo é adequado, a modelagem e a previsão. Caso o modelo não seja adequado, o ciclo é repetido, voltando-se a fase da identificação.

A fase mais crítica do procedimento listado acima é o da identificação, tendo em vista que para uma mesma série temporal, é possível identificar vários modelos diferentes. Caso não seja possível identificar as ordens do modelo ARIMA, pode-se encontra-las através do auto ARIMA, avaliando os critérios de informação como o Bayesian Information Criterion (BIC) ou Akaike Information Criterion (AIC).

2.1.1 Modelos Autorregressivos (AR)

Seja $\{x_t\}$, com $t \in \mathbb{N}$, uma série temporal a tempo discreto. Segundo Shumway e Stoffer (2010), para uma série temporal x_t , um modelo autorregressivo de ordem p , denotado por $AR(p)$, é da seguinte forma:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + w_t. \quad (1)$$

O modelo $AR(p)$ pressupõe que a variável dependente x_t é influenciada pelos seus valores passados $x_{t-1}, x_{t-2}, \dots, x_{t-p}$. A variável w_t representa o ruído branco gaussiano.

O modelo também pode ser escrito através do operador de retardo, tal como segue:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)x_t = w_t. \quad (2)$$

em que o operador de retardos é definido como:

$$B^j = \frac{x_{t-j}}{x_t}, \quad j = 1, 2, \dots, p$$

De forma resumida, o modelo pode ser escrito da seguinte forma:

$$\phi(B)x_t = w_t.$$

2.1.2 Médias Móveis (MA)

O modelo de média móvel de ordem q , para uma série temporal x_t , de acordo com Shumway e Stoffer (2010), denotado por $MA(q)$, é definido como:

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \dots + \theta_q w_{t-q}. \quad (3)$$

Esse modelo assume que o ruído branco do lado direito da equação possui combinação linear com os dados observados.

Da mesma forma que o modelo $AR(p)$, o $MA(q)$ também pode ser reescrito da seguinte maneira:

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q. \quad (4)$$

em que o operador de retardos é definido como:

$$B^j = \frac{w_{t-j}}{w_t}, \quad j = 1, 2, \dots, p$$

De forma resumida,

$$x_t = \theta(B)w_t.$$

2.1.3 Modelos Autorregressivos e de Médias Móveis (ARMA)

A combinação do modelo autorregressivo de ordem p e o de média móvel de ordem q , para uma série temporal x_t , denotado por ARMA(p, q), possui a seguinte expressão:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \dots + \theta_q w_{t-q} \quad (5)$$

Se x_t possui média μ diferente de zero, então $\alpha = \mu(1 - \phi_1 - \dots - \phi_p)$. Desta forma, escrevemos o modelo como:

$$x_t = \alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \dots + \theta_q w_{t-q}.$$

O modelo também pode ser reescrito da seguinte forma:

$$\phi(B)x_t = \theta(B)w_t.$$

2.2.1 Modelos Autorregressivos Integrados e de Médias Móveis (ARIMA)

O modelo ARIMA amplia a classe dos modelos ARMA, onde são incluídas diferenciações na série de interesse. Um processo/série temporal x_t é chamado de ARIMA(p, d, q) se $\nabla^d x_t = (1 - B)^d x_t$ é um ARMA(p, q). Desta forma, o modelo pode ser escrito como:

$$\phi(B)(1 - B)^d x_t = \theta(B)w_t \quad (6)$$

Se a $E(\nabla^d x_t) = \mu$, o modelo é escrito como a seguir:

$$\phi(B)(1 - B)^d x_t = \phi_0 + \theta(B)w_t \quad (7)$$

Onde $\phi_0 = \mu(1 - \phi_1 - \dots - \phi_p)$.

2.2.2 Modelos Sazonais Autorregressivos Integrados e de Médias Móveis (SARIMA)

Considerando a sazonalidade multiplicativa no modelo $ARIMA(p,d,q)$ para uma série temporal x_t , iremos obter um modelo definido como $ARIMA(p,d,q) \times (P,D,Q)_s$ que é dado por:

$$\Phi_p(B^s)\phi(B)\nabla_s^D\nabla^d x_t = \phi_0 + \Theta_Q(B^s)\theta(B)w_t, \quad (8)$$

onde

$$\nabla_s^D = (1 - B^s)^D.$$

2.2.3 Identificação da Ordem do Modelo ARMA

Para identificarmos a ordem do modelo $ARMA(p,q)$ deve-se analisar os gráficos de autocorrelações (FAC) e autocorrelações parciais (FACP) ou o Critério de Informação. A ideia do auto arima, segundo Bueno (2012), é minimizar uma função baseada nos resíduos, penalizada pelo número de regressores no modelo. Ele lista os três principais critérios de informação a serem utilizados: Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC) ou Hannan-Quinn (HQ).

2.3 Modelo de Holt-Winters

Segundo Bueno (2012), existem três formas de suavização de Holt-Winters, o que depende se a série é sazonal ou não, se a sazonalidade é multiplicativa e se há tendência na série.

- i) Sem sazonalidade e com tendência linear:

$$x_t = \alpha y_t + (1 - \alpha)(x_{t-1} + z_{t-1}), \quad (9)$$

$$z_t = \beta(x_t - x_{t-1}) + (1 - \beta)z_{t-1}, \quad 0 < \alpha, \beta < 1. \quad (10)$$

- ii) Com componente sazonal multiplicativo:

$$x_t = \alpha \frac{y_t}{c_{t-q}} + (1 - \alpha)(x_{t-1} + z_{t-1}), \quad (11)$$

$$z_t = \beta(x_t - x_{t-1}) + (1 - \beta)z_{t-1},$$

$$c_t = \gamma \frac{y_t}{x_t} + (1 - \gamma)c_{t-q}. \quad (12)$$

em que $0 < \gamma < 1$ e q é a frequência de sazonalidade.

iii) Sazonalidade aditiva:

$$x_t = \alpha(y_t - c_{t-q}) + (1 - \alpha)(x_{t-1} + z_{t-1}) , \quad (13)$$

$$z_t = \beta(x_t - x_{t-1}) + (1 - \beta)z_{t-1},$$

$$c_t = \gamma(y_t - x_t) + (1 - \gamma)c_{t-q}. \quad (14)$$

2.4 Estatísticas de Erro de Previsão – Acurácia do Modelo

As estatísticas de desempenho dos modelos apresentados neste trabalho são a Root Square Mean Error (RSME), Mean Absolute Error (MAE) e Mean Absolute Percentage Error (MAPE) que estão definidas da seguinte forma:

Root Square Mean Error (RSME):

$$RSME = \sqrt{\frac{\sum_{h=1}^H e^2_{t(h)}}{H}} \quad (15)$$

Mean Absolute Error (MAE):

$$MAE_{t,H} = \frac{\sum_{h=1}^H |e_t(h)|}{H} \quad (16)$$

Mean Absolute Percentage Error (MAPE):

$$MAPE_{t,H} = \sum_{h=1}^H \left| \frac{e_t(h)}{Hy_{t+h}} \right| \quad (17)$$

Quanto menor forem os valores calculados para estas estatísticas, maior é a acurácia do modelo.

2.5 Base de Dados

A base de dados utilizada neste estudo é o desvio (%) do custo relativo de manutenção dos veículos (orçado *versus* realizado) de uma empresa do setor de distribuição de energia elétrica no Brasil, chamado aqui de Empresa A, cujo nome foi omitido por razões confidenciais. A ideia é projetar o custo de manutenção para os próximos meses, visando um melhor planejamento da equipe responsável pela frota da empresa.

O custo de manutenção relativo é mensurado em função do dispêndio em reais (R\$) gastos e orçados para o reparo dos veículos, dividido pela quilometragem rodada no período:

$$CRr_t = \frac{(\text{Valor gasto em reais R\$ com Manutenção})_t}{Km\ rodado_t}, \quad (18)$$

$$CRO_t = \frac{(\text{Valor orçado em reais R\$ com Manutenção})_t}{Km\ rodado_t}. \quad (19)$$

CRr_t representa o custo relativo de manutenção realizado no período t e CRO_t representa o custo relativo de manutenção orçado no período t . Portanto, o desvio percentual do custo orçado em relação ao realizado é dado por:

$$\text{Desvio } CR_t (\%) = \left(\frac{CRr_t}{CRO_t} - 1 \right) \times 100. \quad (20)$$

A série estudada, neste trabalho, contém desvios relativos do custo com manutenção dos veículos no período de janeiro de 2017 até dezembro de 2020.

3 Resultados

Os resultados obtidos e os gráficos demonstrados foram todos gerados através do software R (*R Foundation for Statistical Computing*, 2018) e estão apresentados nos tópicos a seguir.

3.1 Análise Descritiva

A tabela 1 mostra as estatísticas descritivas da série temporal do desvio do custo de manutenção relativo, conforme equação (20). Os dados apresentam leve assimetria, 0,05 aproximadamente, com uma distribuição relativamente platicúrtica, tendo em vista o coeficiente de curtose negativo. O histograma, apresentado na figura 1 traz a visualização de como os dados estão distribuídos.

Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo	Desvio-Padrão	Assimetria	Curtose
-0,130	0,023	0,090	0,089	0,178	0,260	0,106	-0,049	-0,883

Tabela 1: Estatísticas descritivas do desvio do custo relativo de manutenção.

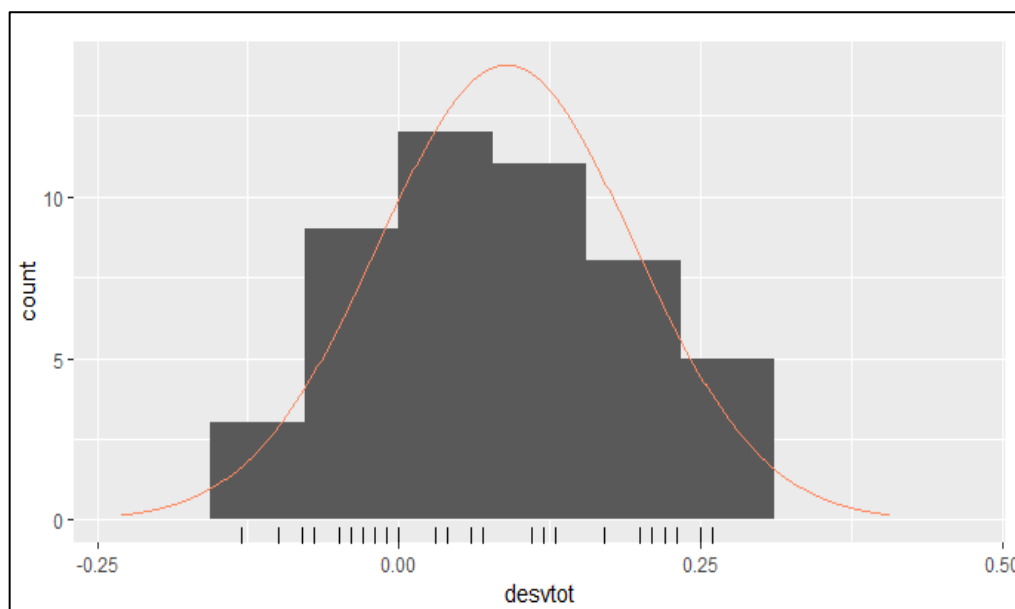


Figura 1: Histograma do desvio do custo relativo de manutenção.

Os dados seguem uma distribuição normal, conforme apontado pelo teste de Shapiro-Wilk na figura 1, em que as hipóteses são:

H_0 : A amostra provém de uma população Normal
 H_1 : A amostra não provém de uma população Normal

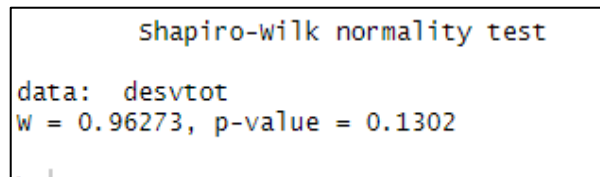


Figura 2: Teste de normalidade de Shapiro-Wilk do desvio do custo relativo de manutenção.

A figura 3 traz o gráfico de normalidade:

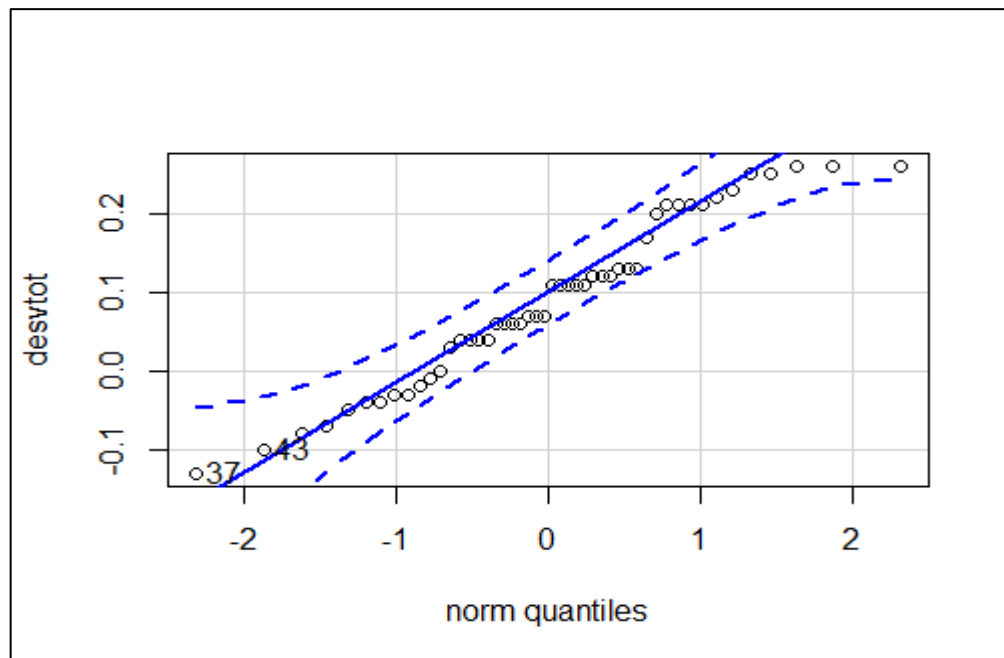


Figura 3: Gráfico de probabilidade normal do desvio do custo relativo de manutenção.

Foi aplicada uma diferenciação na série temporal, tendo em vista que a original não se mostrou estacionária através dos testes previamente realizados. A diferenciação é aplicada para que a série original se torne estacionária, conforme definido na seção de métodos. Após a transformação, os testes de Dickey-Fuller (Figura 4), Dickey-Fuller aumentado com tendência (Figura 5) e Phillips-Perron (Figura 6) mostraram que a série modificada apresenta estacionariedade ao longo do período estudado. As hipóteses dos testes de estacionariedade são:

H_0 : Há uma raiz unitária ou a série é não estacionária
 H_1 : Série temporal estacionária

```

Augmented Dickey-Fuller Test

data: diff(desvtot)
Dickey-Fuller = -3.5161, Lag order = 3, p-value = 0.04999
alternative hypothesis: stationary

```

Figura 4: Teste de Dickey-Fuller da série temporal diferenciada.

```

#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####

Test regression trend

Call:
lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)

Residuals:
    Min       1Q   Median       3Q      Max
-0.132454 -0.042995  0.005763  0.041122  0.090848

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.0076829  0.0253768   0.303  0.763820
z.lag.1     -0.8418220  0.2269739  -3.709  0.000698 ***
tt          -0.0002264  0.0008768  -0.258  0.797667
z.diff.lag  -0.2226779  0.1568274  -1.420  0.164241
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06398 on 36 degrees of freedom
Multiple R-squared:  0.5881,    Adjusted R-squared:  0.5538
F-statistic: 17.13 on 3 and 36 DF,  p-value: 4.447e-07

Value of test-statistic is: -3.7089 4.7255 6.9296

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -4.151 -3.50 -3.18
phi2  7.02  5.13  4.31
phi3  9.31  6.73  5.61

```

Figura 5: Teste de Dickey-Fuller aumentado com tendência da série temporal diferenciada.

```

Phillips-Perron Unit Root Test

data: diff(desvtot)
Dickey-Fuller Z(alpha) = -53.488, Truncation lag parameter = 3, p-value = 0.01
alternative hypothesis: stationary

```

Figura 6: Teste de Phillips-Perron da série temporal diferenciada.

A figura 7, a seguir, apresenta o acompanhamento da série ao longo do tempo, além dos gráficos da FAC e FACP, com o intuito de identificar a ordem do modelo $ARIMA(p, d, q)$ e o padrão sazonal.

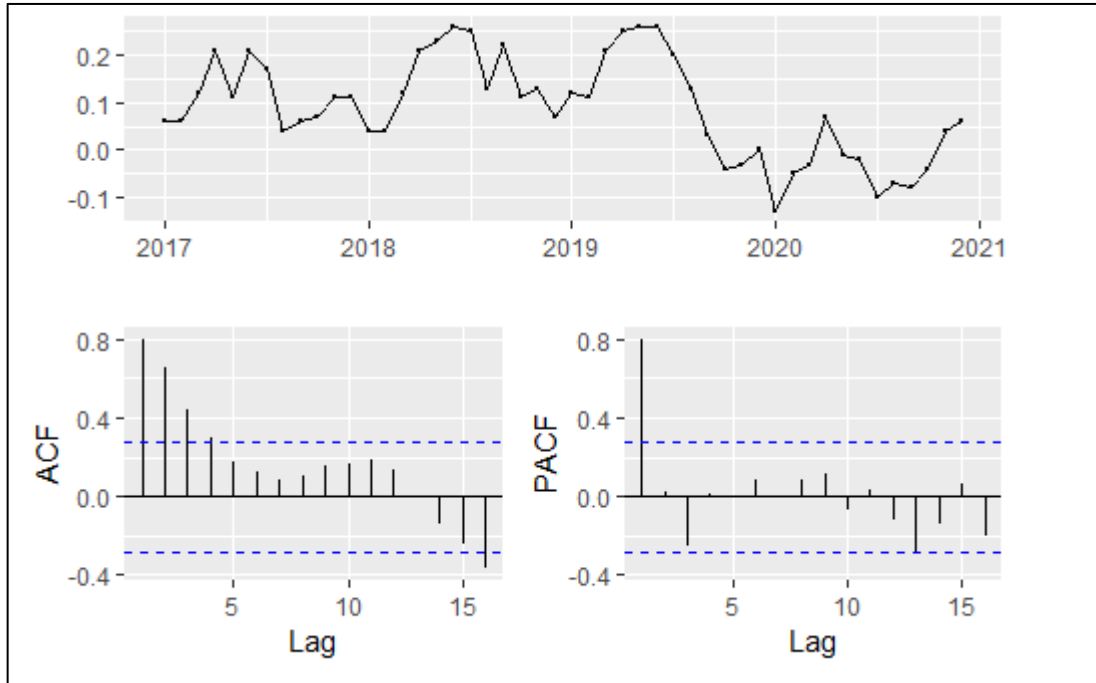


Figura 7: Gráficos da série temporal, FAC e FACP dos dados originais.

O decaimento no gráfico da função de autocorrelação (FAC) e a truncagem na primeira defasagem do gráfico da função de autocorrelação parcial (FACP) indicam que um modelo possível seria o $ARIMA(1,1,0)$. O gráfico da série temporal mostra uma certa sazonalidade ao longo dos anos, sendo importante, portanto, incluir um componente sazonal no modelo. Desta forma, vários modelos foram testados através do auto $ARIMA$, avaliando os critérios de informação como o Bayesian Information Criterion (BIC) ou Akaike Information Criterion (AIC).

Abaixo segue a figura 8 com os gráficos do acompanhamento da série ao longo do tempo, FAC e FACP dos dados transformados (uma diferenciação).

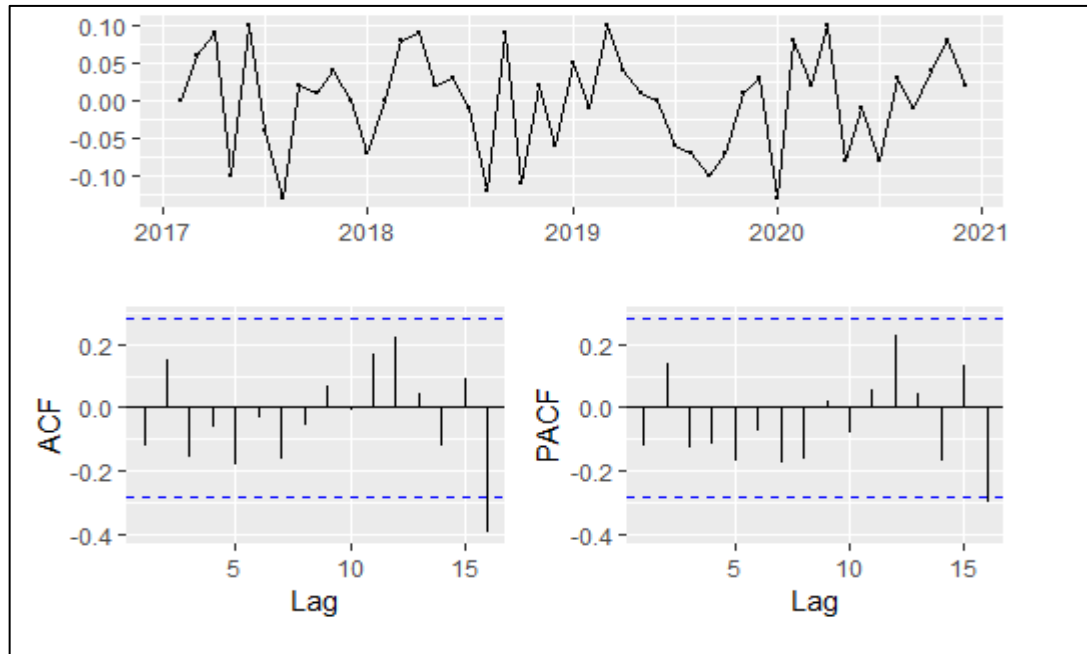


Figura 8: Gráficos da série temporal, FAC e FACP dos dados transformados.

As análises descritivas satisfizeram os pressupostos do modelo SARIMA, conforme metodologia apresentada na seção 2. Os modelos que serão testados, portanto, são o $SARIMA(p,d,q) \times (P,D,Q)_{12}$ e a suavização exponencial de Holt-Winters multiplicativo, tendo em vista a presença de sazonalidade estocástica na série.

3.2 Modelagem Box e Jenkins

O primeiro modelo a ser testado é o ARIMA com componente sazonal, conforme identificado na análise descritiva. Foram ajustados diversos modelos para construir um comparativo dos critérios de informação AIC e BIC, seguindo o conceito do auto ARIMA. Abaixo a tabela 2 mostra o comparativo dos critérios entre os modelos:

Verificação dos Modelos - Auto ARIMA		
Modelos	AIC	BIC
ARIMA(1,1,0)x(1,0,0) ₁₂	-96,74393	-91,83118
ARIMA(1,1,0)x(1,1,0) ₁₂	-64,76761	-60,99332
ARIMA(1,1,1)x(1,0,0) ₁₂	-95,95872	-89,40837
ARIMA(1,1,1)x(1,1,0) ₁₂	-63,63246	-58,60007
ARIMA(1,1,1)x(1,1,1) ₁₂	-62,10721	-55,81673

Tabela 2: Verificação dos modelos ARIMA através do comparativo das estatísticas AIC e BIC.

Tendo em vista que o modelo SARIMA(1,1,0)x(1,0,0)₁₂ apresentou os menores valores para as estatísticas AIC e BIC, além de os demais modelos não satisfazerem todos os pressupostos da modelagem, o primeiro foi o escolhido para ajustarmos a série de interesse. Os parâmetros encontrados estão na tabela 3. A base de treino foi definida em aproximadamente 80% dos dados, ou seja, do período de janeiro de 2017 a março de 2020. A base de validação, portanto, corresponde o período entre abril a dezembro de 2020.

Parâmetros	Estimativa	Desvio-Padrão	P-valor
θ_1	-0,308	0,169	0,069
ϕ_1	0,481	0,178	0,007

Tabela 3: Parâmetros estimados do modelo SARIMA(1,1,0)x(1,0,0)₁₂ dos dados referente ao desvio do custo relativo de manutenção.

Realizando o teste de Wald, o parâmetro autorregressivo se mostrou significativo à 10% e o sazonal à 1%, indicando que o modelo proposto se ajustou bem a série.

A próxima etapa, portanto, consiste na análise residual do modelo. A figura 9 e a tabela 4 trazem os resultados do teste de normalidade de Shapiro-Wilk e o teste de autocorrelação de Ljung-Box.

```
shapiro-wilk normality test
data: resid1
w = 0.95334, p-value = 0.1065
```

Figura 9: Teste de normalidade dos resíduos do modelo SARIMA(1,1,0)x(1,0,0)₁₂.

Teste de Ljung-Box - Resíduos		
Estatística	Lag	P-valor
5,839	12	0,923
12,699	24	0,970
15,775	36	0,999

Tabela 4: Teste de autocorrelação dos resíduos do modelo SARIMA(1,1,0)x(1,0,0)₁₂.

Os resultados da análise residual mostram que o modelo possui um ruído branco e gaussiano, sendo ele adequado aos dados da série temporal.

A tabela 5 traz as estatísticas de acurácia do modelo proposto (MAE, MAPE e RMSE) do período de validação, que corresponde a aproximadamente 20% da série temporal, ou seja, dos meses entre abril/2020 até dezembro/2020.

Acurácia do Modelo – Sarima		
MAE	MAPE	RMSE
0,076	160,056	0,099

Tabela 5: Acurácia do modelo SARIMA(1,1,0)x(1,0,0)₁₂.

A tabela 6 e a figura 10 trazem as estimativas pontuais e intervalares do modelo de previsão, conforme o período de validação proposto (9 meses).

Mês	Estimativa Pontual	Limite Inferior - 95%	Limite Superior - 95%
abr/20	-0,002	-0,123	0,119
mai/20	0,000	-0,147	0,147
jun/20	0,001	-0,174	0,176
jul/20	-0,028	-0,226	0,169
ago/20	-0,062	-0,280	0,156
set/20	-0,110	-0,347	0,127
out/20	-0,144	-0,398	0,111
nov/20	-0,139	-0,409	0,132
dez/20	-0,123	-0,409	0,163

Tabela 6: Estimativas pontuais e intervalares do modelo SARIMA(1,1,0)x(1,0,0)₁₂ no período de 9 meses.

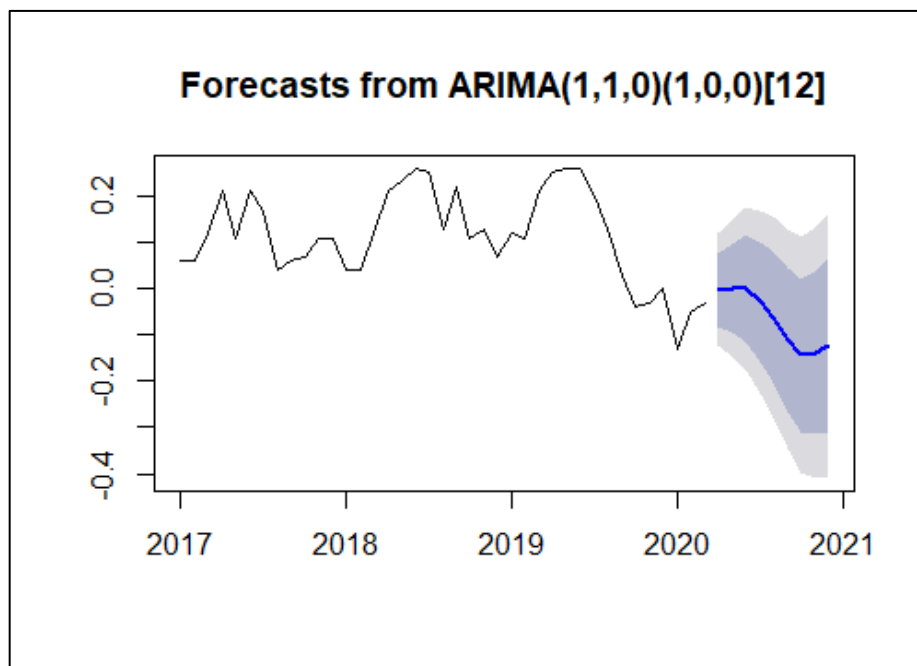


Figura 10: Série temporal e estimativas pontuais e intervalares (previsão) do modelo SARIMA(1,1,0)x(1,0,0)₁₂ no período de 9 meses.

3.3 Modelo de Holt-Winters

A tabela 7 traz os parâmetros de suavização α , β e γ ajustados na base de treino (janeiro de 2017 a março de 2020).

Parâmetros de Suavização	
Parâmetro	Estimativa
α	1,000
β	0,000
γ	0,333

Tabela 7: Parâmetros estimados do modelo de Holt-Winters dos dados referente ao desvio do custo relativo de manutenção.

A tabela 8 traz as estatísticas de acurácia do modelo de Holt-Winters (MAE, MAPE e RSME) do período de validação, que são os meses entre abril/2020 até dezembro/2020.

Acurácia do Modelo - Holt-Winters		
MAE	MAPE	RSME
0,055	115,292	0,063

Tabela 8: Acurácia do modelo de Holt-Winters.

A tabela 9 e a figura 11 trazem as estimativas pontuais e intervalares do modelo de previsão de Holt-Winters, conforme o período de validação proposto.

Mês	Estimativa Pontual	Limite Inferior - 95%	Limite Superior - 95%
abr/20	-0,042	-0,184	0,099
mai/20	-0,037	-0,245	0,171
jun/20	-0,033	-0,309	0,242
jul/20	-0,022	-0,309	0,265
ago/20	-0,003	-0,161	0,154
set/20	-0,002	-0,279	0,275
out/20	0,001	-0,351	0,354
nov/20	0,007	-0,541	0,555
dez/20	0,012	-0,523	0,546

Tabela 9: Estimativas pontuais e intervalares do modelo de Holt-Winters no período de 9 meses.

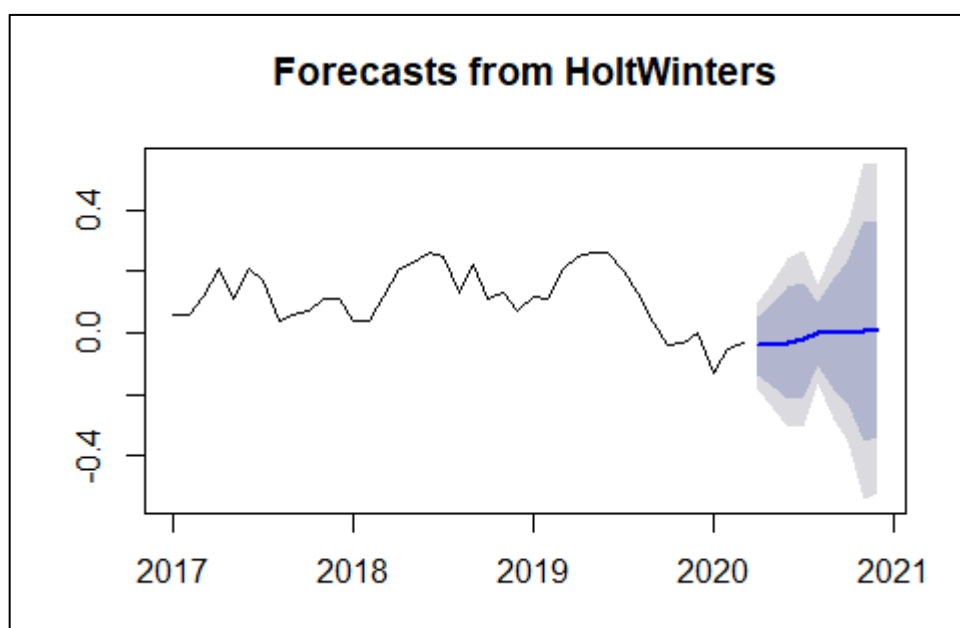


Figura 11: Série temporal e estimativas pontuais e intervalares (previsão) do modelo de Holt-Winters no período de 9 meses.

Podemos observar um estrangulamento no intervalo de previsão em função do parâmetro de suavização estimado para a sazonalidade.

3.4 Comparativo entre os modelos

A figura 12 traz o gráfico comparativo das previsões, conforme segue abaixo:

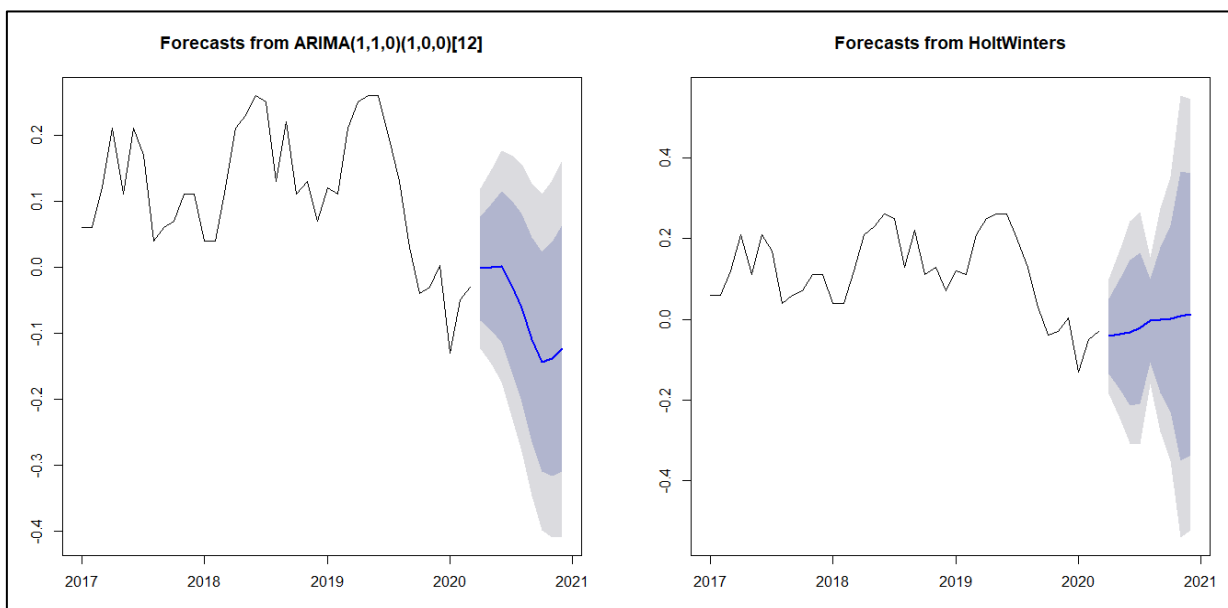


Figura 12: Comparativo dos modelos de previsão para o período de validação de 9 meses.

A análise gráfica nos mostra que o modelo $SARIMA(1,1,0) \times (1,0,0)_{12}$ capturou melhor a sazonalidade multiplicativa, tendo em vista o comportamento constante das previsões do Holt-Winters.

A tabela 10 traz o comparativo entre as estatísticas de acurácia Root Square Mean Error (RSME), *Mean Absolute Error* (MAE) e *Mean Absolute Percentage Error* (MAPE) dos modelos de Holt-Winters e $SARIMA(1,1,0) \times (1,0,0)_{12}$. O modelo que apresentar o menor valor para as estatísticas é o preferível para a série estudada.

Comparativo entre os modelos estimados		
Estatísticas	ARIMA	Holt-Winters
MAE	0,076	0,055
MAPE	160,056	115,292
RSME	0,099	0,063

Tabela 10: Comparativo entre as estatísticas de acurácia dos modelos $SARIMA(1,1,0) \times (1,0,0)_{12}$ e de Holt-Winters no período de 9 meses.

As estatísticas de previsão mostraram que o modelo de Holt-Winters se ajustou melhor a série proposta pelo trabalho, comparando o período de validação com o valor ajustado, tendo em vista que para este modelo os valores calculados foram menores quando comparado ao SARIMA. Isto se deu muito em função dos valores previstos seguirem um formato quase constante, passando mais próximo dos dados reais. A medida em que as novas variáveis vão sendo observadas, os modelos serão novamente testados e comparados.

4 Conclusão

O presente trabalho comparou os modelos autorregressivos e de médias móveis integrados sazonais, SARIMA, proposto por Box e Jenkins, e o de Holt-Winters, para prever o desvio do custo relativo de manutenção dos veículos da frota de uma empresa do setor de distribuição de energia elétrica brasileira.

Os resultados mostraram que ambos os modelos se ajustaram bem a série proposta, sendo que o de Holt-Winters é preferível em termos de previsão em função de estatísticas de erro menores.

Para trabalhos futuros seria interessante avaliar também como a série temporal de interesse se ajustaria a modelos de aprendizado de máquina para comparar com os dois propostos neste trabalho.

5 Referências Bibliográficas

ALON, I.; QI, M.; SADOWSKI, R.; (2001). Forecasting aggregate retail sales: a comparison of artificial neural networks and traditional methods. *Journal of Retailing and Consumer Services* 8 147-156.

BUENO, R.L.S. (2012). *Econometria de Séries Temporais*. São Paulo: Cengage Learning.

Chan, K. F., & Gray, P. (2006). Using extreme value theory to measure value-at-risk for daily electricity spot prices. *International Journal of forecasting*, 22(2), 283-300.

CRANAGE, D.; ANDREW, W.; (1992). A comparison of time series and econometric models for forecasting restaurant sales. *Int. J. Hospitality Management*, Vol. 11 No. 2, 129-142.

DOGANIS, P.; ALEXANDRIDIS, A.; PANAGIOTIS, P.; HARALAMBOS, S.; (2006). Time series sales forecasting for short shelf-life food products based on artificial neural networks and evolutionary computing. *Journal of Food Engineering* 75, 196–204.

HARVEY, A. C.; TODD, P. H. J. Forecasting economic time series with structural and Box-Jenkins models: A case study. *Journal of Business & Economic Statistics*, v.1, n.4, p.299-307, 1983.

LIPPI, M.; BERTINI, M.; FRASCONI, P.; (2013). Short-Term Traffic Flow Forecasting: An Experimental Comparison of Time-Series Analysis and Supervised Learning. *IEEE Transactions on Intelligent Transportation Systems*, 1-12.

MORETTIN, P.A. e TOLOI, C.M.C. (2004). *Análise de Séries Temporais*. São Paulo: Associação Brasileira de Estatística.

Morettin, P.A. (2006). *Econometria Financeira: Um curso em séries temporais financeiras*. São Paulo: Associação Brasileira de Estatística.

NOBRE, F. F.; MONTEIRO, A. B. S.; TELLES, P. R.; WILLIAMSON, G. D.. Dynamic linear model and SARIMA: a comparison of their forecasting performance in epidemiology. *Statistics in medicine*, v.20, n.20, p.3051-3069, 2001.

R CORE TEAM. R: *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, 2018.

SANTOS, T. R.. Uma comparação dos modelos SARIMA e estruturais para a previsão da incidência de Dengue em Belo Horizonte, Minas Gerais. *Revista Brasileira de Biometria*, v. 35, p. 98-114, 2017.

SFETSOS, A.; (1999). A comparison of various forecasting techniques applied to mean hourly wind speed time series. *Renewable Energy* 21, 23-35.

SHUMWAY, R.H. e STOFFER, D.S.; (2010). Time Series Analysis and Its Applications with R Examples. Springer.