

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Exatas – ICEX
Departamento de Estatística

Gustavo Monteiro Pereira

**Aplicação de modelos de séries temporais para previsão de
receita de uma empresa de consultoria**

Belo Horizonte

2022

UFMG/ICE_x/DEST

Gustavo Monteiro Pereira

Aplicação de modelos de séries temporais para previsão de receita de uma empresa de consultoria

Trabalho de Conclusão de Curso apresentado ao Departamento de Estatística do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais, como requisito parcial para a conclusão do curso de Especialização em Estatística - Ênfase em Estatística.

Orientador: Prof. Dr. Guilherme Lopes de Oliveira

Belo Horizonte

2022

2022, Gustavo Monteiro Pereira.
Todos os direitos reservados.

Pereira, Gustavo Monteiro

P436a Aplicação de modelos de séries temporais para previsão de receita de uma empresa de consultoria[recurso eletrônico] / Gustavo Monteiro Pereira —2022.
1 recurso online (48 f. il, color.): pdf.

Orientador: Guilherme Lopes de Oliveira.
Monografia (especialização) - Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Estatística.
Referências: 46-48

1. Estatística. 2. Análise de séries temporais. 3. Previsão - Receita. 4. Sociedades comerciais -Finanças -. I. Oliveira, Guilherme Lopes de. II. Universidade Federal de Minas Gerais I. Instituto de Ciências Exatas, Departamento de Estatística. III. Título.

CDU 519.2 (043)

Ficha catalográfica elaborada pela bibliotecária Belkiz Inez Rezende Costa
CRB 6/1510 Universidade Federal de Minas Gerais – ICEX



Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Estatística
Programa de Pós-Graduação / Especialização
Av. Pres. Antônio Carlos, 6627 - Pampulha
31270-901 – Belo Horizonte – MG

E-mail: pgest@ufmg.br
Tel: 3409-5923 – FAX: 3409-5924

ATA DO 263^a. TRABALHO DE FIM DE CURSO DE ESPECIALIZAÇÃO EM ESTATÍSTICA DE GUSTAVO MONTEIRO PEREIRA.

Aos quinze dias do mês de dezembro de 2022, às 16:30 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 **Gustavo Monteiro Pereira**, intitulado: “Aplicação de modelos de séries temporais para previsão de receita de uma empresa de consultoria”, como requisito para obtenção do Grau de Especialista em Estatística. Abrindo a sessão, o Presidente da Comissão, Professor Guilherme Lopes de Oliveira – 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 o 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, 15 de dezembro de 2022.

Guilherme Lopes de Oliveira

Prof. Guilherme Lopes de Oliveira (Orientador)
DECOM / CEFET-MG

Luis Alberto Toscano Medrano

Prof. Luis Alberto Toscano Medrano
DEMAT-UFRRJ

Documento assinado digitalmente

gov.br

GABRIELA OLIVEIRA

Data: 15/12/2022 20:35:41-0300

Verifique em <https://verificador.itl.br>

Prof. Gabriela Oliveira

Documento assinado digitalmente

gov.br

FABIO ROCHA DA SILVA

Data: 16/12/2022 09:26:55-0300

Verifique em <https://verificador.itl.br>

Prof. Fábio Rocha da Silva
DECOM / CEFET-MG



Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Estatística
Programa de Pós-Graduação / Especialização
Av. Pres. Antônio Carlos, 6627 - Pampulha
31270-901 – Belo Horizonte – MG

E-mail: pgest@ufmg.br
Tel: 3409-5923 – FAX: 3409-5924

DECLARAÇÃO DE CUMPRIMENTO DE REQUISITOS PARA CONCLUSÃO DO CURSO DE ESPECIALIZAÇÃO EM ESTATÍSTICA.

Declaro para os devidos fins que Gustavo Monteiro Pereira, número de registro 2020680380, cumpriu todos os requisitos necessários para conclusão do curso de Especialização em Estatística e que entregou para seu orientador, o professor Guilherme Lopes de Oliveira, que aprovou a versão final. O trabalho foi apresentado no dia 15 de dezembro de 2022 com o título “*Aplicação de modelos de séries temporais para previsão de receita de uma empresa de consultoria*”.

Belo Horizonte, 14 de fevereiro de 2023

A handwritten signature in blue ink, reading "Roberto da Costa Quinino".

Prof. Roberto da Costa Quinino
Coordenador do curso de
Especialização em Estatística
Departamento de Estatística / UFMG

*Ao meu pai,
que foi e sempre será o meu maior exemplo de superação e força de vontade.*

Agradecimentos

Meus principais agradecimentos são direcionados aos meus pais, que sempre foram uma referência de perseverança, esforço, humanidade e amor, por sempre apoiarem incondicionalmente meus estudos.

À minha companheira, que diariamente me incentiva em todos os momentos, me fazendo acreditar em mim e lembrar que juntos vamos sempre mais longe.

À minha família, que mesmo distante fisicamente sempre esteve presente como lembrança do que é realmente importante. Aos meus queridos amigos com quem compartilhei as dificuldades e risadas.

À empresa, que gentilmente cedeu os dados para a realização dessa pesquisa, e aos meus colegas de trabalho que sempre me apoiam em minhas ideias e estão sempre disponíveis para me ajudar. Sou muito feliz com as trocas que temos.

Um agradecimento especial ao meu querido orientador Guilherme Lopes de Oliveira, pelo grande apoio, organização, paciência e conselhos valiosos; para os grandes professores do Departamento de Estatística pela formação e ensinamentos que levarei por toda a vida; e aos técnicos da Universidade, que, mesmo em um tempo conturbado mantiveram toda a estrutura do curso funcionando perfeitamente.

Resumo

Este trabalho tem como objetivo utilizar a análise de séries temporais para realizar previsões para a receita de uma empresa de consultoria. Após análise exploratória dos dados, foram utilizados métodos de previsão baseados nos modelos de Alisamento Exponencial de Holt-Winters e modelos Autorregressivos Integrados de Médias Móveis Sazonais (SARIMA) que são estimados com base na própria série temporal. O processo de construção e identificação dos modelos foi bastante informativo sobre as características da série. Adicionalmente, as previsões feitas pelos modelos SARIMA tiveram boa capacidade preditiva.

Palavras-chave: análise de séries temporais; finanças corporativas; predição de receita; previsão.

Abstract

The aim of this study is to apply time series modelling to predict the revenue of a consulting company. The Holt-Winters exponential smoothing and Integrated Auto Regressive Seasonal Moving Averages (SARIMA) methods were applied to the data, where the SARIMA model had the best result predicting future values for the test dataset. The process of building and identifying the models was very informative about the characteristics of the time series, indicating useful information to conduct the business.

Keywords: corporate finance; forecasting; time series analysis; revenue prediction.

Lista de ilustrações

Figura 1 – Gráfico da FAC para um ruído branco gaussiano simulado pelo autor.	18
Figura 2 – Gráfico da FACP para um ruído branco gaussiano simulado pelo autor.	19
Figura 3 – Gráfico de resíduos, FAC e histograma para um ruído branco gaussiano simulado pelo autor.	25
Figura 4 – Receita da empresa no período de treino.	31
Figura 5 – Diagrama de caixa da receita no período de treino.	31
Figura 6 – Gráfico de sazonalidade da receita no período de treino.	32
Figura 7 – Decomposição do log da receita no período de treino.	33
Figura 8 – Diagrama de caixa do log da receita no período de treino	33
Figura 9 – Log da receita no período de treino.	34
Figura 10 – Log da receita com uma diferença no período de treino.	34
Figura 11 – Resíduos do modelo SARIMA(0,1,1)(1,0,0) no período de treino.	37
Figura 12 – Q-Q plot do modelo SARIMA(0,1,1)(1,0,0)	37
Figura 13 – Resíduos do modelo SARIMA(0,1,1)(0,0,1) no período de treino	38
Figura 14 – Q-Q plot do modelo SARIMA(0,1,1)(0,0,1)	39
Figura 15 – Resíduos do modelo SARIMA(1,1,1)(0,0,1) no período de treino	40
Figura 16 – Q-Q plot do modelo SARIMA(1,1,1)(0,0,1)	41
Figura 17 – Resíduos do modelo SARIMA(0,1,1)(0,1,1) no período de treino	42
Figura 18 – Q-Q plot do modelo SARIMA(0,1,1)(0,0,1)	43
Figura 19 – Resíduos do modelo SARIMA(2,1,0)(1,1,0) no período de treino	43
Figura 20 – Q-Q plot do modelo SARIMA(2,1,0)(1,1,0)	44
Figura 21 – Comparativo entre previsão e real por modelo	46
Figura 22 – Comparativo entre estimativa e real período de testes - SARIMA(0,1,1)(1,0,0)	47

Lista de tabelas

Tabela 1 – Parâmetros para transformação Box-Cox	17
Tabela 2 – Referência para análise de gráficos da FAC e FACP	22
Tabela 3 – Parâmetros estimados para os modelos SARIMA (dados de treino).	36
Tabela 4 – Métricas de avaliação dos modelos ajustados (dados de treino).	36
Tabela 5 – Testes para normalidade dos resíduos - Modelo SARIMA(0,1,1)(1,0,0) . . .	38
Tabela 6 – Testes para normalidade dos resíduos - Modelo SARIMA(0,1,1)(0,0,1) . . .	38
Tabela 7 – Testes para normalidade dos resíduos - Modelo SARIMA(1,1,1)(0,0,1) . . .	40
Tabela 8 – Testes para normalidade dos resíduos - Modelo SARIMA(0,1,1)(0,1,1) . . .	42
Tabela 9 – Testes para normalidade dos resíduos - Modelo SARIMA(2,1,0)(1,1,0) . . .	44
Tabela 10 – Acurácia dos modelos nas duas bases de dados	45

Lista de abreviaturas e siglas

AIC	Akaike Information Criteria - Critério de Informação de Akaike
ARIMA	Autoregressive Integrated Moving Average - modelo autorregressivo integrado de médias móveis
BIC	Bayesian Information Criteria - Critério de Informação Bayesiano
CPI	Consumer Price Index - Índice de Preços ao Consumidor
FAC	Função de Autocorrelação
FACP	Função de Autocorrelação Parcial
KPSS	Kwiatkowski–Phillips–Schmidt–Shin
MAE	Mean absolute error - erro absoluto médio
MAPE	Mean absolute percentage error - erro percentual absoluto médio
MMS	Médias Móveis Simples
MPE	Mean percentage error - erro percentual médio
PIB	Produto Interno Bruto
RMSE	Root mean squared error - Raiz do erro quadrático médio
SARIMA	Seasonal Autoregressive Integrated Moving Average - modelo autorregressivo integrado de médias móveis sazonal
SEH	Suavização Exponencial de Holt
SES	Suavização Exponencial Simples
VAR	Vetor Autoregressivo

Sumário

1	INTRODUÇÃO	14
2	METODOLOGIA	16
2.1	Modelos de Box-Jenkins	20
2.1.1	Processo autorregressivo (AR)	20
2.1.2	Processo de média móvel (MA)	21
2.1.3	Processo autorregressivo de médias móveis (ARMA)	21
2.1.4	Modelo ARIMA	22
2.1.5	Modelo SARIMA	24
2.2	Modelos de Alisamento Exponencial	26
2.3	Crterios para seleção de modelos	27
2.4	Software	29
3	APLICAÇÃO	30
3.1	Base de Dados	30
3.2	Análise Exploratória dos Dados	30
3.3	Modelagem	35
3.3.1	Definição dos Modelos	35
3.3.2	Análise de Resíduos	36
3.3.2.1	SARIMA(0,1,1)(1,0,0)	36
3.3.2.2	SARIMA(0,1,1)(0,0,1)	37
3.3.2.3	SARIMA(1,1,1)(0,0,1)	38
3.3.2.4	SARIMA(0,1,1)(0,1,1)	41
3.3.2.5	SARIMA(2,1,0)(1,1,0)	42
3.3.3	Acurácia dos Modelos	45
4	CONCLUSÃO	48
	REFERÊNCIAS	49

1 Introdução

A previsibilidade é um ponto essencial nos negócios. Gestores diariamente precisam tomar decisões sobre o futuro com base em dados do presente e passado. Considerando a receita de uma empresa, ser capaz de fazer previsões é fundamental, visto que a receita determina o quanto será gerado de caixa para pagar colaboradores, fornecedores e investidores. É também um aspecto essencial para determinar atividades futuras, como expansão de novas unidades, contratação de mais funcionários e investimento em novas ferramentas tecnológicas com o objetivo de aumentar a produtividade.

Neste caso, pode-se definir a receita de uma empresa como o valor que ela fatura vendendo os seus produtos ou serviços aos clientes (BOARD, 2014). No caso de uma empresa de consultoria, os serviços vendidos são as horas dos colaboradores que se dedicam à empresa cliente. O preço dessa hora varia de acordo com a especialidade e nível de experiência da pessoa consultora, sendo maior em nichos mais especializados e escassos - como tecnologias recentes na fronteira do conhecimento - e também aumentando de acordo com a senioridade da pessoa que presta o serviço.

A empresa tem então basicamente duas formas de aumentar a receita: (1) elevando o preço médio vendido por meio de consultores mais seniores ou especializados; e (2) vendendo mais horas, isto é, realizando mais projetos para as empresas clientes.

O processo de avaliação da receita costuma ser constante: as empresas buscam sempre reduzir a incerteza futura por meio de estimativas de receita dos meses à frente. Em geral, isso ocorre de maneira frequente com previsões para todos os meses do ano, as vezes com uma janela fixa de realização das previsões. Por exemplo, pode-se definir que a receita dos próximos 12 meses será estimada mensalmente. Isso varia de acordo com a organização, mas um ponto chave é que as empresas sempre buscam saber o quanto gerarão de receita para constantemente ajustarem as decisões de investimento, como citado anteriormente.

Nesse contexto, a análise de séries temporais possui um grande potencial de reduzir essa incerteza ao propor métodos para essas estimativas futuras. Isso é feito em diversas esferas organizacionais, incluindo, por exemplo, as empresas privadas, como feito por Kodde e Schreuder (1984). Tais autores compararam modelos de séries temporais aplicados à previsão da receita e lucro de empresas com as estimativas feitas pelo time de gestão das mesmas. Nestes casos, os resultados dos modelos estimados superaram as previsões feitas internamente.

Há também a aplicação em entidades governamentais, como no trabalho de Streimikiene et al. (2018). Os autores estimaram a receita de impostos para o governo do Paquistão aplicando modelos estatísticos em dados da arrecadação mensal e também o impacto dos impostos indiretos na classe trabalhadora. Foram estimados três modelos: autorregressivo com **dummies**

sazonais, autorregressivo integrado de médias móveis (autoregressive integrated moving average - ARIMA), e Vetor Autorregressivo (VAR). O modelo com melhor performance foi o ARIMA.

No nível econômico, modelos de séries temporais também são aplicados com pesquisas de expectativas de agentes da economia. No trabalho de [Clar, Duque e Moreno \(2007\)](#), os autores fazem a previsão de indicadores de confiança de negócios e consumidores para a região do Euro usando diversos modelos de séries temporais. Com esses indicadores estimados, os agentes da economia podem antecipar decisões com base na expectativa para o futuro.

Em suma, os modelos de séries temporais possuem uma infinidade de aplicações com temas variados, como a previsão de preços de imóveis ([CHEN; KAWAGUCHI; PATEL, 2004](#)), a estimativa da quantidade de cirurgias em hospitais ([ZINOURI; TAAFFE; NEYENS, 2018](#)), o nível de desigualdade de renda futuro ([ZHENG; XU; TANG, 2000](#)), a expectativa de consumo de energia ([WANG; LI; LI, 2018](#)) e diversos outros.

No presente trabalho, o objetivo é identificar o melhor método para a previsão de receita de uma empresa de consultoria dentro da classe de modelos autorregressivos integrados de médias móveis (do inglês **autoregressive integrated moving average** - ARIMA) e também modelos de alisamento exponencial. Os dados são confidenciais e foram gentilmente cedidos por uma empresa para a execução da pesquisa.

O trabalho está dividido em 4 capítulos, considerando essa introdução, que abordou a contextualização e motivação para a realização do trabalho. O Capítulo 2 aborda a fundamentação teórica da metodologia aplicada na pesquisa e na análise de dados. No Capítulo 3 é feita a análise empírica utilizando os modelos da metodologia nos dados reais fornecidos pela empresa. Por último, no Capítulo 4, é feita uma conclusão baseada no que foi discutido na análise de dados, sob a ótica do objetivo mencionado no parágrafo anterior.

2 Metodologia

A análise de séries temporais é baseada em alguns conceitos fundamentais. O ponto inicial é a definição do que é uma série temporal de fato: uma sequência de observações encadeadas que ocorrem ao longo do tempo (BOX et al., 2015). Essa sequência é chamada de processo estocástico, em que a ocorrência de cada observação é aleatória. De maneira geral, usando a definição de Morettin (2018) uma série temporal pode ser representada como

$$Y_t = T_t + S_t + a_t, \quad (2.1)$$

onde Y_t é a observação da variável Y no período t ; T_t e S_t representam tendência e sazonalidade; e a_t é uma componente aleatória de média zero e variância constante σ^2 .

Com o objetivo de realizar a previsão de séries temporais como a apresentada na Equação (2.1) tomando como base os próprios valores observados para a série no passado, é importante observar diversas características. A primeira delas é a estacionariedade da série. Uma série pode ser definida como estacionária quando sua média e variância são constantes ao longo do tempo (GUJARATI, 2009). Quando a série é estacionária, os seus valores flutuam em torno da média constante e tendem a retornar para ela com o passar do tempo. Esse fenômeno é utilizado por alguns modelos para fornecer estimativas de previsões.

Para checar se uma série é estacionária, podem ser realizados diversos testes de raiz unitária. Dickey e Fuller (1979), em um dos primeiros trabalhos do tema, desenvolveram um teste em que a hipótese nula da existência de uma raiz unitária na série é testada contra a hipótese alternativa de estacionariedade. Kwiatkowski et al. (1992) desenvolveram um teste que baseia-se no contrário: a hipótese nula de estacionariedade dos dados em torno de uma tendência é testada contra a hipótese alternativa de presença de raiz unitária. Sendo este segundo teste - KPSS - mais propício a séries que possuem tendência definida, uma característica que será vista mais adiante.

Quando a série não apresenta comportamento estacionário, algumas transformações podem ser aplicadas a fim de torná-la estacionária. De acordo com Morettin (2018), a transformação mais comum é a diferenciação: as observações sucessivas da série são diferenciadas entre si, isto é, é feita a diferença entre uma observação com a observação imediatamente anterior, como na Equação (2.2), onde é feita a primeira diferença da série Y

$$\Delta Y_t = Y_t - Y_{t-1}. \quad (2.2)$$

A segunda diferença é definida da seguinte forma

$$\Delta^2 Y_t = \Delta[\Delta Y_t] = \Delta[Y_t - Y_{t-1}] = Y_t - 2Y_{t-1} + Y_{t-2}.$$

Segundo Lütkepohl e Xu (2010), uma segunda transformação muito utilizada em séries econômicas e financeiras é a transformação usando a escala logarítmica. Em geral, o objetivo é

reduzir a variabilidade da série entre períodos, tornando-a mais próxima de uma série homocedástica, isto é, fazer com que a variância não mude de acordo com o período no tempo. Essa transformação é feita simplesmente aplicando-se a função \log em todas as observações da série. Há também as transformações propostas por [Box e Cox \(1964\)](#) para estabilizar a variância de uma série, que inclui a transformação logarítmica citada anteriormente. Essas transformações são definidas como:

$$Y_t^x = \begin{cases} \frac{Y_t^x - c}{x} & x \neq 0 \\ \log Y_t & x = 0, \end{cases}$$

onde x e c são parâmetros a serem estimados. A Tabela 1 a seguir apresenta alguns possíveis valores para x e a transformação indicada em cada caso, com $c = 0$:

Tabela 1 – Parâmetros para transformação Box-Cox

x	Transformação
-1,0	$T(Y_t) = 1/Y_t$
-0,5	$T(Y_t) = 1/\sqrt{Y_t}$
0,0	$T(Y_t) = \log Y_t$
0,5	$T(Y_t) = \sqrt{Y_t}$
1,0	$T(Y_t) = Y_t$

Fonte: elaborado pelo autor.

É importante também observar alguns padrões que podem ocorrer nas séries temporais. O primeiro deles é a tendência, isto é, a série tem um comportamento definido de crescimento ou redução ao longo do tempo ([HYNDMAN; ATHANASOPOULOS, 2021](#)). O segundo padrão que também pode ocorrer em uma série temporal é a sazonalidade, em que a série é afetada por acontecimentos específicos a um período de tempo ([HYNDMAN; ATHANASOPOULOS, 2021](#)). Isso pode ocorrer em qualquer frequência fixa, de acordo com a periodicidade da série (diária, semanal, mensal, anual, etc.). O último padrão mencionado por [Hyndman e Athanasopoulos \(2021\)](#) é o cíclico. Nele a série apresenta quedas e crescimentos constantes, mas, diferente da sazonalidade, eles não têm frequência definida e geralmente estão associados a condições econômicas.

Considerando os conceitos mencionados anteriormente, durante a análise de séries temporais, busca-se modelar as estruturas de dependência presentes entre as observações. Em geral, isso é feito com base na análise da sua Função de Autocorrelação (FAC) e Função de Autocorrelação Parcial (FACP). Segundo [Gujarati \(2009\)](#), na prática, como em geral tem-se apenas a amostra de um processo estocástico, é computada a autocorrelação amostral de ordem (defasagem ou lag) k , denotada por $\hat{\rho}_k$, que representa a correlação entre duas variáveis aleatórias

Y_t e Y_{t-k} separadas por k períodos de tempo e é definida como

$$\hat{\rho}_k = \frac{\hat{Cov}(Y_t, Y_{t-k})}{\sqrt{\hat{Var}(Y_t)\hat{Var}(Y_{t-k})}} = \frac{\hat{\gamma}_k}{\hat{\gamma}_0},$$

onde $\hat{\gamma}_k$ é a covariância da amostra com defasagem k definida como

$$\hat{\gamma}_k = \frac{\sum(Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{n},$$

e $\hat{\gamma}_0$ é a variância da amostra estimada por

$$\hat{\gamma}_0 = \frac{\sum(Y_t - \bar{Y})^2}{n},$$

em que n é o tamanho da amostra e \bar{Y} é a média da amostra. Além disso, [Morettin \(2018\)](#) indica algumas propriedades importantes da função de autocorrelação teórica:

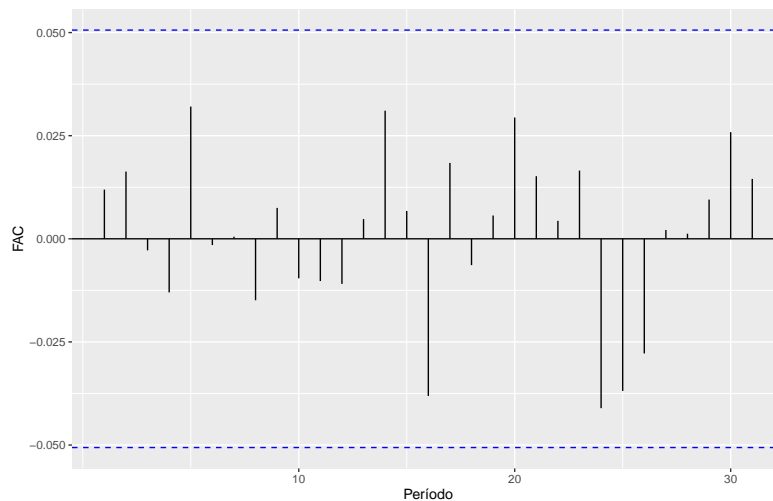
$$\rho_0 = 1$$

$$|\rho_k| \leq 1$$

$$\rho_k = \rho_{-k}.$$

O gráfico de $\hat{\rho}_k$ contra k é conhecido como correlograma amostral e é ilustrado na Figura 1, onde as barras representam a autocorrelação de acordo com as defasagens (lags) e as linhas tracejadas delimitam o intervalo de confiança. O exemplo exposto na Figura 1 refere-se a um processo estocástico chamado de ruído branco gaussiano, que é definido como um processo puramente aleatório onde as observações não são não-correlacionadas e identicamente distribuídas de acordo com uma distribuição Normal com média zero e variância constante ([GUJARATI, 2009](#)).

Figura 1 – Gráfico da FAC para um ruído branco gaussiano simulado pelo autor.



Fonte: elaborado pelo autor.

Para avaliar a significância estatística de $\hat{\rho}_k$, é utilizado o erro-padrão. Conforme demonstrado por [Bartlett \(1946\)](#), uma série temporal puramente aleatória (ruído branco) tem coeficientes de correlação amostral $\hat{\rho}_k$ normalmente distribuídos, $\hat{\rho}_k \approx N(0, S_{\hat{\rho}_k})$, onde

$$S_{\hat{\rho}_k} \approx \sqrt{\frac{1}{n}(1 + 2\hat{\rho}_1^2, \dots, \hat{\rho}_{k-1}^2)}.$$

Logo, é possível adicionar o intervalo de confiança em todos os correlogramas para identificar quais lags possuem autocorrelação amostral estatisticamente significativa.

Diferente da FAC, que mede a correlação amostral entre todas as observações, a FACP busca medir a correlação amostral parcial ρ_{kk} entre as observações da série que estão separadas por k períodos, removendo o efeito das correlações intermediárias ([GUJARATI, 2009](#)), isto é, mede-se a correlação entre Y_t e Y_{t-k} depois de removidas as dependências lineares das variáveis $Y_{t+1} + Y_{t+2}, \dots, Y_{t+k-1}$. Para o lag no período k , a FACP pode ser estimada como:

$$Y_t = \rho_{kk}Y_{t-1} + \dots + \rho_{kk}Y_{t-k} + u_t;$$

$$\hat{\rho}_{kk} = \text{Corr}\left(Y_t, \frac{Y_{t+k}}{Y_{t+1}, \dots, Y_{t+k-1}}\right).$$

Essa função é ilustrada na [Figura 2](#).

Figura 2 – Gráfico da FACP para um ruído branco gaussiano simulado pelo autor.



Fonte: elaborado pelo autor.

2.1 Modelos de Box-Jenkins

Box et al. (2015) propõem uma metodologia para identificar qual tipo de modelo deve ser usado em cada série temporal. Esse processo passa por quatro etapas:

1. Identificação
2. Estimação
3. Verificação do diagnóstico
4. Previsão

A identificação de um possível modelo se baseia nas características das suas funções de FAC e FACP teóricas. O modelo de Box-Jenkins assumem que a série temporal é estacionária e as características principais dos diferentes tipos de modelos serão descritas na sequência.

2.1.1 Processo autorregressivo (AR)

Quando uma série estacionária Y_t apresenta elementos que são dependentes entre si ao longo do tempo, pode-se modelar a sua estrutura por meio de modelos autorregressivos. Esse tipo de modelo varia de acordo com o grau de dependência, aumentando em conjunto com a autocorrelação com elementos anteriores. Podemos representar um modelo autorregressivo de ordem 1, AR(1), da seguinte forma (GUJARATI, 2009):

$$Y_t = \delta + \alpha_1 Y_{t-1} + u_t,$$

em que δ é a média de Y , α_1 é o fator que pondera Y_{t-1} e u_t é um ruído aleatório que se assume serem independentes e identicamente distribuídos com média zero e variância σ_u^2 constante para todo t .

Isso quer dizer que o valor de Y no período t depende do seu valor no período anterior ponderado pelo fator α_1 e de um termo aleatório u_t . Um modelo geral AR(p), apresenta forma igual à Equação (2.3):

$$Y_t = \delta + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + u_t. \quad (2.3)$$

Uma série estacionária representada por um processo AR(p) tem um decaimento exponencial na FAC teórica e as p primeiras autocorrelações significativas na FACP teórica. Tal comportamento pode ser utilizado no passo de identificação de um possível componente autorregressivo para modelar uma série temporal em questão a partir da análise da FAC e FACP estimadas com os dados amostrais.

2.1.2 Processo de média móvel (MA)

No caso do processo de média móvel, considera-se defasagens dos termos de erro (ruídos) para realizar a previsão de Y_t , suposta como uma série estacionária. Isto é, Y_t é dado por uma média móvel dos termos de erro atuais e passados. No caso de uma média móvel de primeira ordem, MA(1), o modelo será definido da seguinte forma (GUJARATI, 2009):

$$Y_t = \mu + u_t + \beta_1 u_{t-1},$$

onde μ é uma constante, β_1 é um fator que pondera u_{t-1} e u_t é um ruído aleatório que se assume serem independentes e identicamente distribuídos com média zero e variância σ_u^2 constante para todo t . O caso geral para um modelo MA(q) é

$$Y_t = \mu + \beta_1 u_{t-1} + \beta_2 u_{t-2} + \dots + \beta_q u_{t-q} + u_t.$$

Teoricamente, uma série estacionária representada por um processo MA(q) tem as q primeiras autocorrelações significativas na FAC e um decaimento exponencial na FACP. Tal comportamento pode ser utilizado no passo de identificação de um possível componente média móvel para modelar a série temporal em questão a partir da análise da FAC e FACP estimadas com os dados amostrais.

2.1.3 Processo autorregressivo de médias móveis (ARMA)

Quando uma série estacionária Y_t possui características de modelos autorregressivos e de modelos de média móvel, ele é modelado como ARMA(p, q), onde p é o grau de ordens autorregressivos e q de média móvel. Para o caso de Y_t seguir um processo ARMA(1,1), o modelo é formulada como na Equação (2.4):

$$Y_t = \theta + \alpha_1 Y_{t-1} + u_t + \beta_1 u_{t-1}, \quad (2.4)$$

onde θ neste caso representa a constante e u_t é um ruído aleatório que se assume serem independentes e identicamente distribuídos com média zero e variância σ_u^2 constante para todo t . O modelo geral para um processo ARMA(p,q) seria

$$Y_t = \theta + \alpha_1 Y_{t-1} + \dots + \alpha_p Y_{t-p} + u_t + \beta_1 u_{t-1} + \dots + \beta_q u_{t-q},$$

onde θ , u_t , β_i e α_i , $i = 1, \dots, p$ ou q são como definidos para o modelo ARMA(1,1).

A FAC e FACP teóricas de um processo ARMA mescla as características usuais destas funções no caso de processos AR e MA.

2.1.4 Modelo ARIMA

Todos modelos citados anteriormente assumem que a série temporal Y_t sendo trabalhada é estacionária, isto é, possuem média e variância constantes e covariância invariante no tempo. Entretanto, especialmente em séries econômicas e financeiras, isso não costuma acontecer sempre. Neste caso elas são chamadas de "integradas", em vez de estacionárias (GUJARATI, 2009). Uma forma de lidar com séries que possuem essa característica é aplicando uma diferença sobre os dados conforme definido na Equação (2.2). Esse processo muitas vezes é capaz de transformar uma série integrada em estacionária. Quando não for o suficiente, é possível aplicar mais de uma diferenciação.

Esse processo de aplicar uma diferença na série é nomeado como diferenciação, sendo escrito como $I(d)$, onde d representa o número de diferenças necessárias para transformar a série em estacionária. Desta forma, as séries que utilizam essa categoria de modelos são chamadas de autoregressivas integradas de médias móveis, $ARIMA(p, d, q)$, composta então por 3 elementos $AR(p)$, $I(d)$ e $MA(q)$.

Assim, voltando às etapas da metodologia de Box-Jenkins citadas anteriormente, busca-se na primeira etapa identificar preliminarmente os valores para p , d e q por meio dos correlogramas de FAC e FACP. Este processo inicia-se pela possível diferenciação da série (d), visto que é por ele que a série se transforma em estacionária, caso ainda não seja. Isso costuma ser visualizado pelo gráfico da própria série ao longo do tempo. Em seguida, seleciona-se os graus de p e q de um modelo ARMA para a série já diferenciada, com base nos gráficos da FAC e da FACP, seguindo, em linhas gerais, a referência da Tabela 2.

Tabela 2 – Referência para análise de gráficos da FAC e FACP

Modelo	FAC	FACP
AR(1)	Decrescimento exponencial	1 pico no lag 1
MA(1)	1 pico no lag 1	Decrescimento exponencial
AR(2)	Exponenciais ou ondas senóides	1 pico no lag 1 e 1 pico no lag 2
MA(2)	1 pico no lag 1 e 1 pico no lag 2	Exponenciais ou ondas senóides
ARMA(1,1)	Decrescimento exponencial	Decrescimento exponencial

Fonte: elaborado pelo autor.

Depois de identificados os valores de p , d e q , estima-se os os parâmetros dos termos autorregressivos e dos termos de média móvel do modelo. Para isso, pode-se considerar um modelo $ARIMA(p,d,q)$ onde seus $p + q + 1$ parâmetros estejam no vetor $\xi = (\alpha, \beta, \sigma_u^2)$, sendo $\alpha = (\alpha_1, \dots, \alpha_p)$, $\beta = (\beta_1, \dots, \beta_q)$. Quando $d > 0$, supõe-se $\mu = 0$. Caso contrário, considera-se μ como mais um parâmetro a ser estimado, tendo então $p + q + 2$ parâmetros para estimação (MORETTIN, 2018).

Para estimar ξ , em geral, utiliza-se o método de máxima verossimilhança: dadas as N observações Y_1, \dots, Y_N , considera-se a função de verossimilhança $L(\xi|Y_1, \dots, Y_N)$. Os

estimadores de máxima verossimilhança (EMV) de ξ serão os valores que maximizam L ou $\log L$. Para se determinar os EMV, supõe-se que o processo $u_t \sim N(0, \sigma_u^2)$, ou seja, assume-se que o u_t é um ruído branco gaussiano. Nessas condições, os EMV serão aproximadamente iguais aos estimadores de mínimos quadrados (EMQ) (MORETTIN, 2018). Tomando-se d diferenças para alcançar a estacionariedade da série, tem-se $n = N - d$ observações W_1, \dots, W_t , onde $W_t = \Delta^d Y_t$. Como o modelo ARMA(p,q) resultante é estacionário e invertível, pode-se definir:

$$u_t = \tilde{W}_t - \alpha_1 \tilde{W}_{t-1} - \dots - \alpha_p \tilde{W}_{t-p} + \beta_1 u_{t-1} + \dots + \beta_q u_{t-q}, \quad (2.5)$$

onde $\tilde{W}_t = W_t - \mu_w$. Para encontrar os valores de u_t através da Equação (2.5), é necessário identificar os valores iniciais dos \tilde{W} 's e dos u 's. Morettin (2018) propõe dois procedimentos para isso: o condicional, em que os valores iniciais desconhecidos são substituídos por valores julgados razoáveis, e o incondicional, em que os valores iniciais são estimados por meio do procedimento de "backforecasting".

No procedimento condicional, assume-se a normalidade dos u_t e a função de densidade conjunta de u_1, \dots, u_n (Equação (2.6)), onde é maximizada a verossimilhança e são substituídos os valores iniciais de W_t e u_t pelos julgados razoáveis.

$$f(u_1, \dots, u_n) = (2\pi)^{-n/2} \sigma_u^{-n} \exp\left\{-\sum_{t=1}^n u_t^2 / 2\sigma_u^2\right\}. \quad (2.6)$$

Já o procedimento incondicional considera que o logaritmo da função de verossimilhança não condicional é dado por

$$l(\xi) \simeq -n \log \sigma_u - \frac{S(\eta)}{2\sigma_u^2},$$

onde

$$S(\eta) = S(\alpha, \beta) = \sum_{t=-\infty}^n [u_t(\eta, W)]^2 \quad (2.7)$$

é a soma de quadrados não-condicional, com

$$[u_t(\eta, W)] = E(u_t | \eta, W). \quad (2.8)$$

Portanto, para obter os estimadores dos mínimos quadrados, minimiza-se a Equação (2.7). Estes estimadores serão boas aproximações para os estimadores de máxima verossimilhança. Com isso, para calcular a Equação (2.7) para um dado η , obtém-se as esperanças condicionais definidas na Equação (2.8) por meio da Equação (2.5). Para dar início a todo este processo, faz-se uma espécie de "previsão" para o passado (**backforecasting**), com o objetivo de calcular valores para W_{-j} e u_{-j} , $j = 0, 1, 2, \dots$, ou seja, valores anteriores ao início da série.

Após ser feita a estimativa, entra-se na terceira etapa de verificação do modelo. Isso é realizado por meio de checagem das estimativas em comparação com os dados reais e por meio da verificação da validade das suposições feitas a respeito do resíduo, ponto que será aprofundado na Seção 2.1.5. Por último, executa-se a previsão para os períodos futuros com base

na estimativa do modelo. É importante reforçar que todas essas etapas compõem um processo bastante iterativo, portanto, ajusta-se o modelo identificado de acordo com o que foi encontrado na etapas de estimativa, de verificação e de previsão.

2.1.5 Modelo SARIMA

Os modelos da classe SARIMA são uma extensão dos modelos ARIMA. Eles incorporam um componente sazonal que não é tratado pelos modelos ARIMA tradicionais. Isso é feito adicionando alguns parâmetros no modelo. A nomenclatura é SARIMA(p, d, q)(P, D, Q), onde P, D, Q são os parâmetros sazonais de um processo autorregressivo (P), diferenças sazonais (D) e médias móveis sazonais (Q). Considerando uma série hipotética com sazonalidade de 12 períodos escrita como

$$Y_t = \mu_t + N_t, \quad (2.9)$$

onde μ_t é uma função determinística periódica que satisfaz $\mu_t - \mu_{t-12} = 0$ ou

$$(1 - B^{12})\mu_t = 0,$$

sendo B^{12} a série Y_t em 12 períodos anteriores, isto é, Y_{t-12} , e N_t um processo estacionário que pode ser modelado por um ARMA(p, q). Assim, N_t satisfaz

$$\alpha(B)N_t = \beta(B)u_t,$$

onde u_t é ruído branco gaussiano. Aplicando a diferença sazonal $(1 - B)^{12}$ à expressão (2.9), chega-se a

$$(1 - B^{12})Y_t = (1 - B^{12})\mu_t + (1 - B^{12})N_t,$$

a qual se transforma em

$$\alpha(B)W_t = \beta(B)(1 - B^{12})u_t,$$

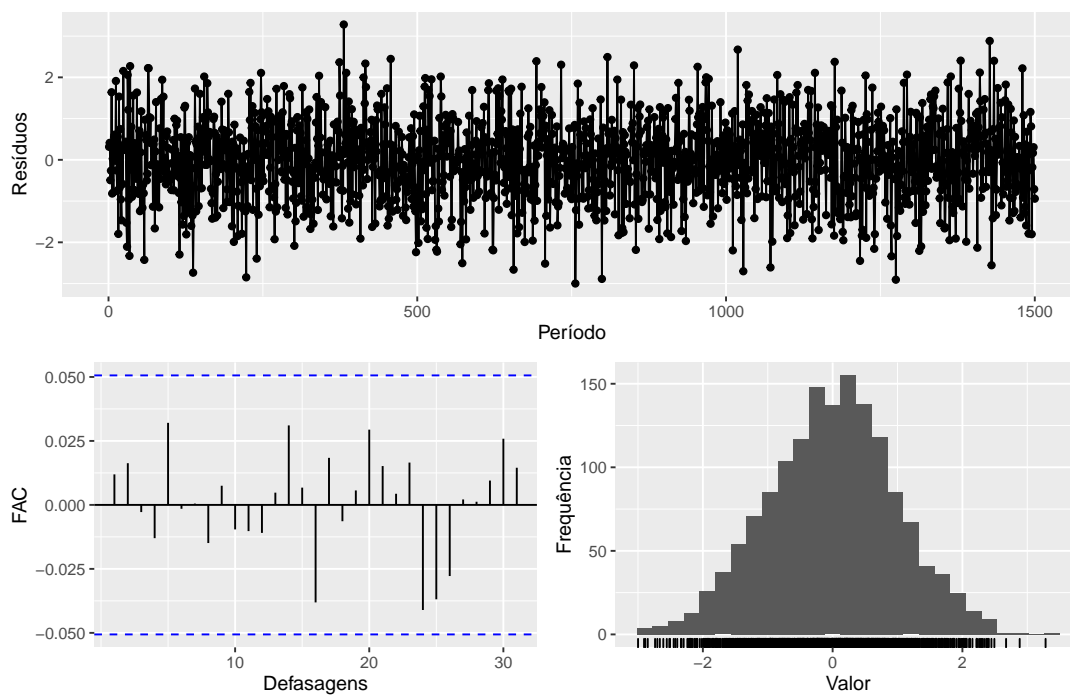
onde $W_t = (1 - B^{12})Y_t$.

Para realizar a estimação de um modelo SARIMA, é importante checar se a série é estacionária. Caso necessário, aplicar a diferenciação e diferenciação sazonal. Com isso, obtém-se os valores de d e D . Em seguida, inspeciona-se os gráficos da FAC e FACP amostrais da série para identificar os valores de P e Q iniciais. Assim, formula-se um modelo SARIMA inicial para descrever a série.

Após a estimação do modelo, é importante verificar algumas suposições dos resíduos. Em geral, assume-se que os resíduos u_t seguem a distribuição normal, isto é, $u_t \sim N(0, \sigma_u^2)$. Visualmente, é possível verificar isso por meio de gráficos dos resíduos como ilustrado na Figura 3, onde o primeiro gráfico mostra a série, o segundo plota a FAC e o terceiro o histograma. Neste caso, por tratar-se de uma série de ruído branco, não há autocorrelação entre as defasagens dos resíduos e o histograma visualmente segue o formato de uma distribuição normal. Uma outra forma, é aplicar os testes de raiz unitária de Kwiatkowski et al. (1992) e Dickey e Fuller (1979)

para checar a estacionariedade, assim como testes de normalidade e independência utilizando os resíduos do modelo ajustado.

Figura 3 – Gráfico de resíduos, FAC e histograma para um ruído branco gaussiano simulado pelo autor.



Fonte: elaborado pelo autor.

2.2 Modelos de Alisamento Exponencial

Há uma outra categoria de modelos chamada de alisamento exponencial. Essa categoria, assim como modelos da classe SARIMA, busca prever o comportamento futuro de uma série temporal com base em padrões repetidos no passado. Para isso, os modelos assumem que valores extremos da série representam aleatoriedade, logo, é possível identificar o padrão básico da série temporal suavizando esses extremos. Esse tipo de modelo é bastante utilizado devido à simplicidade, à eficiência computacional e à razoável precisão (MORETTIN, 2018).

O método mais simples dentro dessa classe de modelos é chamado de médias móveis simples (MMS) e considera uma série localmente constante, isto é, formada por uma média mais um ruído aleatório:

$$Y_t = \mu_t + u_t, \quad t = 1, \dots, N,$$

onde $E(u_t) = 0$, $Var(u_t) = \sigma_u^2$ e μ_t é um parâmetro desconhecido que pode variar com o tempo. Essa técnica consiste basicamente em calcular a média aritmética das r observações mais recentes, ou seja,

$$M_t = \frac{Y_t + Y_{t-1} + \dots + Y_{t-r+1}}{r}.$$

Logo, M_t é uma estimativa da média μ_t que pondera apenas as observações mais atuais.

Há também um segundo tipo de modelo utilizado em séries do tipo localmente constantes: a suavização exponencial simples (SES). Esse tipo de modelo pode ser descrito como

$$\bar{Y}_t = \alpha Y_t + \alpha(1 - \alpha)Y_{t-1} + \alpha(1 - \alpha)^2 Y_{t-2} + \dots,$$

onde $\bar{Y}_0 = Y_1$, $t = 1, \dots, N$, \bar{Y}_t é chamado de valor exponencialmente suavizado e α é a constante de suavização, $0 \leq \alpha \leq 1$. Portanto, a SES é uma média ponderada que atribui pesos maiores às observações mais recentes, o que corrige uma desvantagem do método de MMS.

Em séries com tendência, Holt (2004) propõe o modelo de suavização exponencial de Holt (SEH), visto que o método de SES subestimam ou superestimam continuamente os valores reais da série (MORETTIN, 2018). A SEH é muito similar à SES, a diferença é que a SEH em vez de suavizar apenas o nível, adiciona uma nova constante de suavização para incorporar a tendência da série. Isso é feito pelas duas equações a seguir, sendo o nível estimado pela Equação (2.10) e a tendência pela Equação (2.11):

$$\bar{Y}_t = A Y_t + (1 - A)(\bar{Y}_{t-1} + \hat{T}_{t-1}), \quad 0 < A < 1 \text{ e } t = 2, \dots, N, \quad (2.10)$$

$$\hat{T}_t = C(\bar{Y}_t - \bar{Y}_{t-1} + (1 - C)\hat{T}_{t-1}), \quad 0 < C < 1 \text{ e } t = 2, \dots, N, \quad (2.11)$$

em que A e C são chamadas constantes de suavização.

Para séries com sazonalidade, um dos modelos mais utilizados é a suavização exponencial sazonal de Holt-Winters (WINTERS, 1960). Esse modelo pode ser utilizado por meio de dois

procedimentos diferentes, de acordo com a série trabalhada. No caso de uma série sazonal multiplicativa, Holt-Winters considera F_t como multiplicativo e a tendência aditiva

$$Y_t = \mu_t F_t + T_t + u_t, \quad t = 1, \dots, N.$$

Assim, são propostas três equações de suavização para a série:

$$\hat{F}_t = D \left(\frac{Y_t}{\hat{F}_{t-s}} \right) \hat{F}_{t-s}, \quad 0 < D < 1, \quad t = s + 1, \dots, N, \quad (2.12)$$

$$\bar{Y}_t = A \left(\frac{Y_t}{\hat{F}_{t-s}} \right) + (1 - A)(\hat{Y}_{t-1} + \hat{T}_{t-1}), \quad 0 < A < 1, \quad t = s + 1, \dots, N, \quad (2.13)$$

$$\hat{T}_t = C(\bar{Y}_t - \bar{Y}_{t-1}) + (1 - C)\hat{T}_{t-1}, \quad 0 < C < 1, \quad t = s + 1, \dots, N, \quad (2.14)$$

onde as Equações (2.12), (2.13) e (2.14) representam, respectivamente, a estimativa do fator sazonal, do nível e da tendência, sendo A, C e D as constantes de suavização.

Para o caso da série sazonal aditiva, tem-se

$$Y_t = \mu_t + T_t + F_t + u_t,$$

em que as estimativas do fator sazonal, do nível e da tendência são dadas, respectivamente, pelas Equações (2.15), (2.16) e (2.17):

$$\hat{F}_t = D(Y_t - \bar{Y}_t) + (1 - D)\hat{F}_{t-s}, \quad 0 < D < 1, \quad (2.15)$$

$$\bar{Y}_t = A(Y_t - \hat{Y}_{t-s}) + (1 - A)(\bar{Y}_{t-1} + \hat{T}_{t-1}), \quad 0 < A < 1, \quad (2.16)$$

$$\hat{T}_t = C(\bar{Y}_t - \bar{Y}_{t-1}) + (1 - C)\hat{T}_{t-1}, \quad 0 < C < 1, \quad (2.17)$$

sendo A, C e D são as constantes de suavização.

2.3 Critérios para seleção de modelos

Sob a metodologia de Box-Jenkins, na etapa de identificação são selecionados diversos modelos com base na interpretação da série e dos gráficos da FAC e FACP. Após a estimativa dos modelos, é importante fazer uma verificação do diagnóstico dos resultados estimados. Uma das formas mais utilizadas para isso é a métrica proposta por Akaike (1974), chamada de Critério de Informação de Akaike (AIC). Escolhe-se o modelo cujas ordens k e l minimizam o critério

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

onde $\hat{\sigma}_{k,l}^2$ é a estimativa da variância residual de um modelo ARMA(k, l) para uma série com N observações. Estipula-se valores limites superiores K e L para k e l , calculando então todas as combinações (k, l) com $0 \leq k \leq K$ e $0 \leq l \leq L$.

Um outro método de seleção de modelos é o Critério de Informação Bayesiano (BIC), em que minimiza-se a Equação (2.3), no caso de um processo ARMA (MORETTIN, 2018):

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

em que $\hat{\sigma}_{k,l}^2$ é a estimativa de máxima verossimilhança da variância residual do modelo ARMA(k, l) e N é o tamanho da amostra.

Selecionado o modelo, é importante também fazer uma avaliação da capacidade preditiva para os dados da série. Em geral, essa previsão pode ser avaliada em duas bases diferentes: na amostra utilizada para estimar (dados de treino) e também em dados separados para avaliar a performance preditiva do modelo (dados de teste). As medidas usualmente utilizadas são:

- **MAE** - Erro absoluto médio (**Mean absolute error**):

$$MAE = \frac{\sum_{t=1}^n |Y_t - \hat{Y}_t|}{N}$$

- **MAPE** - Erro percentual absoluto médio (**Mean absolute percentage error**):

$$MAPE = \frac{\sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right|}{N} \cdot 100, Y_t \neq 0$$

- **MPE** - Erro percentual médio (**Mean percentage error**):

$$MPE = \frac{\sum_{t=1}^n \left(\frac{Y_t - \hat{Y}_t}{Y_t} \right)}{N} \cdot 100, Y_t \neq 0$$

- **RMSE** - Raiz do erro quadrático médio (**Root mean squared error**):

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (Y_t - \hat{Y}_t)^2}{N}},$$

onde Y_t é a observação no período t , \hat{Y}_t é o valor estimado para o período t e N é o número de observações da série. Por serem medidas de erro, quanto menor for o valor calculado, mais precisa é a série. Como o interesse da previsão é fora da amostra, para avaliar a performance do modelo nesse quesito deve-se considerar as métricas calculadas para a base de teste (WOOLDRIDGE, 2012).

Sob a abordagem dos modelos de Alisamento Exponencial a avaliação é feita apenas com base nas medidas de performance preditiva (MAE, MAPE, MPE e RMSE), visto que AIC e BIC dependem da existência de uma função de probabilidade associada ao modelo (função de verossimilhança), o que não ocorre para tais métodos.

2.4 Software

O software utilizado para a estimação dos modelos foi o software livre R, usando a interface gratuita para uso individual RStudio. Dentro do R, foram utilizados os pacotes:

- **tidyverse** para manipulação de dados e criação de gráficos (WICKHAM et al., 2019)
- **readxl** para leitura de dados em formato de planilha (WICKHAM; BRYAN, 2022)
- **tsibble** para criar dados na estrutura de séries temporais (WANG; COOK; HYNDMAN, 2020)
- **feasts** para métodos de decomposição e extração de medidas de séries temporais (O'HARA-WILD; HYNDMAN; WANG, 2022)
- **forecast** para realização de previsões (HYNDMAN; KHANDAKAR, 2008)
- **gridExtra** para organização de múltiplos gráficos (AUGUIE, 2017)
- **kableExtra** para criação de tabelas em formato $\text{L}^{\text{T}}\text{E}^{\text{X}}$ (ZHU, 2021)

3 Aplicação

3.1 Base de Dados

A base de dados que será considerada neste trabalho é composta por dados da receita mensal de uma empresa de consultoria em tecnologia. A receita é o valor recebido pela empresa pela prestação dos serviços que são fornecidos por ela. Portanto, trata-se de uma série temporal financeira com frequência mensal. Os dados foram disponibilizados para a pesquisa e possuem intervalo de data entre **Janeiro de 2012** e **Junho de 2022**.

Por tratar-se de uma série financeira internacional, os dados foram reajustados pela inflação utilizando o índice de preços ao consumidor do **Federal Reserve Bank of St. Louis (Consumer Price Index - CPI)**, com dados extraídos do website da própria organização para o mesmo período de dados da base de receita. A base de dados com a receita deflacionada foi dividida em duas:

- **Treino:** Dados de Janeiro/2012 a Junho/2021 ($n = 114$);
- **Teste:** Dados de Julho/2021 a Junho/2022 ($n = 12$).

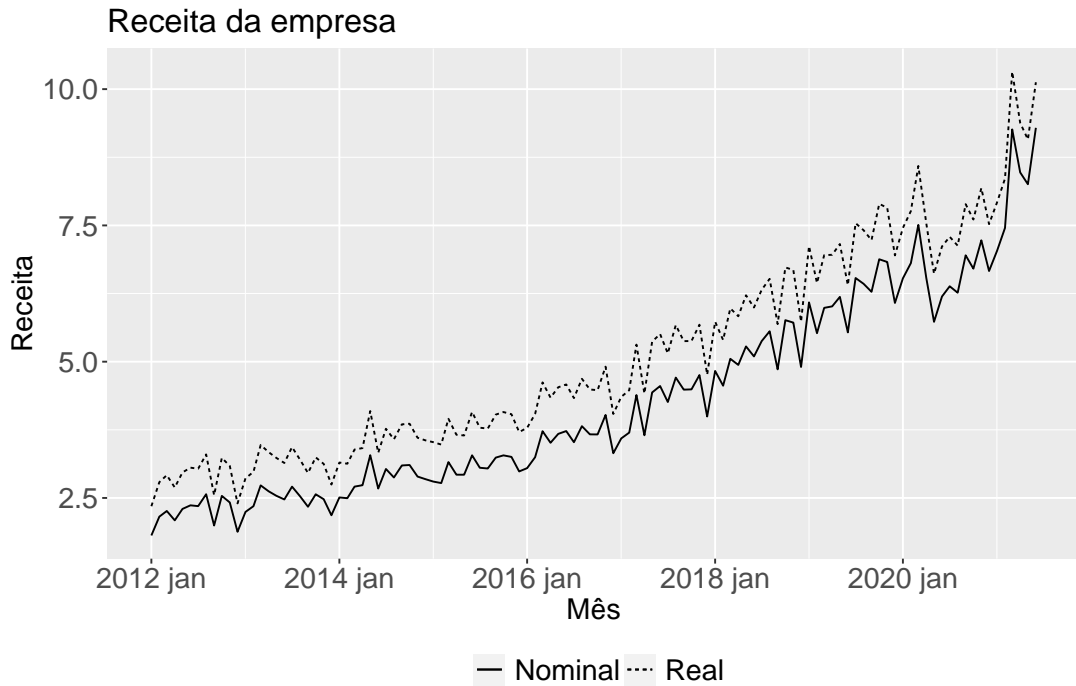
A base de treino foi utilizada para a realização de toda a análise exploratória e modelagem, enquanto a base de teste foi usada para avaliar a capacidade preditiva dos modelos SARIMA e Alisamento Exponencial que foram aplicados aos dados.

3.2 Análise Exploratória dos Dados

Para a análise, o primeiro ponto observado é o comportamento da série ao longo do tempo. A Figura 4 demonstra a evolução da receita mensal no formato nominal, sem reajuste pela inflação, e no formato real, com a série deflacionada. É possível observar que a série apresenta tendência positiva, pequenas variações sazonais e também uma ligeira quebra na tendência em março e abril de 2020, que logo retorna à tendência anterior em junho do mesmo ano. Este fenômeno muito provavelmente ocorreu por efeitos da pandemia da COVID-19 nas atividades da companhia.

Em seguida, é necessário analisar com mais detalhes a sazonalidade da receita da empresa. Na Figura 5 é possível visualizar o comportamento da série de acordo com os meses do ano. Nota-se que há um aumento da mediana nos meses de março, junho e novembro. Há também uma ligeira redução em janeiro e dezembro. O aumento ocorrido em março e junho possivelmente é causado pelo fim de trimestre, quando pode-se haver o reconhecimento de

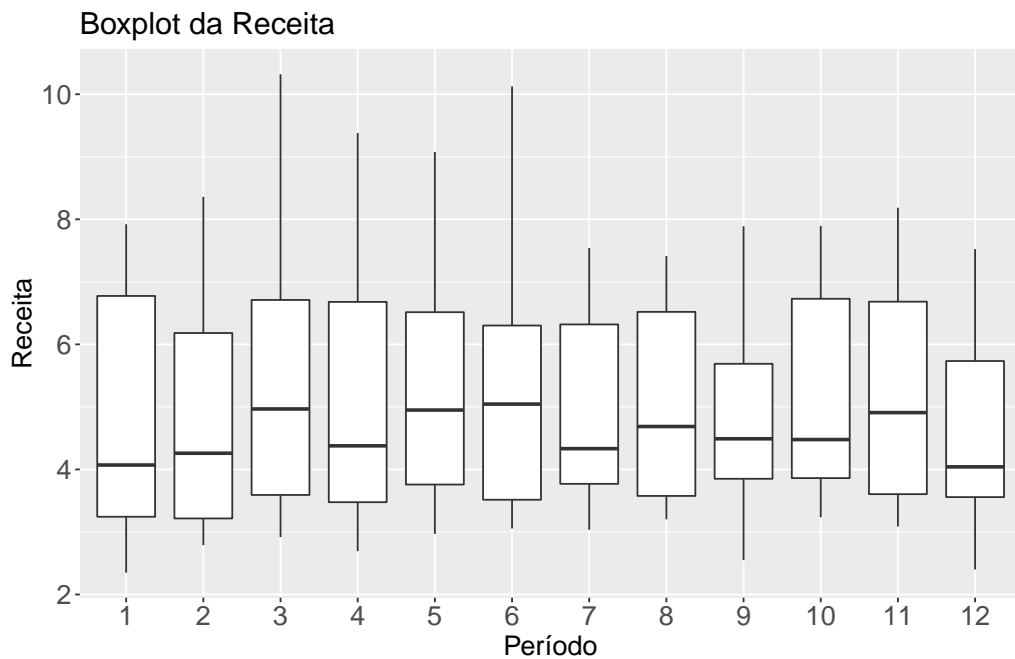
Figura 4 – Receita da empresa no período de treino.



Fonte: elaborado pelo autor.

receitas ocorridas nos meses anteriores mas não cobradas. Em novembro, há a possibilidade de acontecer uma aceleração nas vendas como antecipação às férias que ocorrem encerramento do ano, em dezembro, e também no início do ano, em janeiro.

Figura 5 – Diagrama de caixa da receita no período de treino.

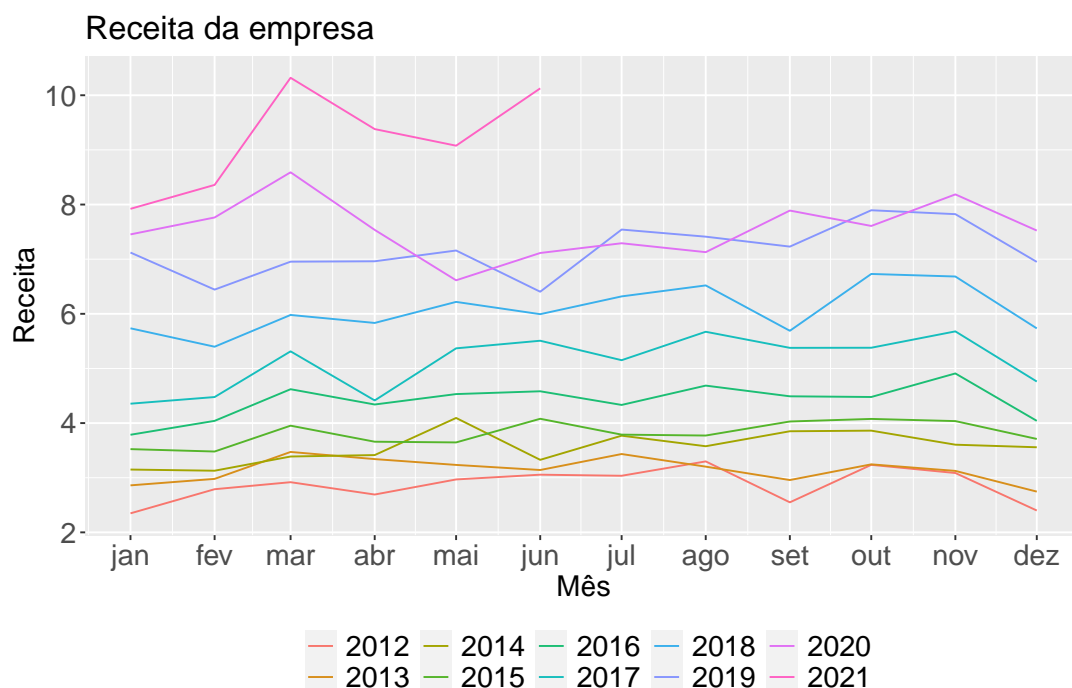


Fonte: elaborado pelo autor.

As variações citadas anteriormente também são visualizadas na Figura 6, que mostra as

séries anuais em linhas separadas. Ocorrem pequenos saltos no mês de março em praticamente todos os anos. O aumento de junho já ocorre de forma menos consistente. Para as quedas, os padrões de dezembro são bastante consistentes, enquanto em janeiro não é tão perceptível, tendo em vista a quebra no gráfico.

Figura 6 – Gráfico de sazonalidade da receita no período de treino.



Fonte: elaborado pelo autor.

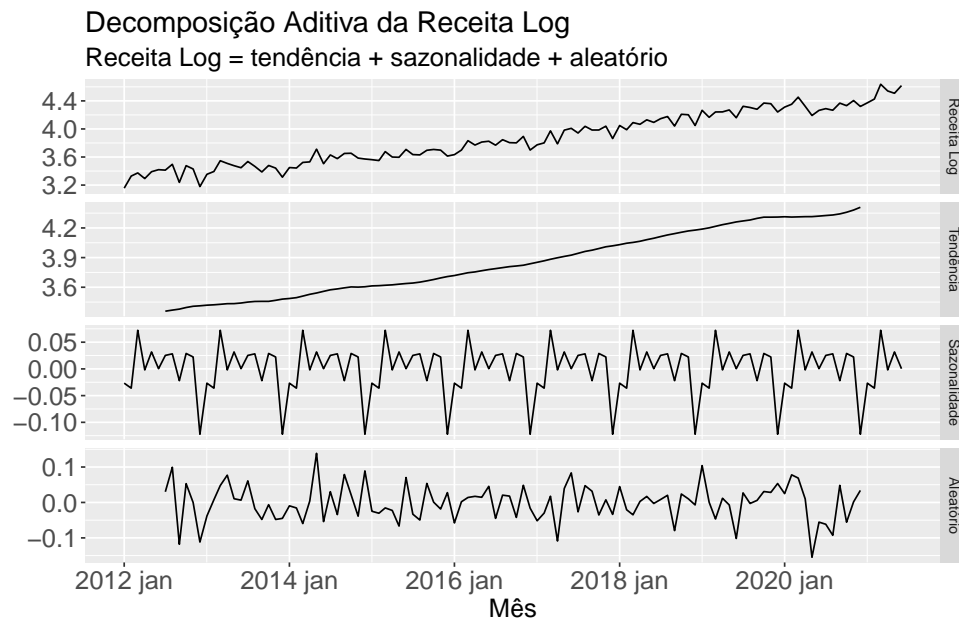
Considerando que a variância da série não é constante ao longo do tempo, é possível transformá-la aplicando o *log* nos dados. A série transformada e a decomposição dela podem ser vistas na Figura 7, que indica aspectos similares aos pontos observados anteriormente. A tendência positiva da série é visualmente evidente, enquanto há também um padrão sazonal bem definido nos meses de março e dezembro.

O boxplot na Figura 8 indica um padrão semelhante na sazonalidade da série antes da transformação, na qual existem picos nos meses de março, junho e novembro, e quedas nos meses de janeiro e dezembro.

Apesar de reduzida a variância, a série ainda não aparenta ser estacionária, como é possível observar na Figura 9, em que há um decaimento muito lento nas barras de acordo com os lags no gráfico de FAC. Isso indica ser necessário transformá-la novamente transformando a série com uma diferença. Essa diferença é aplicada e visualizada na Figura 10, onde a série visualmente aparenta ser mais estacionária e tem poucos picos pontuais fora do intervalo de confiança nos gráficos de FAC e FACP. No gráfico de FAC, os picos concentram-se nos lags 1 e 12, o que pode estar associado à sazonalidade mencionada anteriormente.

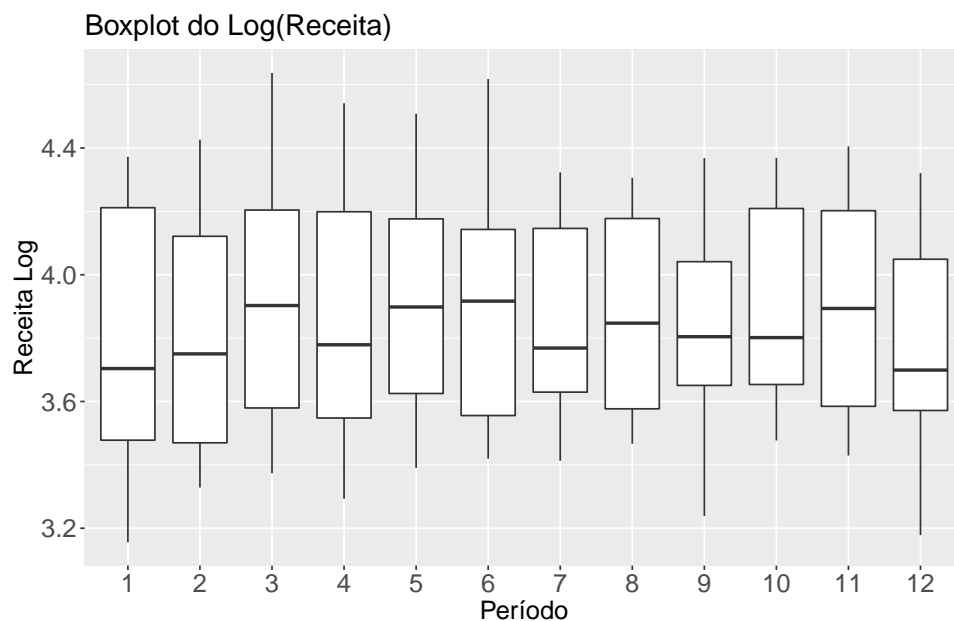
O teste de raiz unitária de Kwiatkowski–Phillips–Schmidt–Shin (KPSS) foi aplicado

Figura 7 – Decomposição do log da receita no período de treino.



Fonte: elaborado pelo autor.

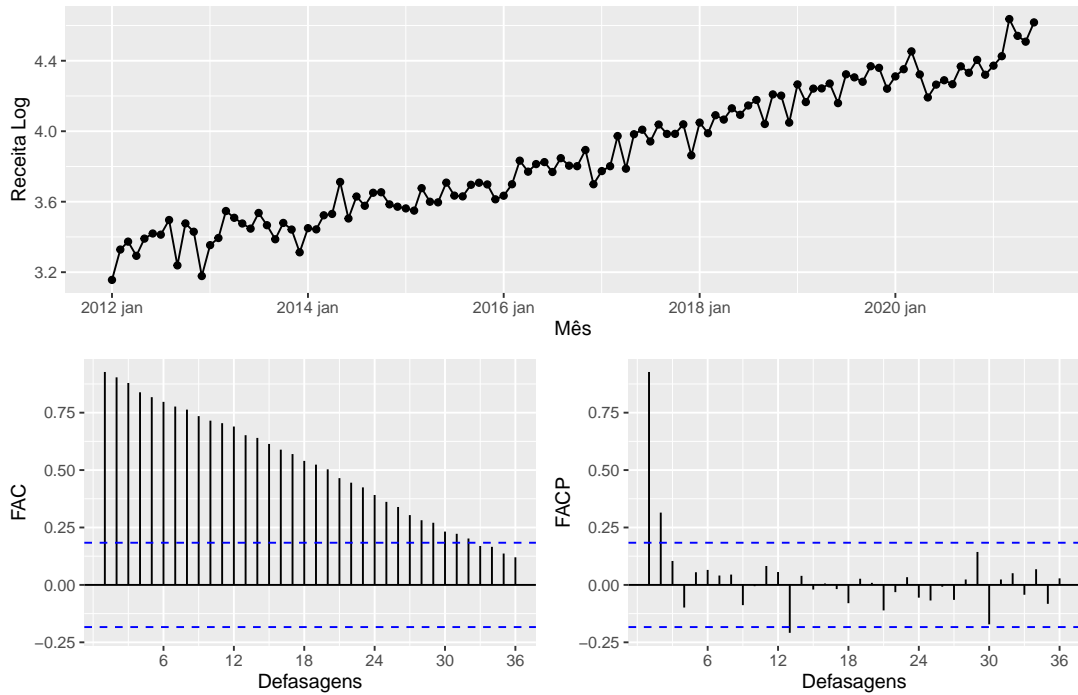
Figura 8 – Diagrama de caixa do log da receita no período de treino



Fonte: elaborado pelo autor.

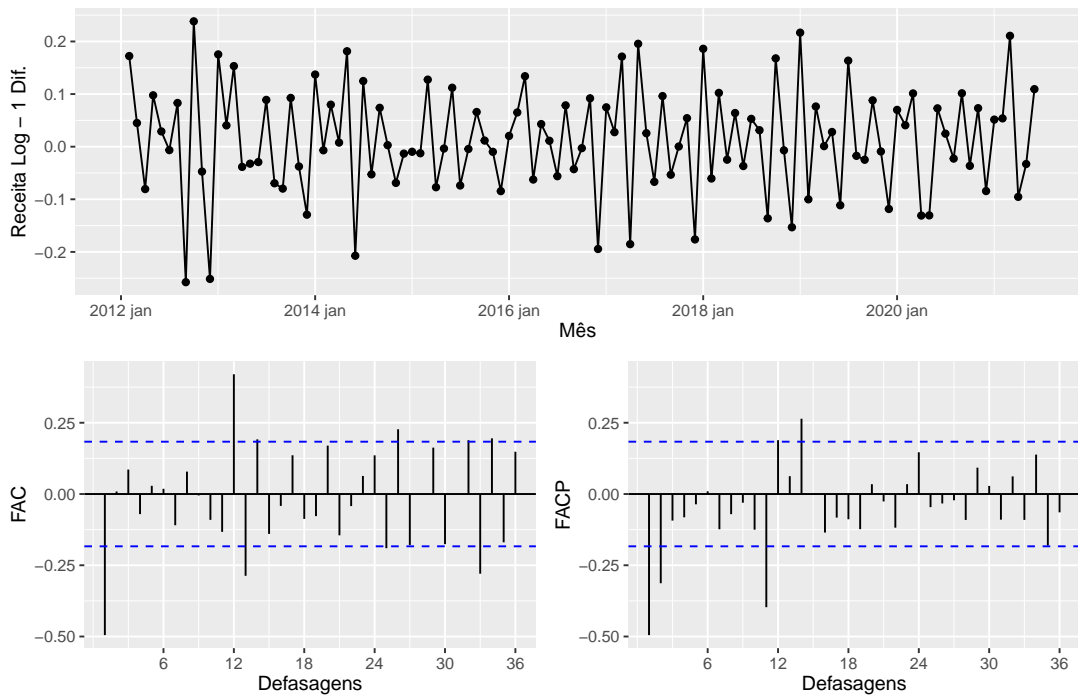
para as duas séries (*log* e *log* diferenciada). No teste da série *log*, a estatística é de 2,37 e o p-valor é de 0,01, portanto, a hipótese nula de que os dados são estacionários é rejeitada ao nível de 5% de significância. Por outro lado, para o teste com a série *log* diferenciada, a estatística é de 0,0397 e o p-valor é de 0,1, o que indica que não rejeitamos a hipótese nula e, portanto, a série é estacionária.

Figura 9 – Log da receita no período de treino.



Fonte: elaborado pelo autor.

Figura 10 – Log da receita com uma diferença no período de treino.



Fonte: elaborado pelo autor.

3.3 Modelagem

3.3.1 Definição dos Modelos

Como mencionado anteriormente, a série precisou de uma diferença para tornar-se estacionária, portanto $d = 1$. O gráfico de FAC apresenta pontos acima do intervalo de confiança nos lags 1, 12 e 13. Dessa forma, serão considerados modelos com vetor AR de ordem 0, 1 e 2. Por ter duas autocorrelações parciais negativas no gráfico de FACP, podemos supor que a série possui comportamento de média móvel. Logo, serão testados modelos com ordem MA 0 e 1. Para o componente sazonal, serão testados modelos com ordem 0 e 1 para todos os fatores (P, D, Q) , tendo em vista a presença de pontos sazonais significantes nos gráficos de FAC e FACP.

O software R possui uma função chamada `auto.arima()` que identifica o modelo com melhor ajuste de acordo com a minimização do AIC. O modelo identificado para os dados foi um SARIMA(0,1,1)(1,0,0). Além dos modelos SARIMA, será estimado também um modelo de alisamento exponencial para a realização de previsões na base de treino, visto que essa categoria de modelos costuma apresentar bom poder preditivo para alguns tipos de série, como visto na seção de metodologia. Dessa forma, todos os modelos a seguir foram estimados para comparação:

- Alisamento Exponencial
- `auto.arima()`: SARIMA(0,1,1)(1,0,0)
- SARIMA(0,1,1)(0,0,1)
- SARIMA(1,1,1)(0,0,1)
- SARIMA(1,1,0)(1,1,0)
- SARIMA(2,1,0)(1,1,0)
- SARIMA(0,1,1)(0,1,1)

Os parâmetros estimados podem ser visualizados na Tabela 3 e as métricas de avaliação estão na Tabela 4. O modelo com menor AIC é o modelo estimado automaticamente pelo software, o que já era esperado, como citado anteriormente. Entretanto, outros três modelos apresentaram performance bastante similar tanto em AIC, quanto em RMSE. Dessa forma, será feita a análise de resíduos dos quatro modelos SARIMA e do modelo de alisamento exponencial, que funcionará como uma espécie de **benchmark**.

Tabela 3 – Parâmetros estimados para os modelos SARIMA (dados de treino).

Modelo	Termo	Estimativa	Erro Padrão	Estatística	P-valor
SARIMA(0,1,1)(1,0,0)	ma1	-0.6816	0.0860	-7.9260	0.0000
	sar1	0.5068	0.0913	5.5521	0.0000
	constant	0.0060	0.0020	3.0208	0.0031
SARIMA(0,1,1)(0,0,1)	ma1	-0.6820	0.0909	-7.5027	0.0000
	sma1	0.4898	0.1034	4.7372	0.0000
	constant	0.0120	0.0032	3.7197	0.0003
SARIMA(1,1,1)(0,0,1)	ar1	-0.1295	0.1604	-0.8072	0.4212
	ma1	-0.5799	0.1468	-3.9487	0.0001
	sma1	0.5008	0.1017	4.9260	0.0000
	constant	0.0140	0.0043	3.2939	0.0013
SARIMA(1,1,0)(1,1,0)	ar1	-0.5114	0.0865	-5.9112	0.0000
	sar1	-0.2174	0.1034	-2.1022	0.0380
SARIMA(2,1,0)(1,1,0)	ar1	-0.7673	0.0901	-8.5206	0.0000
	ar2	-0.4833	0.0889	-5.4388	0.0000
	sar1	-0.2850	0.1064	-2.6790	0.0086
SARIMA(0,1,1)(0,1,1)	ma1	-0.7004	0.0842	-8.3189	0.0000
	sma1	-0.8050	0.1511	-5.3287	0.0000

Fonte: elaborado pelo autor.

Tabela 4 – Métricas de avaliação dos modelos ajustados (dados de treino).

Modelo	AIC	BIC	MAE	MAPE	MPE	RMSE
SARIMA(0,1,1)(1,0,0)	-263.17	-252.26	0.0547	1.4376	0.0003	0.0712
SARIMA(0,1,1)(0,0,1)	-259.82	-248.91	0.0563	1.4775	-0.0063	0.0724
SARIMA(1,1,1)(0,0,1)	-258.36	-244.72	0.0553	1.4521	-0.0116	0.0722
SARIMA(0,1,1)(0,1,1)	-233.94	-226.09	0.0492	1.2663	0.0446	0.0651
SARIMA(2,1,0)(1,1,0)	-230.53	-220.07	0.0506	1.3042	0.0291	0.0693
SARIMA(1,1,0)(1,1,0)	-206.81	-198.96	0.0606	1.5639	0.0172	0.0791
Alisamento Exponencial	-	-	0.0492	1.2860	0.0064	0.0610

Fonte: elaborado pelo autor.

3.3.2 Análise de Resíduos

3.3.2.1 SARIMA(0,1,1)(1,0,0)

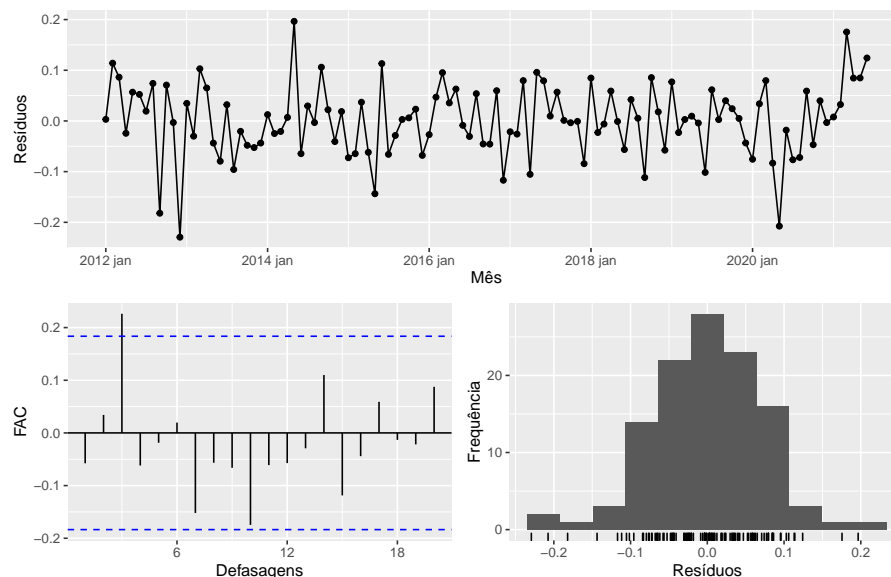
Visualmente, os resíduos do modelo parecem ser não correlacionados, como visto na Figura 11. Há somente uma pequena diferença na variância por volta de 2013 e 2020. No gráfico de FAC, todos os lags estão dentro do intervalo, com exceção do terceiro lag, que está ligeiramente acima. No histograma, a distribuição aparenta ser muito similar à Normal. Há uma pequena diferença no extremo negativo, onde há um pequeno pico.

Já o Q-Q plot (Figura 12) do modelo apresenta valores bem distribuídos sobre a linha teórica. As variações estão mais presentes nos extremos, possivelmente associadas às maiores variações da série em determinados períodos.

Os testes de Shapiro-Wilk, Ljung-Box e Box-Pierce todos possuem p-valor maior que

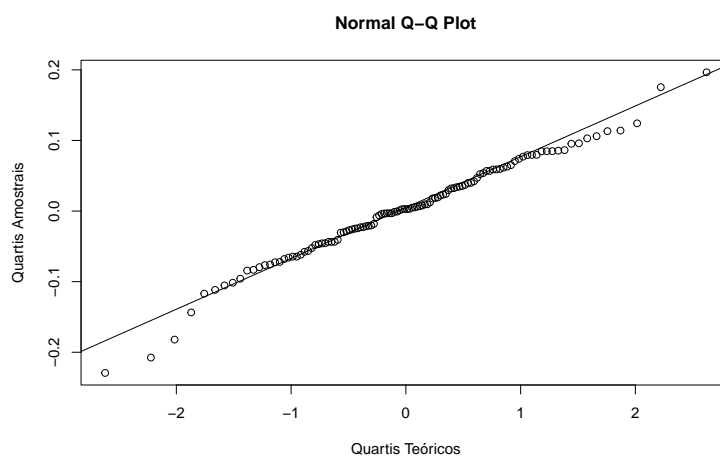
0,05, o que indica que não é rejeitada a hipótese nula de normalidade dos resíduos (Shapiro-Wilk) e de ausência de autocorrelação nos resíduos (Ljung-Box e Box-Pierce).

Figura 11 – Resíduos do modelo SARIMA(0,1,1)(1,0,0) no período de treino.



Fonte: elaborado pelo autor.

Figura 12 – Q-Q plot do modelo SARIMA(0,1,1)(1,0,0)



Fonte: elaborado pelo autor.

3.3.2.2 SARIMA(0,1,1)(0,0,1)

A distribuição dos resíduos do modelo SARIMA(0,1,1)(0,0,1) possui formato similar à normal, porém com uma ligeira assimetria à direita, como visto no terceiro gráfico da Figura 13. A série de resíduos possui alguns pontos extremos, basicamente nos mesmos pontos que o modelo anterior. Já o gráfico de FAC, indica que todos os pontos estão dentro do intervalo de confiança.

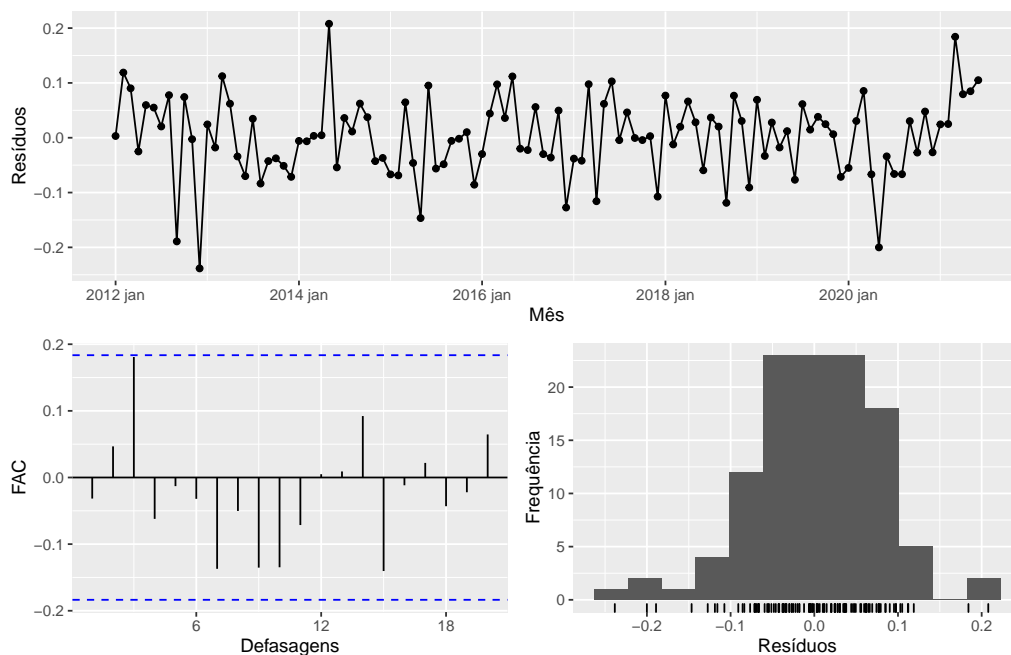
Neste caso, o Q-Q plot (Figura 14) do modelo apresenta mais pontos fora da linha teórica, quando comparado ao modelo anterior. Estes pontos, entretanto, concentram-se nos valores

Tabela 5 – Testes para normalidade dos resíduos - Modelo SARIMA(0,1,1)(1,0,0)

Teste	Estatística	P-valor
Shapiro	0.9835497	0.1763137
Ljung-Box	0.3890062	0.5328223
Box-Pierce	0.3789456	0.5381680

Fonte: elaborado pelo autor.

Figura 13 – Resíduos do modelo SARIMA(0,1,1)(0,0,1) no período de treino



Fonte: elaborado pelo autor.

Tabela 6 – Testes para normalidade dos resíduos - Modelo SARIMA(0,1,1)(0,0,1)

Teste	Estatística	P-valor
Shapiro	0.9822073	0.1346307
Ljung-Box	0.1170020	0.7323091
Box-Pierce	0.1139761	0.7356621

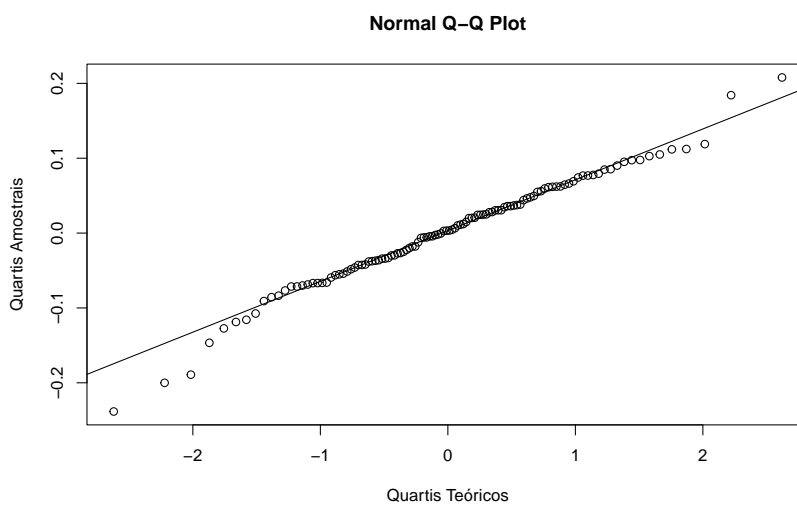
Fonte: elaborado pelo autor.

mais extremos. Além disso, os testes vistos na Tabela 6 todos não rejeitam a hipótese nula de normalidade, no caso do teste de Shapiro-Wilk, ainda que com p-valor menor, e de ausência de autocorrelação, no caso dos testes de Ljung-Box e de Box-Pierce.

3.3.2.3 SARIMA(1,1,1)(0,0,1)

O modelo SARIMA(1,1,1)(0,0,1) possui distribuição dos resíduos visualmente muito próxima à distribuição Normal (Figura 15). Há somente uma ligeira assimetria à direita, porém muito menor que a encontrada no modelo SARIMA(0,1,1)(0,0,1). O gráfico de FAC também apresenta todos os dados dentro do intervalo de confiança, o que nos indica inicialmente que os resíduos estão seguindo a distribuição Normal.

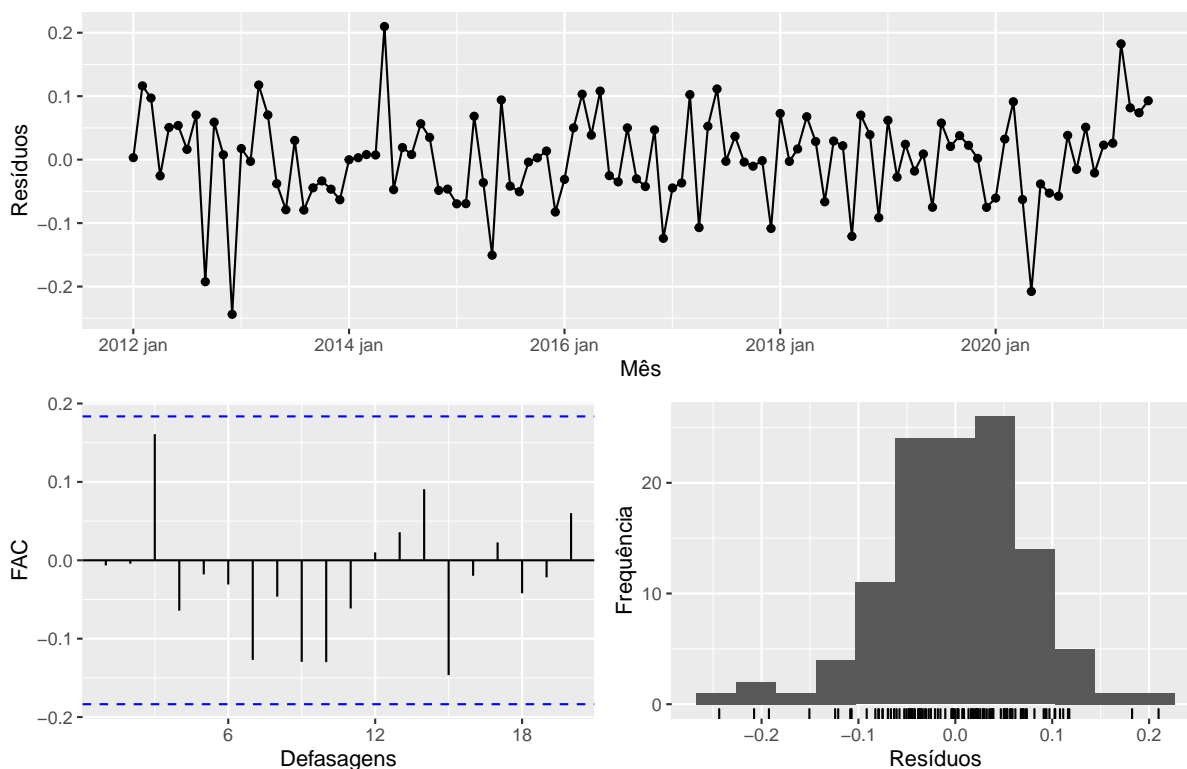
Figura 14 – Q-Q plot do modelo SARIMA(0,1,1)(0,0,1)



Fonte: elaborado pelo autor.

O Q-Q plot (Figura 16) está bastante similar ao modelo anterior, o que de certa forma era esperado, tendo em vista que ambos possuem especificações muito similares, mudando apenas o fator AR que neste caso é igual a 1. Os pontos fora da reta concentram-se mais nos extremos.

Figura 15 – Resíduos do modelo SARIMA(1,1,1)(0,0,1) no período de treino



Fonte: elaborado pelo autor.

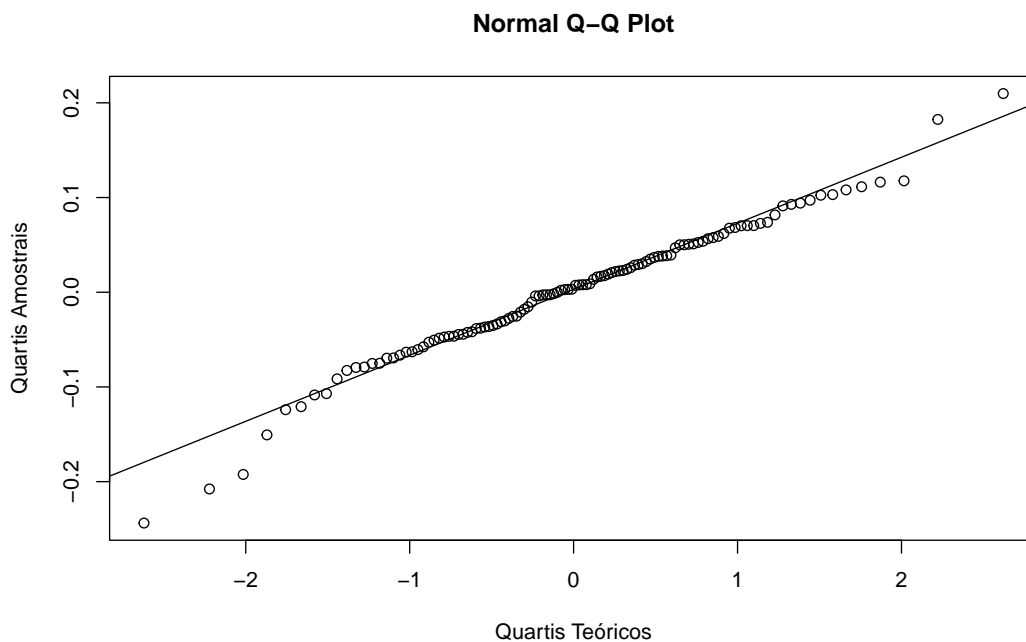
Tabela 7 – Testes para normalidade dos resíduos - Modelo SARIMA(1,1,1)(0,0,1)

Teste	Estatística	P-valor
Shapiro	0.9782703	0.0604072
Ljung-Box	0.0050914	0.9431160
Box-Pierce	0.0049597	0.9438552

Fonte: elaborado pelo autor.

A maior diferença deste modelo encontra-se nos resultados dos testes. No teste de normalidade de Shapiro-Wilk, o p-valor é relativamente menor que o dos outros modelos, chegando muito próximo ao nível de 0,05 de significância. Entretanto, mesmo sendo mais próximo, o teste falha em rejeitar a hipótese nula, não sendo necessário descartar o modelo por essa ótica. Nos testes de ausência de independência dos resíduos Ljung-Box e Box-Pierce, o p-valor é bastante alto e falha em descartar a hipótese nula de ausência de autocorrelação.

Figura 16 – Q-Q plot do modelo SARIMA(1,1,1)(0,0,1)



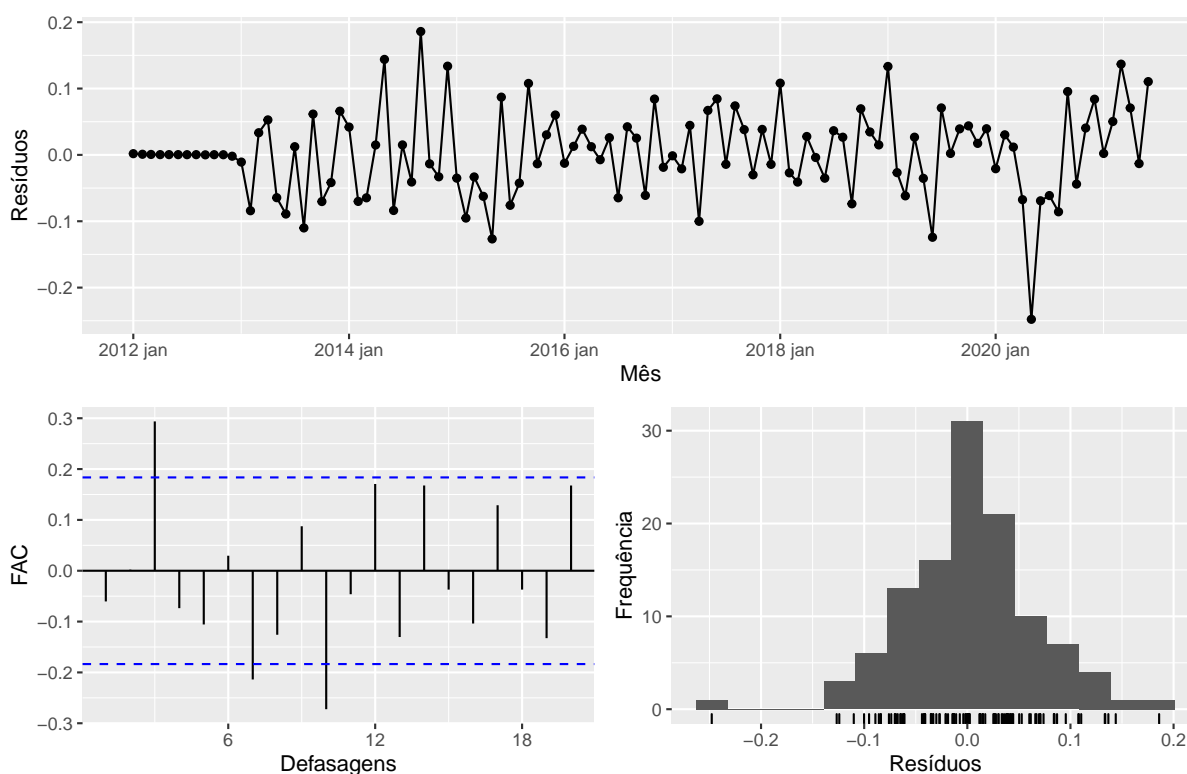
Fonte: elaborado pelo autor.

3.3.2.4 SARIMA(0,1,1)(0,1,1)

O modelo SARIMA(0,1,1)(0,1,1) também possui distribuição dos resíduos visualmente muito próxima à distribuição Normal (Figura 17), a diferença é que os dados concentram-se mais no centro, com ligeira assimetria. O gráfico de FAC apresenta três defasagens fora do intervalo de confiança, o que pode indicar autocorrelação nos resíduos, um ponto a ser validado pelos testes.

O Q-Q plot (Figura 18) possui pontos com um leve deslocamento, especialmente nos quartis positivos, entretanto, ainda sobrepõem em grande parte a reta. Mesmo com este deslocamento, o teste de Shapiro-Wilk falha em rejeitar a hipótese nula de normalidade dos dados. Já nos teste de Ljung-Box e Box-Pierce, para autocorrelação dos resíduos, possuem alto p-valor, não sendo rejeita então a hipótese nula de ausência de autocorrelação nos resíduos.

Figura 17 – Resíduos do modelo SARIMA(0,1,1)(0,1,1) no período de treino



Fonte: elaborado pelo autor.

Tabela 8 – Testes para normalidade dos resíduos - Modelo SARIMA(0,1,1)(0,1,1)

Teste	Estatística	P-valor
Shapiro	0.9823809	0.139437
Ljung-Box	0.4267925	0.513567
Box-Pierce	0.4157548	0.519063

Fonte: elaborado pelo autor.

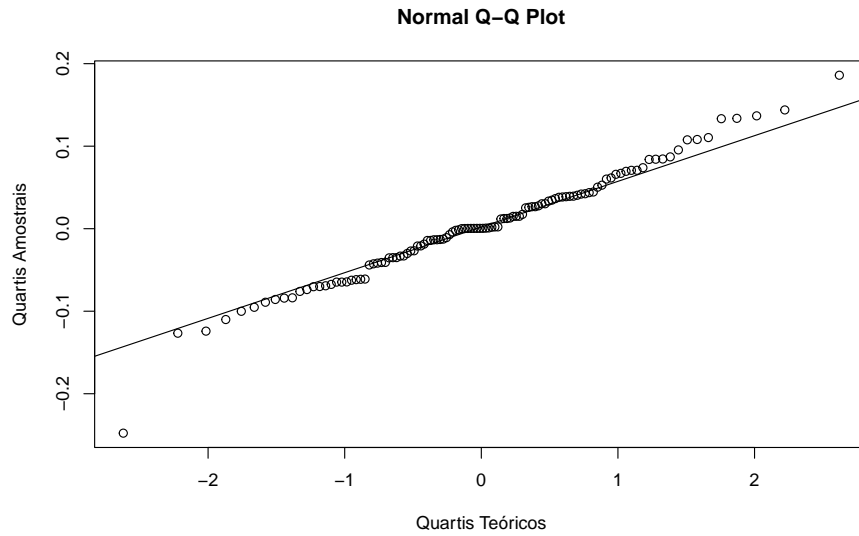
3.3.2.5 SARIMA(2,1,0)(1,1,0)

O último modelo analisado apresenta resíduos com distribuição graficamente (Figura 19) muito similar à normal. Os únicos pontos de atenção são os extremos, em que há algumas observações **outliers** mais distantes. O gráfico de FAC mostra que todas as autocorrelações estão dentro do intervalo de confiança.

O Q-Q plot (Figura 20) possui pontos bastante alinhados com a linha teórica, com exceção do centro que tem um ligeiro deslocamento na reta. Há também cerca de três pontos nos extremos que se afastam ligeiramente, possivelmente associados aos **outliers** citados anteriormente.

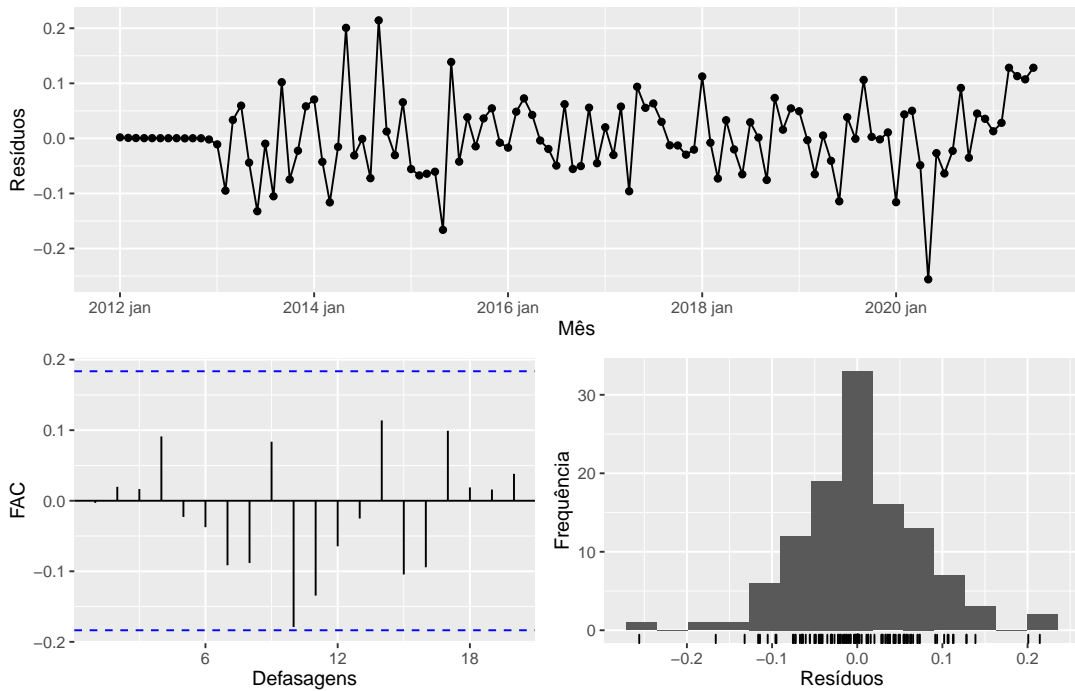
Nos testes, o modelo foi afetado pela redução na amostra pelas duas diferenças. Além disso, as observações nos extremos da distribuição de resíduos impactaram diretamente no teste de Shapiro-Wilk, que apresentou p-valor abaixo de 0,05, rejeitando-se então a hipótese nula de

Figura 18 – Q-Q plot do modelo SARIMA(0,1,1)(0,0,1)



Fonte: elaborado pelo autor.

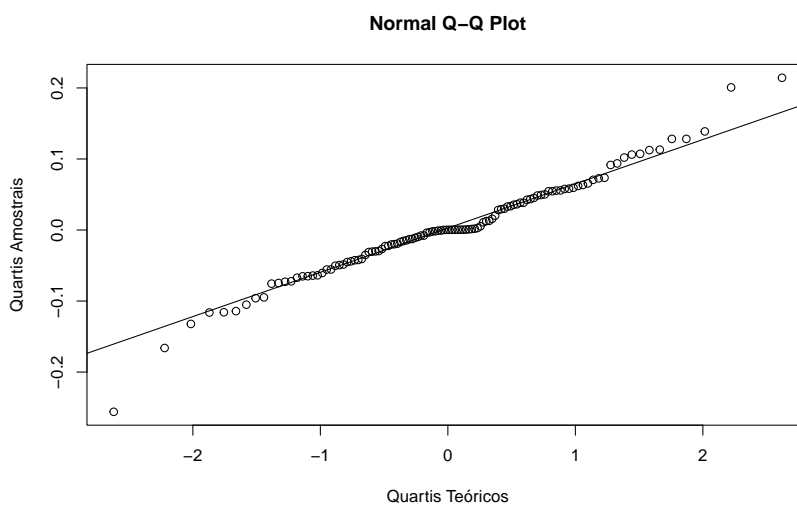
Figura 19 – Resíduos do modelo SARIMA(2,1,0)(1,1,0) no período de treino



Fonte elaborado pelo autor.

normalidade dos resíduos. Entretanto, nos outros dois testes, Ljung-Box e Box-Pierce, a hipótese nula de independência dos resíduos não foi rejeitada. Dessa forma, o modelo não será descartado de imediato. Será feita ainda uma avaliação da acurácia no período de testes.

Figura 20 – Q-Q plot do modelo SARIMA(2,1,0)(1,1,0)



Fonte elaborado pelo autor.

Tabela 9 – Testes para normalidade dos resíduos - Modelo SARIMA(2,1,0)(1,1,0)

Teste	Estatística	P-valor
Shapiro-Wilk	0.9759021	0.0373348
Ljung-Box	0.0010295	0.9744038
Box-Pierce	0.0010029	0.9747369

Fonte: elaborado pelo autor.

3.3.3 Acurácia dos Modelos

Na base de treino, o modelo de alisamento exponencial apresentou o melhor resultado nas métricas **MAE**, **MAPE** e **RMSE**, enquanto o modelo SARIMA(1,1,1)(0,0,1) teve melhor performance na métrica **MPE**.

Já na base de testes, o modelo com melhor desempenho foi o SARIMA(0,1,1)(1,0,0), com melhores números nas métricas **MAE**, **MAPE** e **RMSE**, sendo o SARIMA(2,1,0)(1,1,0) o melhor na métrica **MPE**.

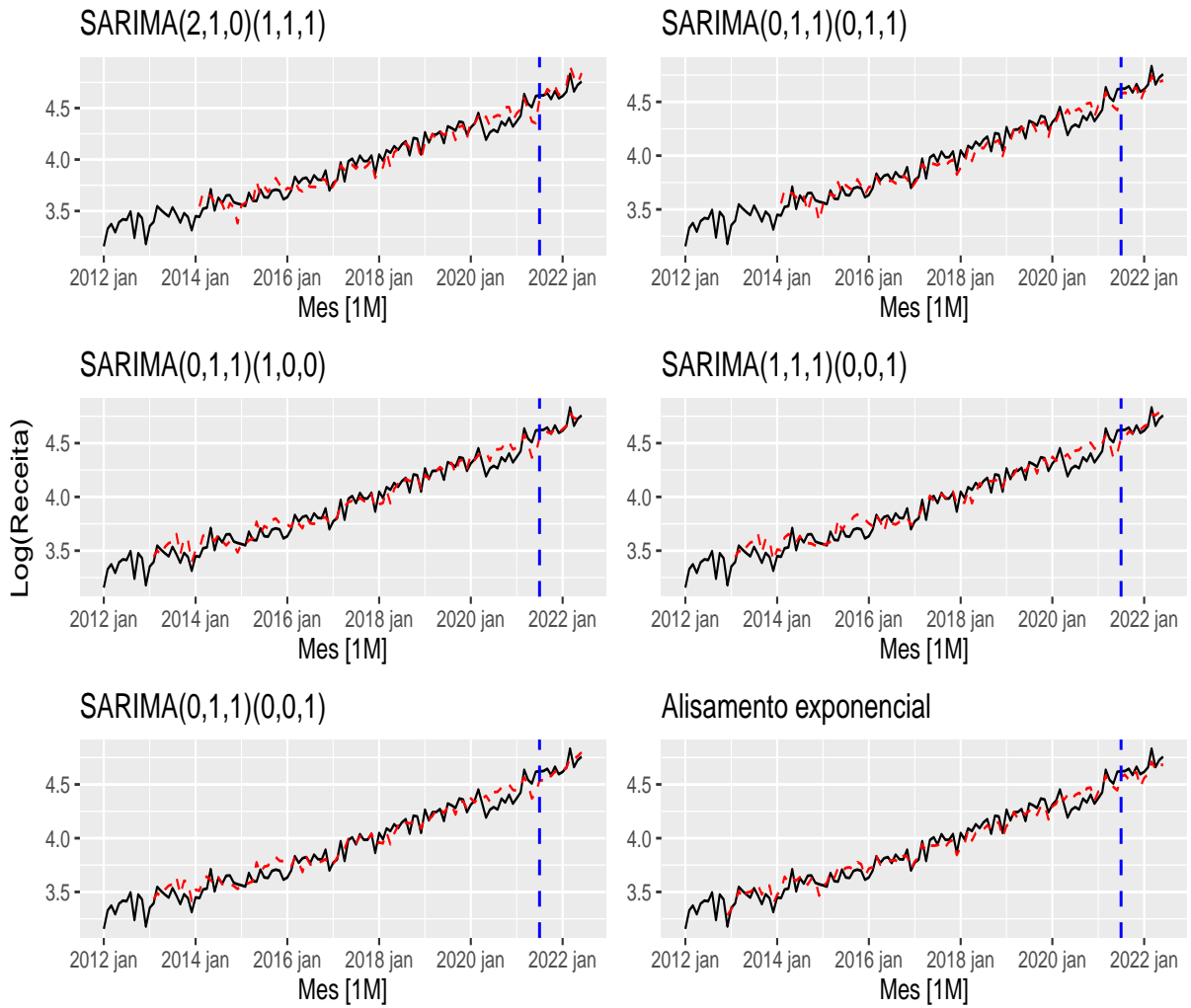
Tabela 10 – Acurácia dos modelos nas duas bases de dados

Modelo	Tipo	MAE	MAPE	MPE	RMSE
Alisamento Exponencial	Treino	0.0492	1.2860	0.0064	0.0610
SARIMA(0,1,1)(1,0,0)	Treino	0.0547	1.4376	0.0003	0.0712
SARIMA(0,1,1)(0,0,1)	Treino	0.0563	1.4775	-0.0063	0.0724
SARIMA(1,1,1)(0,0,1)	Treino	0.0553	1.4521	-0.0116	0.0722
SARIMA(2,1,0)(1,1,0)	Treino	0.0506	1.3042	0.0291	0.0693
Alisamento Exponencial	Teste	0.0630	1.3462	1.2670	0.0715
SARIMA(0,1,1)(1,0,0)	Teste	0.0337	0.7205	0.2837	0.0440
SARIMA(0,1,1)(0,0,1)	Teste	0.0466	0.9947	0.3475	0.0552
SARIMA(1,1,1)(0,0,1)	Teste	0.0485	1.0371	-0.0863	0.0549
SARIMA(2,1,0)(1,1,0)	Teste	0.0567	1.2140	-0.9919	0.0650

Fonte: elaborado pelo autor.

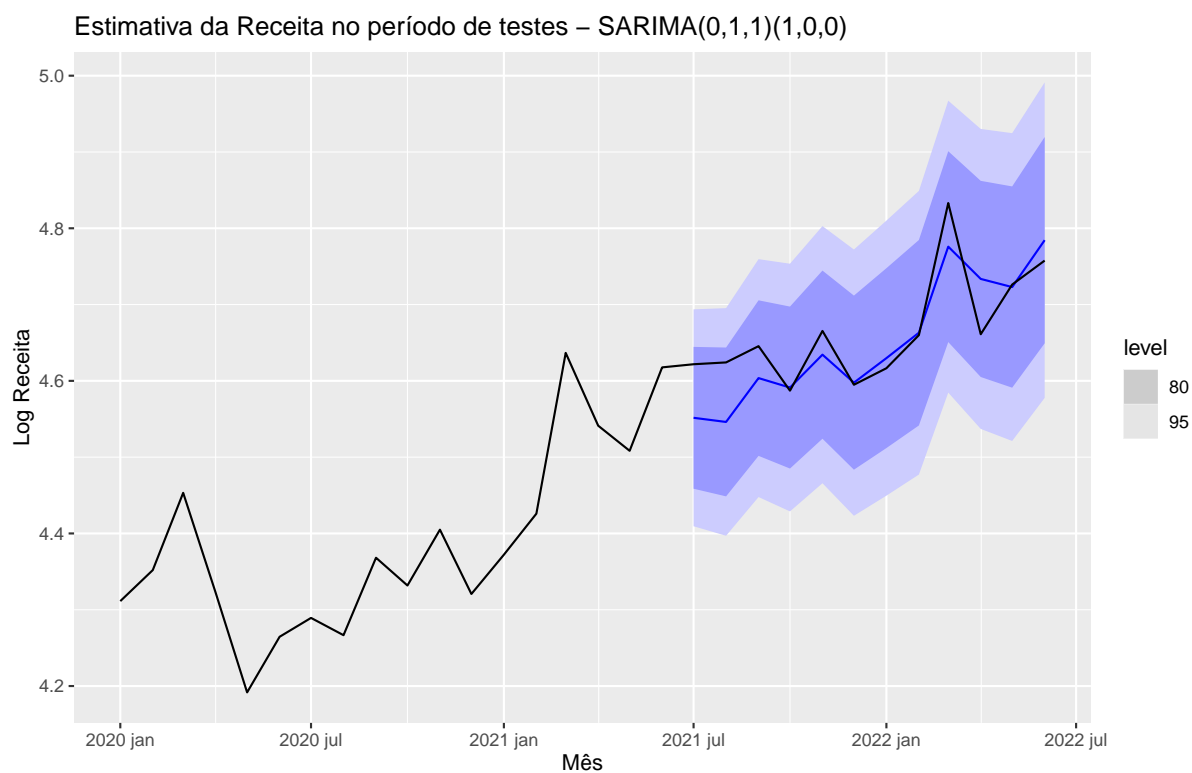
Na Figura 21 é possível visualizar a série real e as séries estimadas para os modelos selecionados, onde a linha tracejada indica o início do período de testes. É possível perceber que o modelo SARIMA(0,1,1)(1,0,0) retornou previsões muito próximas dos valores reais nos dados de teste. Isso também é mostrado pelo gráfico da Figura 22, que foca na previsão para o período de testes. Vale destacar que, apesar de nenhum dos modelos ter retornado boas previsões para a série durante o ano de 2020 devido a uma possível quebra estrutural neste ano, a qual foi possivelmente causada pelos efeitos da pandemia do COVID-19, os valores observados voltaram ao patamar usual a partir de 2021 e a estrutura de dependência de modelos como o SARIMA(0,1,1)(1,0,0) foi capaz de retornar boas previsões.

Figura 21 – Comparativo entre previsão e real por modelo



Fonte elaborado pelo autor.

Figura 22 – Comparativo entre estimativa e real período de testes - SARIMA(0,1,1)(1,0,0)



Fonte elaborado pelo autor.

4 Conclusão

O desenvolvimento deste estudo possibilitou a reflexão sobre como a previsão da receita mensal de uma empresa pode ser desenvolvida utilizando a análise de séries temporais. Isso tem impacto direto na expansão do negócio e visão estratégica de futuro da empresa, considerando a melhora na assertividade e previsibilidade.

Em termos de acurácia, o modelo com melhor resultado é o SARIMA(0,1,1)(1,0,0), que à primeira vista não seria o modelo escolhido, considerando apenas o AIC e a análise de resíduos. Isso mostra que é importante também avaliar a utilidade do modelo de acordo com o potencial preditivo, afrouxando ligeiramente o rigor de algumas hipóteses. Tal modelo retornou previsões muito próximas dos valores reais nos dados de teste, mostrando que ele pode retornar boas estimativas para a receita mensal da empresa.

Tendo escolhido o melhor modelo, pode-se incorporar o processo de previsão como algo frequente na empresa. Ainda que sejam utilizados outros processos informais/manuais, a estimativa via modelagem estatística oferece um segundo ponto de vista com ótima previsão. Isso pode trazer maior confiança nas estimativas, possibilitando análise dos erros e variabilidade envolvidas no processo. Na rotina prática da empresa, tanto os métodos informais utilizados atualmente quando os resultados obtidos com ajustes de modelos estatísticos podem ser analisados em conjunto nas tomadas de decisões futuras.

Um tópico de interesse para estudos futuros é a incorporação de quebras estruturais aos modelos da classe SARIMA. Uma série temporal financeira pode passar por mudanças bruscas, por exemplo, os efeitos ocasionados pela pandemia do COVID-19, sendo relevante estar preparado para acomodar tal informação no modelo e informá-lo de possível modificação de nível, evitando previsões muito fora da realidade momentânea. Outro ponto de evolução constante é o acréscimo de novas observações na modelagem assim que elas se tornarem disponíveis na base de dados da empresa, visto que isso incorporará novos fenômenos à série e poderá mudar as estruturas de dependências autorregressivas e de médias móveis que foram observadas neste trabalho.

Referências

- AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, Institute of Electrical and Electronics Engineers (IEEE), v. 19, n. 6, p. 716–723, dec 1974. Citado na página 27.
- AUGUIE, B. *gridExtra: Miscellaneous Functions for "Grid"Graphics*. [S.l.], 2017. R package version 2.3. Disponível em: <<https://CRAN.R-project.org/package=gridExtra>>. Citado na página 29.
- BARTLETT, M. S. On the theoretical specification and sampling properties of autocorrelated time-series. *Supplement to the Journal of the Royal Statistical Society*, JSTOR, v. 8, n. 1, p. 27, 1946. Citado na página 19.
- BOARD, I. A. S. *IFRS 15 Revenue from Contracts with Customers*. [S.l.]: IFRS Foundation, 2014. Citado na página 14.
- BOX, G. E. P.; COX, D. R. An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, Wiley, v. 26, n. 2, p. 211–243, jul 1964. Citado na página 17.
- BOX, G. E. P. et al. *Time Series Analysis: Forecasting and Control*. Wiley series in probability and statistics. [S.l.]: John Wiley & Sons, 2015. ISBN 9781118674925. Citado 2 vezes nas páginas 16 e 20.
- CHEN, M.-C.; KAWAGUCHI, Y.; PATEL, K. An analysis of the trends and cyclical behaviours of house prices in the asian markets. *Journal of Property Investment & Finance*, Emerald, v. 22, n. 1, p. 55–75, feb 2004. Citado na página 15.
- CLAR, M.; DUQUE, J.-C.; MORENO, R. Forecasting business and consumer surveys indicators—a time-series models competition. *Applied Economics*, Informa UK Limited, v. 39, n. 20, p. 2565–2580, nov 2007. Citado na página 15.
- DICKEY, D. A.; FULLER, W. A. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, Informa UK Limited, v. 74, n. 366a, p. 427–431, jun 1979. Citado 2 vezes nas páginas 16 e 24.
- GUJARATI, D. N. *Basic Econometrics*. [S.l.]: McGraw Hill, 2009. ISBN 9780073375779. Citado 7 vezes nas páginas 16, 17, 18, 19, 20, 21 e 22.
- HOLT, C. C. Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, Elsevier BV, v. 20, n. 1, p. 5–10, jan 2004. Citado na página 26.
- HYNDMAN, R. J.; ATHANASOPOULOS, G. *Forecasting Principles and Practice: Principles and practice*. [S.l.]: OTexts, 2021. ISBN 9780987507136. Citado na página 17.
- HYNDMAN, R. J.; KHANDAKAR, Y. Automatic time series forecasting: the forecast package for r. *Journal of Statistical Software*, Foundation for Open Access Statistic, v. 27, n. 3, 2008. Citado na página 29.

KODDE, D. A.; SCHREUDER, H. FORECASTING CORPORATE REVENUE AND PROFIT: TIME-SERIES MODELS VERSUS MANAGEMENT AND ANALYSTS. *Journal of Business Finance & Accounting*, Wiley, v. 11, n. 3, p. 381–395, sep 1984. Citado na página 14.

KWIATKOWSKI, D. et al. Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, Elsevier BV, v. 54, n. 1-3, p. 159–178, oct 1992. Citado 2 vezes nas páginas 16 e 24.

LÜTKEPOHL, H.; XU, F. The role of the log transformation in forecasting economic variables. *Empirical Economics*, Springer Science and Business Media LLC, v. 42, n. 3, p. 619–638, dec 2010. Citado na página 16.

MORETTIN, C. M. C. T. P. A. *Análise de Séries Temporais Modelos Lineares Univariados: Modelos lineares univariados*. [S.l.]: Blucher, 2018. ISBN 9788521213512. Citado 6 vezes nas páginas 16, 18, 22, 23, 26 e 28.

O'HARA-WILD, M.; HYNDMAN, R.; WANG, E. *feasts: Feature Extraction and Statistics for Time Series*. [S.l.], 2022. R package version 0.3.0. Disponível em: <<https://CRAN.R-project.org/package=feasts>>. Citado na página 29.

STREIMIKIENE, D. et al. Forecasting tax revenues using time series techniques – a case of pakistan. *Economic Research-Ekonomska Istraživanja*, Informa UK Limited, v. 31, n. 1, p. 722–754, jan 2018. Citado na página 14.

WANG, E.; COOK, D.; HYNDMAN, R. J. A new tidy data structure to support exploration and modeling of temporal data. *Journal of Computational and Graphical Statistics*, Informa UK Limited, v. 29, n. 3, p. 466–478, jan 2020. Citado na página 29.

WANG, Q.; LI, S.; LI, R. Forecasting energy demand in china and india: Using single-linear, hybrid-linear, and non-linear time series forecast techniques. *Energy*, Elsevier BV, v. 161, p. 821–831, oct 2018. Citado na página 15.

WICKHAM, H. et al. Welcome to the tidyverse. *Journal of Open Source Software*, The Open Journal, v. 4, n. 43, p. 1686, nov 2019. Citado na página 29.

WICKHAM, H.; BRYAN, J. *readxl: Read Excel Files*. [S.l.], 2022. R package version 1.4.1. Disponível em: <<https://CRAN.R-project.org/package=readxl>>. Citado na página 29.

WINTERS, P. R. Forecasting sales by exponentially weighted moving averages. *Management Science*, Institute for Operations Research and the Management Sciences (INFORMS), v. 6, n. 3, p. 324–342, apr 1960. Citado na página 26.

WOOLDRIDGE, J. M. *Introductory econometrics a modern approach: a modern approach*. [S.l.]: South-Western Cengage Learning, 2012. 881 p. ISBN 9781111531041. Citado na página 28.

ZHENG, F.; XU, L. D.; TANG, B. Forecasting regional income inequality in china. *European Journal of Operational Research*, Elsevier BV, v. 124, n. 2, p. 243–254, jul 2000. Citado na página 15.

ZHU, H. *kableExtra: Construct Complex Table with 'kable' and Pipe Syntax*. [S.l.], 2021. R package version 1.3.4. Disponível em: <<https://CRAN.R-project.org/package=kableExtra>>. Citado na página 29.

ZINOURI, N.; TAAFFE, K. M.; NEYENS, D. M. Modelling and forecasting daily surgical case volume using time series analysis. *Health Systems*, Informa UK Limited, v. 7, n. 2, p. 111–119, jan 2018. Citado na página [15](#).