



Universidade Federal de Uberlândia
Faculdade de Matemática

Bacharelado em Estatística

**PREVISÃO DO PREÇO DE AÇÕES DE
EMPRESAS DO SETOR DE ENERGIA
ELÉTRICA DA [B]³ (BRASIL, BOLSA,
BALCÃO), POR MEIO DE MODELOS
AUTOREGRESSIVOS**

João Pedro Peres Pereira

Uberlândia-MG

2023

João Pedro Peres Pereira

**PREVISÃO DO PREÇO DE AÇÕES DE
EMPRESAS DO SETOR DE ENERGIA
ELÉTRICA DA [B]³ (BRASIL, BOLSA,
BALCÃO), POR MEIO DE MODELOS
AUTOREGRESSIVOS**

Trabalho de conclusão de curso apresentado à Co-
ordenação do Curso de Bacharelado em Estatística
como requisito parcial para obtenção do grau de
Bacharel em Estatística.

Orientador: Prof. Dr. Marcelo Tavares

**Uberlândia-MG
2023**



**Universidade Federal de Uberlândia
Faculdade de Matemática**

Coordenação do Curso de Bacharelado em Estatística

A banca examinadora, conforme abaixo assinado, certifica a adequação deste trabalho de conclusão de curso para obtenção do grau de Bacharel em Estatística.

Uberlândia, _____ de _____ de 20_____

BANCA EXAMINADORA

Prof. Dr. Marcelo Tavares

Proa. Dra. Nadia Giaretta Biase

Proa. Dra. Tatiane Bento da Costa

**Uberlândia-MG
2023**

AGRADECIMENTOS

Gostaria de expressar minha profunda gratidão a todas as pessoas que contribuíram de maneira significativa para a conclusão deste trabalho, tornando possível a realização deste sonho acadêmico.

Primeiramente, dedico este espaço de agradecimentos à minha amada esposa, Marina Gonçalves Alves. Sua constante compreensão, paciência e incentivo foram fundamentais ao longo de toda a minha jornada acadêmica e vida pessoal. Agradeço por ser meu pilar de apoio, fonte de inspiração e por compartilhar as alegrias e desafios desta caminhada.

À minha família, meus pais e irmão, José Roberto Pereira, Alessandra Peres e José Roberto Pereira Junior, expresso meu reconhecimento pela inabalável confiança, apoio emocional e incentivo ao longo dos anos. Seu apoio e incentivo foram fundamentais para o meu progresso acadêmico.

Quero estender meus agradecimentos especiais ao meu dedicado professor orientador, Marcelo Tavares. Sua orientação perspicaz, conhecimento especializado e incentivo constante foram cruciais para o desenvolvimento deste trabalho. Agradeço por compartilhar sua expertise, proporcionar direcionamento valioso e pela inspiração contínua que proporcionou ao longo de toda minha graduação.

A todos os amigos, colegas, e demais pessoas que de alguma forma contribuíram para este percurso, meu sincero agradecimento. Cada interação, conselho e momento compartilhado foram peças essenciais na construção desta conquista.

Este trabalho é fruto do apoio, amor e dedicação de cada pessoa mencionada. A todos, o meu mais profundo agradecimento.

RESUMO

O setor de energia elétrica no Brasil representa um dos pilares fundamentais da economia e do cotidiano nacional. Este estudo tem como objetivo analisar e realizar previsões utilizando modelos de séries temporais sobre ativos do setor de energia elétrica da Brasil, Bolsa, Balcão [B]³ para as 10 primeiras semanas de 2023. A seleção do modelo mais adequado aos dados ocorrerá mediante testes de hipóteses e comparações utilizando critérios de ajuste de modelos. A metodologia de construção dos modelos de séries temporais segue os fundamentos da teoria de Box e Jenkins. Após a definição dos modelos ARIMA (2,1,2) e ARIMA(3,1,4) para os ativos, optando pela versão logarítmica da série temporal original como a melhor abordagem. Como resultado, as evidências empíricas encontradas demonstraram que os melhores ajustes de modelos para realizar previsões eficientes e confiáveis para a série foram os modelos ARIMA.

Palavras-chave: Análise Preditiva, Séries Temporais, ARIMA, EDP Brasil, Engie Brasil.

ABSTRACT

The electric power sector in Brazil stands as one of the fundamental cornerstones of the national economy and daily life. This study aims to analyze and make predictions using time series models on assets within the electric power sector of the Brasil, Bolsa, Balcão [B]³ for the first 10 weeks of 2023. The selection of the most appropriate model for the data will be determined through hypothesis testing and comparisons using model fitting criteria. The methodology for constructing time series models follows the principles of the Box e Jenkins theory. After definition the of ARIMA models (2,1,2) and ARIMA (3,1,4) models for the assets, opting for the logarithmic version of the original time series as the best approach. As a result, the empirical evidence found demonstrated that the best model adjustments to make efficient and reliable forecasts for the series were the ARIMA models.

Keywords: Predictive Analytics, Time Series, ARIMA, EDP Brazil, Engie Brazil.

SUMÁRIO

Lista de Figuras	I
Lista de Tabelas	II
1 Introdução	1
2 Revisão de Literatura	3
2.1 Energia Elétrica	3
2.2 Bolsa de Valores - Brasil, Bolsa, Balcão [B] ³	4
2.3 Tecnologia	5
2.4 Comportamento das Ações no Tempo	5
2.5 Série Temporal	6
2.5.1 Processos Estocásticos	7
2.5.2 Modelo Clássico	7
2.5.3 Estacionariedade	8
2.5.4 Sazonalidade	8
2.5.5 Ruído Branco	9
2.6 Autocorrelograma	9
2.7 Modelo ARIMA	10
2.8 Significância dos Parâmetros	12
2.9 Seleção do Modelo	13
2.9.1 Critério de informação de AKAIKE	13
2.9.2 Erro Percentual Médio Absoluto	13
2.9.3 Erro Quadrático Médio de Previsão	14
2.10 Análise de Resíduos	14
2.10.1 Normalidade	14
2.10.2 Autocorrelação	15
2.10.3 Teste do Espectrograma	15
3 MATERIAIS E MÉTODOS	16
3.1 Procedimentos estatísticos realizados na análise	16
4 Resultados	18
4.1 Engie Brasil Energias S.A. (EGIE3)	18
4.2 Energias do Brasil S.A. (ENBR3)	25
5 Conclusões	33
Referências Bibliográficas	34

LISTA DE FIGURAS

4.1	Gráfico de linha da série histórica da ação EGIE3	18
4.2	Gráfico de linha da diferença da série EGIE3	19
4.3	Gráfico de linha de log da série histórica EGIE3	19
4.4	Gráfico de linha da diferença de log da série histórica EGIE3	19
4.5	Função de Autocorrelação e Função de Autocorrelação Parcial, para Diferença de EGIE3 e Diferença de Log	21
4.6	Gráfico de linha dos Resíduos do modelo ARIMA (2,1,2)	23
4.7	Teste do Espectrograma para o modelo ARIMA (2,1,2)	24
4.8	Previsão de fechamento de EGIE3 para as primeiras semanas de 2023	24
4.9	Fechamento EGIE3 realizado 2023 (em vermelho) versus previsão de fechamento EGIE3 (em azul)	25
4.10	Gráfico de linha da série histórica da ação ENBR3	26
4.11	Gráfico de linha da diferença da série ENBR3	26
4.12	Gráfico de linha de log da série histórica ENBR3	26
4.13	Gráfico de linha da diferença de log da série histórica ENBR3	27
4.14	Função de Autocorrelação e Função de Autocorrelação Parcial, para Diferença de ENBR3 e Diferença de Log	28
4.15	Gráfico de linha dos Resíduos do modelo ARIMA (3,1,4)	30
4.16	Teste do Espectrograma para o modelo ARIMA (3,1,4)	31
4.17	Previsão de fechamento de ENBR3 para as primeiras semanas de 2023	31
4.18	Fechamento ENBR3 realizado 2023 (linha vermelha) versus previsão de fechamento ENBR3 (linha azul)	32

LISTA DE TABELAS

4.1	Testes ADF para estacionariedade (EGIE3)	20
4.2	Teste t de Student para significância dos parâmetros nos modelos ajustados (EGIE3)	22
4.3	Critério de Seleção do Modelo (EGIE3)	23
4.4	Teste para normalidade dos resíduos do modelo ARIMA (2,1,2)	23
4.5	Fechamento EGIE3 realizado, Previsão do modelo, limites inferiores e superiores do modelo para as primeiras 10 semanas de 2023	25
4.6	Testes ADF para estacionariedade (EGIE3)	27
4.7	Teste t de Student para significância dos parâmetros nos modelos ajustados (ENBR3)	29
4.8	Critério de Seleção do Modelo (ENBR3)	30
4.9	Teste para normalidade dos resíduos do modelo ARIMA (3,1,4)	30
4.10	Fechamento ENBR3 realizado, Previsão do modelo, limites inferiores e superiores do modelo para as primeiras 10 semanas de 2023	32

1. INTRODUÇÃO

A exploração da análise de série temporal no contexto dos preços das ações das empresas do setor de energia elétrica na [B]³ [4] revela-se como uma área de estudo crucial e promissora. No âmbito do Plano Decenal de Expansão de Energia 2027, traçado pelo Ministério de Minas e Energia [15], destaca-se um investimento substancial de aproximadamente R\$ 1,8 trilhão, no período entre 2018 e 2027. Esse investimento tem o propósito de não somente suprir o aumento da demanda por energia, mas também de manter a natureza renovável da matriz energética.

Prevê-se que o consumo de energia elétrica cresça à uma taxa média de 2,3% anuais entre 2017 e 2027, onde ao longo da década, projeta-se redução gradual da intensidade energética, pela incidência de ganhos de eficiência energética bem como pela mudança na participação dos setores no consumo de energia [15].

No contexto do mercado financeiro, que é caracterizado por uma complexa interação entre diversos agentes orientados pela busca por resultados, órgãos reguladores responsáveis pelo controle do risco sistêmico e agentes econômicos que gerenciam a economia em larga escala, surgem desafios significativos, onde os riscos financeiros e operacionais se manifestam devido às operações que compõem o portfólio das instituições financeiras [40].

Portanto, é essencial obter uma compreensão precisa dos resultados financeiros, uma vez que esses resultados refletem o impacto de formação das carteiras e estratégias das instituições [40]. Essa análise é realizada por meio de diversas ferramentas financeiras utilizadas nas interações com clientes e mercados, sendo fundamental para monitorar e controlar o progresso em direção aos objetivos institucionais. A compreensão dos resultados financeiros são fonte de informação para tomada de decisão dos usuários das informações, entre estes destacam-se os investidores, que usam as informações financeiras para entenderem a ‘saúde’ da empresa e decidirem sobre investir em quais ações e o valor que estão dispostos a pagar “nos preços das ações”, além dos possíveis dividendos distribuídos conforme divulgação de resultados das empresas em determinados períodos.

O entendimento das diferentes carteiras e abordagens adotadas pelas instituições financeiras são fundamentais para uma compreensão abrangente do negócio e do contexto no qual ele opera. Portanto, é imperativo ter conhecimento da série temporal relacionada aos valores dos ativos e prever seus valores futuros.

Uma série temporal consiste em um conjunto de observações sequenciais ao longo do tempo, sendo fundamental observar que as observações próximas são interdependentes. Como muitos métodos estatísticos foram desenvolvidos para analisar dados independentes, a análise de séries

temporais também requer a aplicação de técnicas específicas [7].

A utilização de séries temporais, são encontradas em diversos campos, tais como, Economia (preços diários de ações, taxas mensais de desemprego), Medicina (níveis de eletrocardiograma, batimentos cardíacos por minuto), Epidemiologia (novos casos mensais de doenças, casos semanais de surtos), Meteorologia (precipitação pluviométrica, temperatura diária), entre outros.

A previsão dos dados é de suma importância para acionistas ou investidores do setor de energia elétrica, na trajetória de crescimento do ativo para identificar maiores ganhos ou possíveis perdas com determinado investimento, podendo auxiliar na diversificação da carteira do investidor e projetar uma margem de valores e variações da ação [21].

Ativos do setor de energia elétrica como Engie Brasil Energia S.A. [12] e Energias do Brasil S.A. [14] são negociados na [B]³ [4] e possuem um volume médio de 1 e 2 milhões de ações negociadas diariamente, respectivamente. Em função destes volumes de ações negociadas, o comportamento da série histórica e previsões de curto e longo prazo, devem ser estabelecidas por meio de modelos de séries temporais. A compreensão dessas dinâmicas é vital para a tomada de decisões informadas, permitindo que investidores estejam preparados para aproveitar oportunidades de mercado e mitigar riscos em um setor estratégico e dinâmico como o de energia elétrica.

Segundo Stevenson [39], um modelo de séries temporais analisa a existência de quatro elementos básicos na composição da série, os quais são: tendência, variações cíclicas, variações sazonais e as variações irregulares. Ao identificar os elementos básicos é possível construir uma série livre de tendência e sazonalidade [19].

Box e Jenkins [6] introduziram um método utilizado na análise de modelos de séries temporais, no qual ajusta modelos autorregressivos integrados de médias móveis, ARIMA (p, d, q) a um conjunto de dados. É possível que vários modelos sejam ajustados para uma mesma série. Prontamente, o recomendado é de que seja escolhido um modelo que forneça o menor erro de previsão. Uma característica importante da série é que ela deve ser estacionária, para que possa ser analisada pelo modelo ARIMA. Uma série estacionária é uma série de dados na qual as propriedades estatísticas, como média, variância e covariância, permanecem constantes ao longo do tempo. Em outras palavras, uma série estacionária não exibe tendências claras, variações sazonais ou mudanças sistemáticas em suas características estatísticas [35].

Dessa forma, o estudo busca antecipar os preços das ações de empresas do setor de energia elétrica da [B]³ [4], por meio de modelos autorregressivos, através da análise de séries temporais como uma ferramenta fundamental, procurando identificar padrões e características futuras de valores dos ativos, com o objetivo de fornecer informações preditivas essenciais para orientar investidores na tomada de decisões.

2. REVISÃO DE LITERATURA

Com o propósito de aprimorar a base teórica, a revisão de literatura desempenha um papel fundamental na análise de séries temporais nos preços de ações do setor de energia elétrica. Esta revisão busca explicitamente fundamentar os argumentos e resultados por meio de contribuições provenientes da literatura especializada em estatística, proporcionando uma sólida base metodológica para a análise de séries temporais. Neste contexto, esta revisão se propõe a abordar de forma abrangente os pilares teóricos abordados nesta pesquisa: empresas do setor de energia elétrica e sua relação com a bolsa de valores, as transformações do mercado financeiro em relação aos preços de ativos, o comportamento histórico das ações ao longo do tempo e o uso de análise de séries temporais como ferramenta essencial para a análise e previsão dos movimentos de mercado de capitais.

2.1 ENERGIA ELÉTRICA

As ações vinculadas ao planejamento energético têm como objetivo compreender o comportamento futuro da energia elétrica e buscar ferramentas úteis para auxiliar nas decisões estratégicas.

No estudo conduzido por Santos [34], a autora discute a importância de um planejamento preciso diante do impacto significativo que uma produção de energia elétrica acima da demanda ou abaixo dela pode ter. Em seu trabalho, foi realizado o modelamento da série mensal de demanda de energia elétrica nas regiões sul e sudeste do Brasil, abrangendo o período de janeiro de 2002 a julho de 2018. Vale ressaltar que a região sudeste, devido ao seu nível mais elevado de desenvolvimento industrial, precede a região sul.

Diversos modelos estatísticos foram empregados, como SARIMA, SARIMAX e ETS, incorporando variáveis exógenas como o PIB e o IBCR para compreender o comportamento da previsão da demanda industrial de energia elétrica nessas regiões. A avaliação da acurácia dos modelos baseou-se em medidas como o Erro Percentual Absoluto Médio (MAPE) e o Erro Percentual Absoluto (MAE). Importante mencionar que todos os modelos exibiram taxas de erro reduzidas. Ao comparar as métricas de acurácia para cada região, o modelo ETS foi escolhido devido ao seu desempenho superior nas séries analisadas.

O setor de energia elétrica no Brasil desempenha um papel vital na economia e na vida cotidiana do país. Conforme discutido por Goldemberg et al. [18], o Brasil tem uma das matrizes energéticas mais limpas do mundo, com uma significativa parcela de sua eletricidade

gerada a partir de fontes renováveis, como a hidroeletricidade. No entanto, o setor também enfrenta desafios, incluindo a necessidade de diversificação das fontes de energia para garantir a segurança energética e a expansão da infraestrutura para atender à crescente demanda. O governo brasileiro tem implementado políticas e regulamentações para promover o desenvolvimento sustentável do setor, incentivando investimentos em energias renováveis e promovendo a eficiência energética [26].

No mercado financeiro, o setor de energia elétrica atrai investidores devido à sua dimensão e ao potencial de retorno. Conforme abordado por Fonseca et al. [17], o mercado de energia elétrica brasileiro envolve uma série de oportunidades de investimento, desde a geração de energia por meio de fontes variadas até a infraestrutura de transmissão e distribuição. A privatização e a desregulamentação parcial do setor nas últimas décadas abriram caminho para a participação do setor privado, atraindo investidores nacionais e internacionais que veem o mercado de energia elétrica como uma classe de ativos com potencial de crescimento [17].

A relação entre o mercado financeiro e o setor de energia elétrica é evidenciada pelo investimento em ações de empresas de energia, emissões de títulos vinculados ao setor e negociações de contratos futuros de eletricidade. Portanto, para entender as dinâmicas do mercado de energia elétrica brasileiro, é essencial considerar não apenas as questões regulatórias e operacionais, mas também as implicações financeiras e as oportunidades de investimento [3].

2.2 BOLSA DE VALORES - BRASIL, BOLSA, BALCÃO [B]³

No Brasil, atualmente, existe uma única instituição financeira dedicada à negociação de ativos financeiros, como ações, títulos e outros ativos, denominada Brasil, Bolsa, Balcão [4]. Esta é a evolução da antiga BM&F Bovespa e é uma empresa de capital aberto cujas ações (B3SA3) são listadas no Novo Mercado. O Novo Mercado é uma lista de empresas de capital aberto cujas ações são negociadas na [B]³ [4], que apresentam melhores níveis de governança corporativa, com isso garantindo maior transparência para quem deseja investir nessas instituições. A [B]³ [4] tem sua sede na cidade de São Paulo e é uma das maiores bolsas de valores do mundo em termos de valor de mercado, ocupando uma posição de destaque internacional no setor de negociação de ativos financeiros [4].

De acordo com Bodie et al. [5] a bolsa de valores desempenha um papel importante no sistema financeiro, permitindo que empresas levantem capital e que investidores comprem e vendam ativos. Os preços dos ativos na bolsa são determinados pela oferta e demanda, refletindo a percepção dos investidores sobre o desempenho e o potencial das empresas, facilitando a captação de capital por parte das empresas e oferece um ambiente para que os investidores comprem e vendam ativos financeiros.

Uma das principais funções da bolsa de valores é proporcionar liquidez, ou seja, a facilidade com que os ativos podem ser comprados ou vendidos. Isso permite que investidores comprem ações de empresas e, ao mesmo tempo, oferece às empresas a oportunidade de obter financiamento por meio da emissão de ações.

As bolsas de valores também desempenham um papel vital na economia, pois contribuem para a alocação eficiente de recursos, fornecendo informações sobre o desempenho das empresas e incentivando a transparência e a prestação de contas.

A globalização do mercado financeiro aumenta a interdependência entre os mercados, criando um ambiente em que eventos locais e globais podem ter impactos significativos nos preços dos ativos. Para gerenciar esses riscos e buscar oportunidades de investimento, os investidores reconhecem a importância da diversificação de portfólio.

A globalização exerce impacto significativo no mercado financeiro, pois está interligado em escala global. Isso se deve ao fato de que eventos e desenvolvimentos em uma parte do mundo podem ter efeitos imediatos e amplos em diferentes mercados financeiros. Por exemplo, notícias econômicas, políticas ou eventos inesperados em uma região podem desencadear reações em cadeia que afetam os preços de ativos em diversas partes do globo [24].

2.3 TECNOLOGIA

A tecnologia desempenha um papel fundamental no mercado de capitais, impulsionando a eficiência operacional, facilitando transações instantâneas e promovendo a transparência, o que influencia diretamente a dinâmica e a liquidez do mercado [37]. Da mesma forma, no segmento de energia elétrica, a adoção de tecnologias inovadoras não apenas aprimora a gestão da infraestrutura e a eficiência na distribuição, mas também permite a integração de fontes renováveis, promovendo a sustentabilidade e a resiliência do setor [23].

Além disso, a tecnologia possibilitou o surgimento de novos instrumentos financeiros e estratégias de negociação, como algoritmos de alta frequência. Conforme destacado por O'Hara [30]; "a automação de negociações por meio de algoritmos permite uma execução mais rápida e precisa, resultando em mudanças significativas na dinâmica de preços". Isso pode levar a movimentos bruscos e volatilidade em curtos espaços de tempo, afetando a avaliação de ativos.

A tecnologia também influencia a análise de dados e a modelagem financeira, proporcionando *insights* mais profundos e precisos. Autores como Hull [21] observam que "a modelagem quantitativa e a análise de *big data* permitem a identificação de padrões ocultos e a previsão de movimentos de preços". Portanto, a tecnologia não apenas afeta a eficiência operacional, mas também desempenha um papel fundamental na interpretação e previsão de preços de ativos no mercado financeiro.

2.4 COMPORTAMENTO DAS AÇÕES NO TEMPO

De acordo com Bodie et al. [5], a volatilidade é uma característica marcante das ações, ou seja, afetando o comportamento das ações ao longo do tempo. Ao longo de décadas, as ações geralmente apresentam uma tendência de crescimento no longo prazo, mas esse crescimento é pontuado por flutuações significativas de curto prazo. Essas flutuações são influenciadas por uma variedade de fatores, incluindo notícias econômicas, políticas, eventos globais e mudanças

nas perspectivas da empresa. É importante notar que as ações individuais podem se comportar de maneira diferente, com algumas empresas exibindo maior volatilidade do que outras. A volatilidade é uma medida de risco que se refere-se a variação dos preços de ativos ao longo do tempo, indicando o grau de flutuação dos preços e a instabilidade do mercado.

É importante observar que, em horizontes temporais mais longos, como décadas, a tendência histórica tem sido de crescimento no mercado de ações. Essa perspectiva de crescimento é apoiada por autores como Siegel [36], que destacam o papel dos dividendos e do crescimento dos lucros das empresas como impulsionadores do retorno sobre o investimento em ações ao longo do tempo. No entanto, os investidores devem estar preparados para enfrentar períodos de volatilidade e possíveis correções de mercado ao longo de suas jornadas de investimento.

2.5 SÉRIE TEMPORAL

Séries temporais são conjuntos de observações registradas em intervalos de tempo regulares e são amplamente utilizadas na análise de dados financeiros, econômicos, climáticos e muitos outros campos.

Segundo Brookwell e Davis [7], uma série temporal é um conjunto de dados em que as observações são registradas em função do tempo, e sua estrutura temporal permite estudar tendências, padrões sazonais e ciclos subjacentes. Essas séries são úteis para fazer previsões, identificar padrões e entender como os dados evoluem ao longo do tempo, incluindo a aleatoriedade inerente aos processos estocásticos [7].

A análise de séries temporais envolve técnicas estatísticas e matemáticas para modelar e extrair informações úteis desses dados. É fundamental para a previsão de tendências econômicas, gerenciamento de riscos financeiros e tomada de decisões em diversos setores. Portanto, a compreensão das séries temporais e suas aplicações desempenha um papel fundamental em muitos campos de pesquisa e negócios.

O elemento distintivo desses dados é a dependência entre observações próximas, ao contrário de um modelo de regressão, no qual a ordem das observações não tem relevância para a análise. O objetivo principal é compreender e modelar essa dependência, a qual pode estar relacionada a variáveis além do tempo, como localização geográfica, profundidade, entre outras [13].

É possível classificar séries temporais como contínuas ou discretas. No primeiro caso, as observações ocorrem de forma ininterrupta ao longo do tempo, enquanto no segundo, as observações são feitas em momentos específicos. Observações contínuas podem ser discretizadas, sendo importante destacar que a classificação da série não está diretamente relacionada à classificação da variável observada como contínua ou discreta [13].

A aplicação de técnicas de séries temporais tem como finalidade compreender eventos passados e realizar previsões para o futuro, contribuindo para o processo de tomada de decisão por parte de responsáveis e autoridades [11]. Além da previsão de valores futuros, outros objetivos incluem a investigação do processo gerador da série, a análise do comportamento da série e a identificação de periodicidades relevantes [29].

2.5.1 PROCESSOS ESTOCÁSTICOS

Os processos estocásticos desempenham um papel fundamental na modelagem de séries temporais, pois capturam a aleatoriedade e a incerteza inerentes aos dados sequenciais ao longo do tempo. Segundo Brookwell e Davis [7], os processos estocásticos são modelos matemáticos que descrevem a evolução probabilística dos dados temporais, considerando a natureza aleatória das observações. Eles incluem propriedades como estacionariedade, dependência temporal e propriedades de distribuição, permitindo a compreensão de como os dados evoluem ao longo do tempo em um ambiente probabilístico. A utilização de processos estocásticos é essencial para capturar a dinâmica temporal dos dados, permitindo previsões e projeções futuras com base na aleatoriedade dos eventos.

A compreensão dos processos estocásticos é crucial na análise de séries temporais. Conforme destacado por Chatfield [8], os modelos estocásticos permitem representar e analisar a variabilidade e a incerteza dos dados ao longo do tempo. Isso fornece uma estrutura teórica para a modelagem e a interpretação das séries temporais, sendo essencial para a identificação de padrões e a geração de previsões precisas em diversas áreas, desde economia e finanças até climatologia e epidemiologia.

Žitković [41] apresenta uma definição fundamental de processos estocásticos e suas características. Ele explica que uma família de variáveis aleatórias $\{X_t\}_{t \in T}$, onde T é um subconjunto de $[0, \infty)$, é denominada processo estocástico ou aleatório. Quando $T = \mathbb{N}$ ($T = \mathbb{N}_0$), $\{X_t\}_{t \in T}$ é chamado de processo de tempo discreto, e quando $T = [0, \infty)$, é chamado de processo de tempo contínuo.

Žitković [41] também esclarece que quando T é único (ou seja, $T = \{1\}$), o processo $\{X_t\}_{t \in T}$ reduz-se a uma variável aleatória única, e quando T é finito (ou seja, $T = \{1, 2, \dots, n\}$), temos um vetor aleatório. Portanto, os processos estocásticos representam uma generalização dos vetores aleatórios. A interpretação desses processos difere, uma vez que os componentes de um vetor aleatório geralmente, embora não sempre, representam coordenadas espaciais, enquanto o índice $t \in T$ frequentemente é interpretado como tempo.

Žitković [41] ressalta que os processos estocásticos são frequentemente utilizados para modelar a evolução de um fenômeno aleatório ao longo do tempo. Quando T abrange o intervalo $[0, \infty)$ (processo de tempo contínuo), o valor do processo pode mudar de forma contínua a qualquer momento. Por outro lado, quando $T = \mathbb{N}$ (processo de tempo discreto), as mudanças ocorrem em instantes discretos no tempo.

2.5.2 MODELO CLÁSSICO

No estudo de séries temporais, geralmente há um interesse em analisar e modelar essas séries com o objetivo de descrever suas características mais relevantes, identificar possíveis relações com outras séries e realizar previsões com base em dados históricos. O número de períodos futuros para os quais se deseja fazer previsões é denominado horizonte de previsão. De acordo com Ehlers [13], as propriedades de uma série temporal X_t podem ser compreendidas por meio

da seguinte decomposição:

$$X_t = T_t + C_t + R_t \text{ ou } X_t = T_t \cdot C_t \cdot R_t \quad (2.1)$$

Nessa equação apresentada anteriormente, " t " representa o tempo, definido como $t = 1, 2, \dots, n$. T_t é uma componente de tendência; C_t é uma componente cíclica (ou sazonal) e R_t é uma componente aleatória ou ruído. A questão central do modelo clássico é determinar a forma da equação que relaciona essas componentes com a variável em questão, ou seja, se o modelo será aditivo (soma das componentes) ou multiplicativo (produto das componentes).

2.5.3 ESTACIONARIEDADE

Segundo Shuumway e Stoffer [35], estacionariedade em séries temporais refere-se à constância das propriedades estatísticas da série ao longo do tempo. Em termos mais simples, uma série temporal é estacionária quando suas características, como média, variância e autocorrelações, não mudam de forma sistemática com o tempo. Essa propriedade é importante porque muitas técnicas de análise de séries temporais, como previsões e modelagem, pressupõem que a série seja estacionária.

Quando uma série não é estacionária, pode ser desafiador aplicar métodos estatísticos tradicionais, pois as propriedades da série podem variar ao longo do tempo, tornando as previsões e análises menos confiáveis. Portanto, a detecção da estacionariedade é uma etapa determinante na análise de séries temporais.

O Teste de Dickey-Fuller (ADF) é comumente utilizado para verificar a estacionariedade de uma série temporal [19]. Ele se baseia na hipótese nula de que a série possui raiz unitária, o que implica que a série não é estacionária, enquanto a hipótese alternativa sugere estacionariedade.

O teste ADF avalia estatisticamente se a série temporal pode ser transformada em uma série estacionária aplicando diferenciação (ou outras transformações). A estatística do teste é baseada na regressão de uma série temporal X_t sobre seus próprios valores defasados (ou seja, X_{t-1} , X_{t-2} , etc.). A ideia é verificar se a série após a aplicação de diferenças (ou seja, tornando-a estacionária) pode ser explicada de forma significativa por seus próprios valores anteriores.

$$ADF = \frac{\text{coeficiente estimado de } X_{t-1}}{\text{Erro Padro do coeficiente}} \quad (2.2)$$

Essa estatística é comparada com valores críticos para determinar se a série é estacionária ou não. Se a estatística de teste for menor do que o valor crítico correspondente ao nível de significância escolhido, rejeita-se a hipótese nula de raiz unitária, indicando estacionariedade. Caso contrário, não é possível rejeitar a hipótese nula, sugerindo que a série não é estacionária.

2.5.4 SAZONALIDADE

Os ciclos se manifestam por meio de oscilações ascendentes e descendentes nas séries, exibindo padrões suaves e repetitivos ao longo da tendência principal, resultando em flutuações

periódicas nos valores da variável. A sazonalidade, presente em uma série, refere-se a essas oscilações previsíveis que ocorrem regularmente em determinados intervalos, como em um período anual, mensal, semanal ou diário. A presença da sazonalidade pode ser identificada por meio da análise espectral do correlograma. É importante notar que a diferença fundamental entre os componentes sazonais e cíclicos reside na previsibilidade dos movimentos; a sazonalidade segue padrões regulares, ao passo que os movimentos cíclicos tendem a ser irregulares [20].

2.5.5 RUÍDO BRANCO

Variações irregulares se referem a flutuações imprevisíveis e não explicadas em uma série temporal. Elas são o resultado de eventos imprevistos ou fatores que não podem ser planejados. A decomposição de uma série temporal ajuda a identificar quais componentes estão contribuindo para as variações na série, permitindo a criação de índices e equações para prever futuros períodos da série.

Quando um modelo é apropriado, espera-se que os resíduos, que são as diferenças entre os valores observados e os valores previstos pelo modelo, não exibam nenhum padrão discernível. Isso é importante para que possam ser considerados como ruído branco. De acordo com Morettin e Tolo [28], uma sequência de resíduos R_t é caracterizada como ruído branco quando cada valor na série de erros de estimação segue uma distribuição normal, tem média zero, variância constante e não apresenta correlação entre os valores.

Em resumo, o ruído branco é uma condição desejada nos resíduos de um modelo, pois indica que as flutuações restantes na série não são previsíveis e não possuem estrutura sistemática, tornando o modelo apropriado para análise e previsão.

2.6 AUTOCORRELOGRAMA

Uma ferramenta essencial para analisar as características de uma série temporal envolve uma série de medidas chamadas coeficientes de autocorrelação amostral. Isso é semelhante ao conceito do coeficiente de correlação comum que usamos para medir a relação entre pares de observações de variáveis, como x e y . A função de autocorrelação é uma função que associa a cada valor um intervalo de tempo e sua representação gráfica é chamada de autocorrelograma ou simplesmente correlograma. Geralmente, espera-se que essa função diminua à medida que aumentamos o intervalo de tempo entre as observações, o que significa que, com o passar do tempo, as observações tendem a se tornar menos relacionadas. Em resumo, a função de autocorrelação é uma medida que nos ajuda a entender o grau de dependência entre as diferentes observações em uma série temporal. Valores próximos no tempo tendem a ter maior semelhança entre si do que valores que estão muito afastados no tempo. A função de autocorrelação será dada por:

$$\rho_k = \frac{X_k}{X_0} = \frac{\sum_{t=1}^{n-k} (x_t - \bar{X})(x_{t+k} - \bar{X})}{\sum_{t=1}^n (x_t - \bar{X})^2} \quad (2.3)$$

Na equação 2.3, x_t representa o valor da série temporal no tempo t , \bar{X} é a média da série temporal, n é o número total de observações na série e k é o atraso para o qual será calculado a autocorrelação. A fórmula mede a covariância normalizada entre a série temporal no tempo t e a série temporal no tempo $t+k$, dividida pela variância da série temporal.

Conforme Franco [9], o conceito de autocorrelação pode ser ampliado. Quando avaliamos a correlação entre duas observações em série, X_t e x_{t+k} , enquanto eliminamos a influência das observações intermediárias, como $x_{t+1}, x_{t+2}, \dots, x_{t+k-1}$, estamos lidando com o que é conhecido como autocorrelação parcial. Essa medida é representada por:

$$\begin{bmatrix} 1 & \rho_1 & \dots & \rho_{k-1} \\ \rho_1 & 1 & \dots & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k_1} \\ \phi_{k_2} \\ \vdots \\ \phi_{k_k} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix} \quad (2.4)$$

A autocorrelação parcial de um processo é representada pela sequência dos valores de ϕ_{k_k} , que são obtidos através da solução das equações mencionadas anteriormente 2.4. De acordo com o teorema de Cramer, podemos estabelecer o seguinte resultado:

$$\phi_{k_1} = \rho_1; \phi_{k_2} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}}; \phi_{k_3} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} \dots; \phi_{k_k} = \frac{|\rho_k^*|}{|\rho_k|} \quad (2.5)$$

Onde $|\cdot|$ é o determinante da matriz, P_k é a matriz de autocorrelação e P_k^* é a matriz P_k com a última coluna substituída pelo vetor de autocorrelações. Para determinar a quantidade de parâmetros de Média Móveis, é realizada uma análise do autocorrelograma. Os possíveis valores de "q" correspondem aos "lags" que se encontram acima do intervalo de confiança no gráfico de autocorrelação, representam o número de termos de Médias Móveis anteriores que serão incluídos no modelo. Da mesma forma, para os parâmetros Auto-Regressivos, a análise é feita considerando a Função de Autocorrelação Parcial (FACP) e os possíveis valores de "p" correspondem aos "lags" que se encontram acima do intervalo de confiança na FACP, e que serão incluídos no modelo.

2.7 MODELO ARIMA

A utilização de modelos de séries temporais, como modelos autoregressivos (AR), modelos de médias móveis (MA), ou modelos mais avançados como ARIMA (*Autoregressive Integrated Moving Average*) e modelos de suavização exponencial, permitem capturar e interpretar a estrutura temporal dos dados, levando em consideração padrões anteriores para fornecer previsões precisas e *insights* valiosos [22]. A escolha do modelo adequado depende da natureza

dos dados e do comportamento observado ao longo do tempo, permitindo a interpretação e projeção precisa de séries temporais em diversas aplicações.

Conforme Gutierrez [20], os modelos ARIMA são ferramentas estatísticas lineares utilizadas na análise de séries temporais. A sigla, em inglês, corresponde a "*Auto-Regressive Integrated Moving Average model*", ou seja, um modelo que combina componentes auto-regressivos, que representam as defasagens da série transformada (geralmente tornada estacionária por meio de diferenciação), com componentes de médias móveis, que refletem as defasagens dos erros aleatórios. O termo "integrado" descreve o processo de diferenciação aplicado à série original, com o objetivo de torná-la estacionária, ou seja, eliminando tendências e padrões não desejados.

De acordo com Morettin e Toloí [28] explicam, a modelagem ARIMA requer que a série temporal em análise seja estacionária. Isso implica que a série deve apresentar uma distribuição aleatória em torno de uma média constante, demonstrando uma certa estabilidade ao longo do tempo. Ao construir o correlograma, se os coeficientes de autocorrelação diminuírem rapidamente em direção a zero, isso indica que a série é estacionária. No entanto, caso os dados originais não atendam a esse critério, é necessário aplicar uma transformação. A transformação mais comum envolve a aplicação de diferenças sucessivas à série original até que se obtenha uma série estacionária.

$$\Delta Z(t) = Z(t) - Z(t - 1) \quad (2.6)$$

A estimativa da variável S em um dado momento t é calculada como uma combinação linear de p termos da série observada, que incluem o termo aleatório $\varepsilon(t)$, que representa o ruído branco. Os coeficientes φ_i são parâmetros que atribuem pesos aos valores da série S desde o instante imediatamente anterior ($t - 1$) até o mais distante ($t - p$). Esses coeficientes são determinados por meio de técnicas que visam minimizar o erro na estimativa.

De acordo com Pindyck & Rubinfeld [31] o modelo Auto-Regressivo $AR(p)$ é definido por:

$$S(t) = \sum_{i=1}^p \varphi_i S(t - i) + \varepsilon(t) \quad (2.7)$$

O modelo $AR(p)$, ou modelo autorregressivo de ordem p , apresenta algumas limitações significativas. Primeiramente, ele parte da suposição de que existe uma relação linear entre os elementos da sequência temporal. Em outras palavras, o modelo presume que as observações estão interligadas de maneira linear.

Além disso, o modelo $AR(p)$ se baseia na hipótese de que a série é estacionária. Isso significa que ele assume que a média e o desvio padrão das observações permanecem constantes ao longo do tempo. Em termos mais simples, ele considera que não há mudanças significativas na média ou na variabilidade das observações com o passar do tempo.

O modelo de médias móveis, conhecido como $MA(q)$, opera sob a suposição de que a série que está sendo modelada é gerada por meio de uma combinação linear de q termos anteriores, juntamente com o termo aleatório $\varepsilon(t)$, que representa o ruído branco. Pindyck e Rubinfeld [31] afirma que "no processo de médias móveis de ordem q , cada observação de X_t é gerada por

uma média ponderada de perturbações aleatórias que vão para trás q períodos”. Makridakis [25] corrobora essa ideia, destacando que a relação de dependência entre os erros e a equação é referida como $MA(q)$. Em resumo, o termo MA é a média móvel dos erros ε . O processo $MA(q)$ é representado por:

$$S(t) = - \sum_{i=1}^q \theta_i \varepsilon(t-1) + \varepsilon(t) \quad (2.8)$$

Segundo Makridakis [25], os parâmetros $\theta_1, \theta_2, \dots, \theta_i$ são determinantes das ponderações da média móvel, podendo ser positivos ou negativos. Quando combinamos os modelos $AR(p)$ e $MA(q)$, obtemos o modelo $ARMA(p, q)$. As mesmas limitações mencionadas anteriormente para o modelo $AR(p)$, relacionadas à linearidade e estacionariedade do fenômeno modelado, também se aplicam aos modelos $MA(q)$ e $ARMA(p, q)$.

Além dos modelos $AR(p)$ e $MA(q)$, apresentamos o modelo $ARIMA$, que é representado como $ARIMA(p, d, q)$, onde p representa a ordem dos termos Auto-Regressivos, d indica o número de diferenciações necessárias para tornar a série temporal estacionária e q corresponde a ordem dos termos de Médias Móveis.

A adição do componente de diferenciação, representado por d , é uma característica distintiva do $ARIMA$ em relação ao $ARMA$. Essa etapa de diferenciação é aplicada para remover a tendência da série temporal, tornando-a estacionária.

O processo de aplicação dos modelos de Box & Jenkins para a previsão de séries temporais envolve três etapas principais: identificação, estimação e verificação. Após a conclusão dessas etapas, se o modelo for considerado satisfatório, procedemos com a estimação dos parâmetros. Neste estudo, utilizaremos o método de mínimos quadrados para estimar o vetor autorregressivo e o vetor de parâmetros de médias móveis. Esses parâmetros são escolhidos de forma a minimizar a soma dos quadrados das diferenças entre os valores observados na amostra e as estimativas obtidas com esses parâmetros estimados [38].

Essa etapa de estimação dos parâmetros foi realizada com o auxílio de um software específico para análise de séries temporais. Vale ressaltar que existem diferentes métodos para a estimação dos parâmetros do modelo, e todos eles tendem a produzir resultados semelhantes, embora possam variar em termos de eficiência computacional [28].

2.8 SIGNIFICÂNCIA DOS PARÂMETROS

Para determinar se um parâmetro tem influência significativa em um modelo, realizamos um teste de hipótese para cada parâmetro estimado. Esse teste pressupõe que os erros sejam independentes e sigam uma distribuição idêntica. Para realizar o teste, utilizamos o teste t-Student, com $n - 1 - d$ graus de liberdade, seja $\hat{\theta}_i$ um parâmetro estimado de Médias Móveis, a Estatística de teste é calculada da seguinte maneira:

$$T_{\hat{\theta}_i} = \frac{\hat{\theta}_i - 0}{\sqrt{\sigma_{\hat{\theta}_i}^2}} \quad (2.9)$$

O mesmo vale para $\hat{\varphi}_i$, os parâmetros estimados Auto-Regressivos. Logo, rejeitamos H_0 a um nível de confiança de $(1 - \alpha)$ se $|T_0| > t_{(1-\frac{\alpha}{2}; n-1-d)}$ [2].

Após a identificação do processo estocástico que gera a série temporal em questão e a estimação dos parâmetros do modelo, avançamos para a etapa de previsão de valores futuros. Neste estudo, utilizaremos o software R [10] para realizar os processos de identificação e previsão dos valores.

2.9 SELEÇÃO DO MODELO

Neste estudo, utilizamos três critérios de seleção de modelo: o Critério de Akaike, o Erro Percentual Absoluto Médio (MAPE) e o Erro Quadrático Médio de Previsão (EQMp). É importante destacar que essas medidas são comparativas e não possuem significado por si só quando avaliadas individualmente para um único modelo. Portanto, a seleção do melhor modelo é determinada considerando o modelo que apresentar os valores mais baixos em todos os três critérios.

No caso deste estudo, os três métodos de seleção convergiram, o que significa que o modelo escolhido foi o mesmo de acordo com o Critério de Akaike, o MAPE e o EQMP. No entanto, em situações em que esses critérios podem divergir, a escolha será baseada no modelo que obtiver o menor valor no Erro Quadrático Médio de Previsão.

2.9.1 CRITÉRIO DE INFORMAÇÃO DE AKAIKE

O Critério de Informação de Akaike (AIC) é definido como

$$AIC_p = -2\text{Log}(L_p) + 2[(p + 1) + 1] \quad (2.10)$$

Em que L_p é a função de máximo verossimilhança do modelo e p é o número de variáveis explicativas consideradas no modelo [20]. O critério de informação de Akaike (AIC) é uma métrica que pode assumir qualquer valor, inclusive valores negativos, uma vez que sua magnitude depende da forma da função de verossimilhança do modelo. Para efetuar comparações significativas de modelos usando o AIC, é fundamental que os modelos ajustados sejam estimados com base no mesmo período amostral. Isso garante que as comparações sejam feitas em condições equitativas e que o AIC seja uma métrica válida para determinar qual modelo é mais apropriado.

2.9.2 ERRO PERCENTUAL MÉDIO ABSOLUTO

O Erro Percentual Absoluto Médio (MAPE) é uma métrica que expressa a precisão de uma previsão como uma porcentagem do erro. Essa abordagem pode ser mais fácil de compreender

do que outras estatísticas de erro. A equação que calcula o MAPE é a seguinte:

$$\sum \frac{|(x_i - \hat{x}_i)|}{x_i}, (x_i \neq 0) \quad (2.11)$$

Nessa fórmula, x_i representa o valor real da observação, \hat{x}_i é o valor previsto do modelo e n é o número total de observações[27].

2.9.3 ERRO QUADRÁTICO MÉDIO DE PREVISÃO

O Erro Quadrático Médio de Previsão (EQM_p) é definido por:

$$\frac{\sum_{i=1}^s (x_{t-s+i} - \hat{x}_{t-s+i})^2}{s} \quad (2.12)$$

Onde s representa o horizonte de previsão, x_{t-s+i} denota o valor real das s últimas observações, \hat{x}_{t-s+i} refere-se ao valor previsto em um modelo construído a partir dos dados, excluindo as s últimas observações, e n é igual ao número total de observações [28].

A finalidade dessa comparação é avaliar a precisão do modelo, comparando os s valores retirados com os valores previstos pelo modelo que foi treinado sem considerar essas s últimas observações. Isso permite verificar o quão bem o modelo é capaz de fazer previsões para um horizonte de previsão de s períodos à frente.

2.10 ANÁLISE DE RESÍDUOS

Conforme discutido por Morettin & Toloí [28], ao ajustar um modelo ARIMA, é fundamental que os resíduos desse modelo bem ajustado se comportem como uma sequência de ruídos brancos. Isso implica que esses resíduos devem atender a várias condições, incluindo: distribuição normal, média zero, variação constante ao longo do tempo e não autocorrelacionados. Para verificar se essas suposições são atendidas, realiza-se análises gráficas e testes de hipóteses.

2.10.1 NORMALIDADE

De acordo com Abdi & Molin [1], o teste de Lilliefors é uma modificação do teste de Kolmogorov-Smirnov, utilizado para verificar se um conjunto de dados segue uma distribuição normal. A distinção fundamental entre esses testes reside no fato de que o teste de Kolmogorov-Smirnov avalia a normalidade com base na média e no desvio padrão fornecidos, enquanto o teste de Lilliefors utiliza a média e o desvio padrão calculados diretamente a partir do conjunto de dados em análise. O objetivo principal do teste de Lilliefors é testar a hipótese de normalidade dos dados, considerando as estatísticas de média e desvio-padrão obtidas a partir da própria amostra de dados.

2.10.2 AUTOCORRELAÇÃO

Um teste usado para verificar se os resíduos são independentes e identicamente distribuídos, ou seja, se não apresentam autocorrelação, é o teste de Box e Pierce, conforme explicado por Priestley [32]. Esse teste é baseado nas primeiras k autocorrelações, \hat{r}_k , dos resíduos. No contexto de um modelo ARIMA (p, d, q) , se o modelo ajustado é apropriado, a estatística do teste segue uma distribuição aproximadamente qui-quadrado (X^2). A hipótese testada é que os resíduos são independentes e identicamente distribuídos (i.i.d.). A hipótese nula (H_0) é rejeitada se o valor da estatística Q for maior que o valor crítico da distribuição X^2 com $(k - p - q)$ graus de liberdade. Nesse contexto, k representa o número de “lags”, p é a ordem da parte auto-regressiva do modelo e q é a ordem da parte de médias móveis do modelo.

Para uma avaliação informal da independência e identidade das distribuições dos resíduos, como sugerido por Morettin & Toloi [28], você pode construir um correlograma à um intervalo de confiança $(1 - \alpha)$, e $(1 - \alpha)\%$ das correlações deve estar dentro deste intervalo de confiança. Isso pode ser usado como um indicativo de independência dos resíduos.

2.10.3 TESTE DO ESPECTROGRAMA

O espectrograma foi introduzido nos estágios iniciais da análise espectral de séries temporais com o propósito de identificar regularidades cíclicas nos dados. Esse teste envolve a avaliação do espectro acumulado por meio de um estimador imparcial, ou seja, um método estatístico que não possui viés na estimativa do espectro, representado por:

$$I(f) = \frac{1}{n} \left| \sum_{t=1}^n X_t e^{-2\pi i f t} \right|^2 \quad (2.13)$$

Onde $I(f)$ é o valor do periodograma para uma determinada frequência f , n é o tamanho da amostra da série temporal, X_t são os valores da série temporal e f é a frequência em que o periodograma está sendo calculado. Para realizar o Teste do Espectrograma, geralmente se compara o periodograma da série com intervalos de confiança, obtidos através da aplicação do teste Smirnov-Kolmogorov, para determinar se existem componentes de frequência significativas que podem indicar sazonalidade ou padrões importantes na série temporal [38].

3. MATERIAIS E MÉTODOS

Os dados coletados neste trabalho são dos preços das ações de dois ativos do setor de energia elétrica na B3 [4], observados semanalmente entre janeiro de 2015 e dezembro de 2022 os valores em reais. Os dados são relacionados as cotações dos preços de fechamento das ações obtidos através da plataforma de negociação ProfitChart [33].

Os ativos observados foram da empresa Engie Brasil Energias S.A. (EGIE3) [12] negociada como ação ordinária e atuando na implantação e operação de usinas geradoras de eletricidade, sendo também agente ativo na atividade de comercialização e Energias do Brasil S.A. (ENBR3) [14] negociada como ação ordinária e atuando na geração, distribuição e comercialização de energia elétrica para consumo residencial, industrial, rural e comercial.

Para realização da análise de séries temporais torna-se fundamental possuir conhecimentos em processos estocásticos. Isso se deve ao fato de que um processo estocástico representa um conjunto de variáveis aleatórias que, em sua maioria, são empregadas para examinar a evolução de fenômenos observados ao longo do tempo. Em contraste com a abordagem determinística, na qual, com uma condição inicial dada, se conhece todo o curso do sistema, os processos estocásticos são utilizados. Estes processos, quando uma condição inicial é estabelecida, ainda apresentam diversas trajetórias possíveis para a evolução do sistema [16].

Desta forma, serão abordadas posteriormente, as conceituações de processos estocásticos e series temporais com suas características, além de teste associados a estas características e parâmetros para seleção e adequação de modelos escolhidos.

3.1 PROCEDIMENTOS ESTATÍSTICOS REALIZADOS NA ANÁLISE

- a. Coleta de dados: Realização da coleta de dados referentes aos preços das ações das empresas do setor de energia elétrica.
- b. Diferenciação da série: Aplicação da diferenciação para remover a tendência e reduzir a variabilidade da série, tornando-a estacionária.
- c. Verificação de estacionariedade: Utilização do teste de Dickey-Fuller para verificar a estacionariedade da série temporal.
- d. Análise de autocorrelação (FAC e FACP): Realização da Função de Autocorrelação (FAC)

- e Função de Autocorrelação Parcial (FACP) para identificar ordens de médias móveis e autorregressivas nos modelos.
- e. Teste de significância dos parâmetros: Teste t-Student para avaliar a significância dos estimadores dos parâmetros, com uma significância de 5%. Modelos cuja hipótese nula foi rejeitada foram selecionados.
- f. Seleção do modelo: Utilização dos testes de critério de Akaike, EQMp e MAPE para a seleção dos modelos, com preferência pelos que apresentaram menores resultados nos três critérios.
- g. Teste de normalidade e homoscedasticidade: Aplicação do teste de Lilliefors para avaliar a normalidade dos resíduos e geração de um gráfico de resíduos para verificar a homoscedasticidade da variação temporal.
- h. Análise do espectrograma: Teste do modelo por meio da análise do espectrograma, permitindo uma visualização da distribuição da energia nos diferentes componentes de frequência.
- i. Verificação de ruído branco: Verificação final para assegurar se o modelo resultante é um ruído branco.
- j. Geração de previsões: Produção das previsões dos modelos e comparação com os resultados reais das primeiras 10 semanas de 2023.

4. RESULTADOS

A primeira fase compreendeu a coleta de dados das empresas, no fechamento semanal durante o período de janeiro 2015 a dezembro 2022.

4.1 ENGIE BRASIL ENERGIAS S.A. (EGIE3)

Na Figura 4.1 é apresentado o gráfico da série histórica da ação EGIE3 [12], no qual podemos notar que os dados não possuem sazonalidade e nem ciclicidade, e devido a série apresentar tendência, pode-se dizer que é uma série não estacionária. Segundo MORETTIN & TOLOI [28] para retirarmos a tendência faz-se a diferença da série, apresentada na Figura 4.2, onde foi realizada a primeira diferença e nota-se média constante com baixa variabilidade. Foi realizada uma transformação logarítmica nos dados com intuito de reduzir a variabilidade, com resultados da série apresentados na Figura 4.3, e também a mesma transformação foi realizada para a diferença, apresentados na Figura 4.4, verificando-se que a variabilidade se tornou mais homogênea em torno da média zero, com indícios de estacionariedade, demonstrando assim não necessidade de realizar outra diferença.

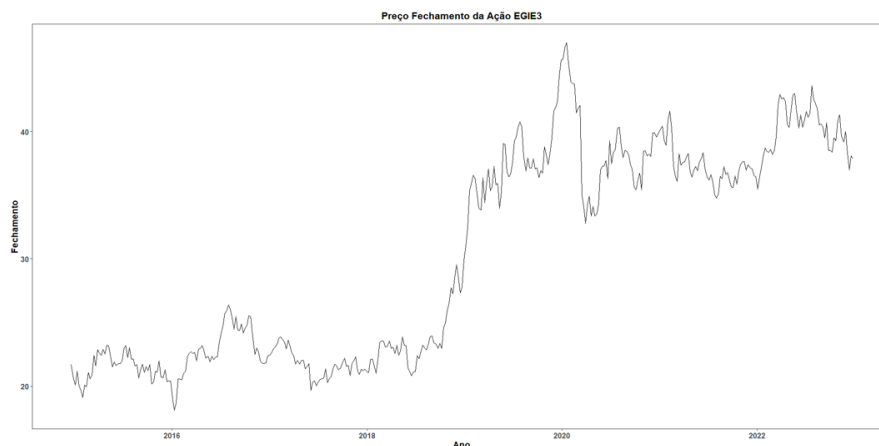


Figura 4.1: Gráfico de linha da série histórica da ação EGIE3
Fonte: Elaborado pelo autor.

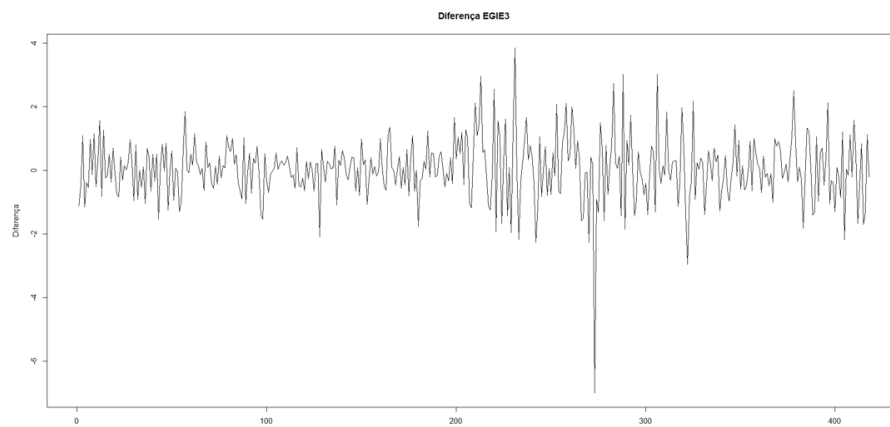


Figura 4.2: Gráfico de linha da diferença da série EGIE3
Fonte: Elaborado pelo autor.

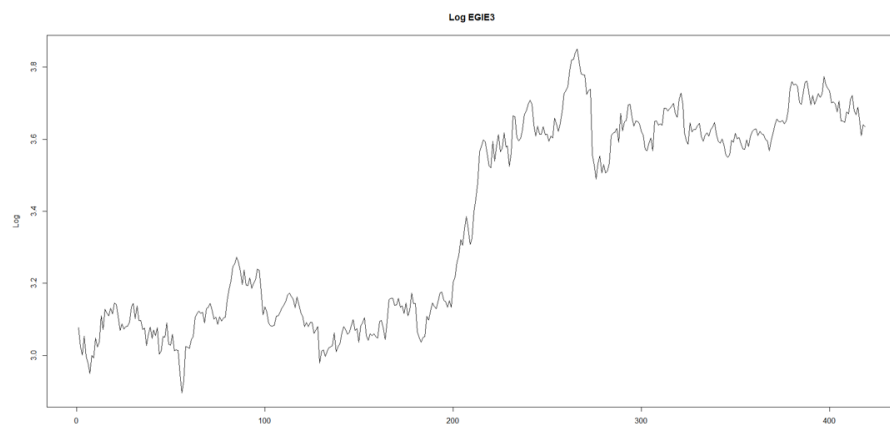


Figura 4.3: Gráfico de linha de log da série histórica EGIE3
Fonte: Elaborado pelo autor.

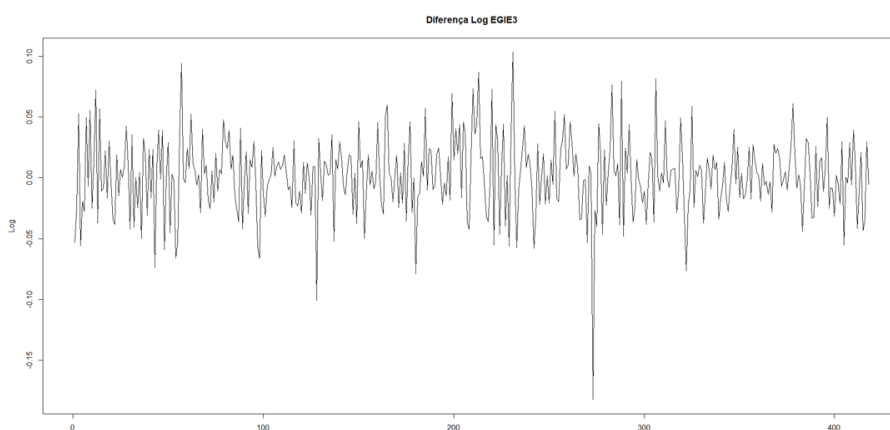


Figura 4.4: Gráfico de linha da diferença de log da série histórica EGIE3
Fonte: Elaborado pelo autor.

Na Tabela 4.1 o teste ADF (Augmented Dickey-Fuller), foi realizado para testar a hipótese de que a série possui raiz unitária, ou seja, não é estacionária. Essa hipótese não foi rejeitada para as séries de EGIE3 e Log, pois apresentaram valores-p maiores do que 0,05. Porém para

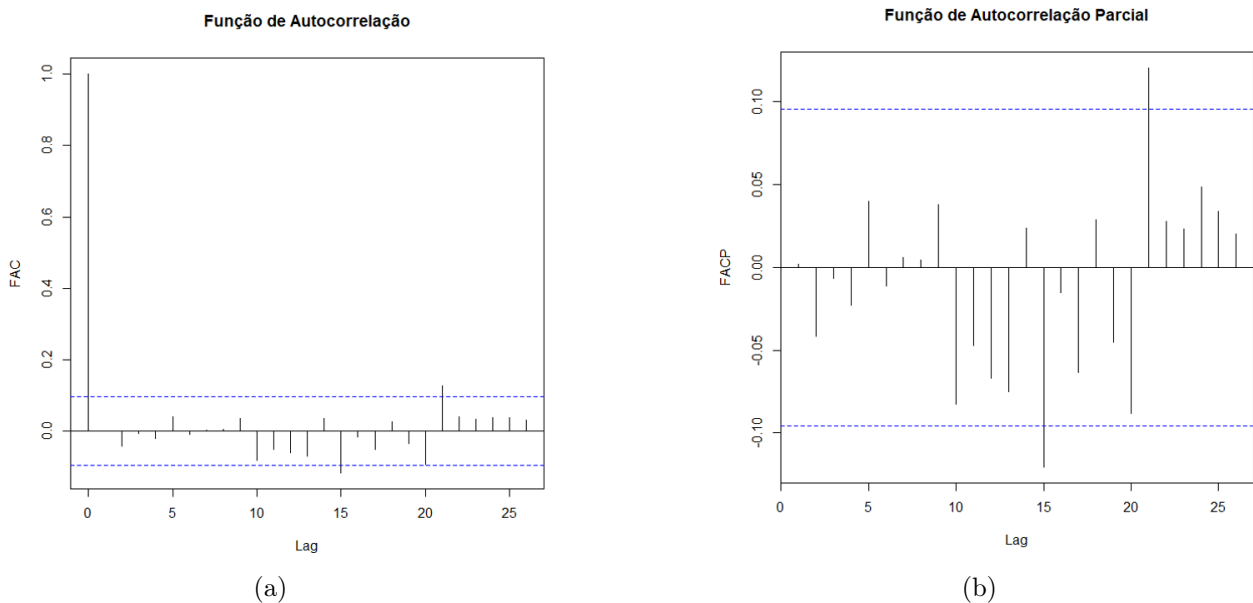
a diferença da série e diferença de Log, podemos ver que a hipótese foi rejeitada, ou seja, a diferença da série e a diferença de Log são estacionárias.

Tabela 4.1: Testes ADF para estacionariedade (EGIE3)

Série Temporal	Estatística	Valor-p
EGIE3	-2,3782	0,4177
Diferença de EGIE	-7,0258	0,0100
Log	-2,1229	0,5255
Diferença de Log	-7,5609	0,0100

Fonte: Elaborado pelo autor.

Na análise das autocorrelações, verifica-se na Figura 4.5(a) que o gráfico de autocorrelação para a diferença de EGIE3 os “lags” 15 e 21 ultrapassam os limites, isso indica que os “lags” possuem correlações significativas com as observações atuais, o que apontam para a possibilidade de considerar ordens específicas para a parte de Médias Móveis, e também na Figura 4.5(b) os “lags” 15 e 21, o que apontam para a possibilidade de considerar ordens específicas da parte Auto-Regressiva. Para a diferença do logaritmo de EGIE3 na figura 4.5(c) não nos dá pistas da ordem de Médias Móveis, e na figura 4.5(d) o mesmo para a ordem de Auto-Regressiva. A partir disso, através do software R [10] foi realizado ajustes dos modelos com até dois parâmetros de Médias Móveis e dois Auto-Regressivos.



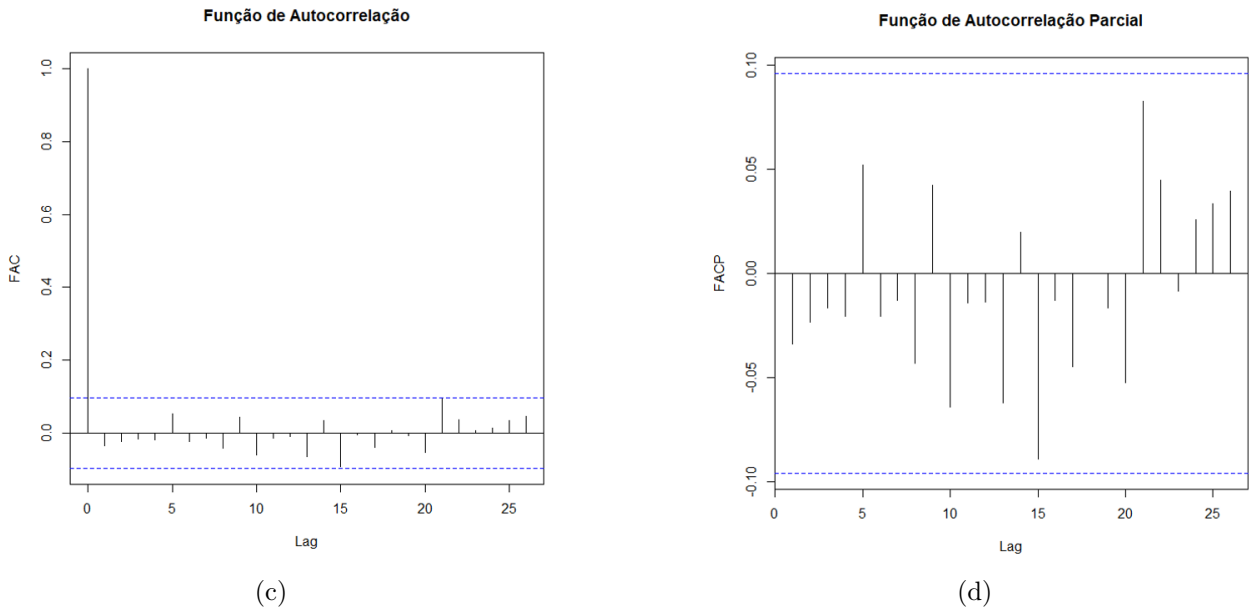


Figura 4.5: Função de Autocorrelação e Função de Autocorrelação Parcial, para Diferença de EGIE3 e Diferença de Log

Fonte: Elaborado pelo autor.

Considerando modelos com dois parâmetros de Médias Móveis e dois Auto-Regressivos, testou-se a significância dos estimadores dos parâmetros por meio do teste t de Student, considerando um nível nominal de significância de 5%, considerando os dados da diferença da série e diferença de log da série.

Os modelos que tiveram a hipótese nula rejeitada, ou seja, que possuem todos os parâmetros significativos foram os modelos da primeira diferença da série EGIE3 ARIMA: (1,1,1) e (2,1,2); e da série com diferença de log ARIMA: (2,1,2).

Tabela 4.2: Teste t de Student para significância dos parâmetros nos modelos ajustados (EGIE3)

ARIMA	Parâmetro	Diferença em EGIE3			Diferença de Log		
		Estatística	Valor-p	Desv. Pad.	Estatística	Valor-p	Desv. Pad.
(0,1,1)	θ_1	0,0039	0,3975	0,0510	-0,0334	0,3191	0,0501
(0,1,2)	θ_1	0,0035	0,3976	0,0489	-0,0333	0,3164	0,0490
	θ_2	-0,0416	0,2815	0,0499	-0,0218	0,3621	0,0497
(0,1,3)	θ_1	0,0035	0,3976	0,0489	-0,0333	0,3164	0,0490
	θ_2	-0,0414	0,2832	0,0501	-0,0200	0,3682	0,0502
	θ_3	-0,0016	0,3984	0,0496	-0,0125	0,3868	0,0508
(1,1,0)	φ_1	0,0036	0,3976	0,0489	-0,0320	0,3220	0,0490
(1,1,1)	φ_1	0,9033	0,0000	0,1087	0,4101	0,3589	0,8956
	θ_1	-0,9234	0,0000	0,0984	-0,4465	0,3509	0,8845
(1,1,2)	φ_1	0,1221	0,3970	1,3325	0,2166	0,3941	1,4159
	θ_1	-0,1186	0,3971	1,3324	-0,2497	0,3925	1,4158
	θ_2	-0,0413	0,2833	0,0500	-0,0156	0,3886	0,0690
(1,1,3)	φ_1	0,2088	0,3975	2,7303	0,1871	0,3956	1,5042
	θ_1	-0,2052	0,3976	2,7304	-0,2202	0,3944	1,5037
	θ_2	-0,0420	0,2876	0,0520	-0,0139	0,3908	0,0696
	θ_3	0,0490	0,3984	0,1214	-0,0093	0,3949	0,0671
(2,1,1)	φ_1	-0,7645	0,0054	0,2599	-1,0141	0,0000	0,0538
	φ_2	-0,0295	0,3386	0,0517	-0,0447	0,2655	0,0496
	θ_1	0,7699	0,0044	0,2556	0,9851	0,0000	0,0235
(2,1,2)	φ_1	-1,8397	0,0000	0,0325	-0,6605	0,0000	0,0275
	φ_2	-0,9645	0,0000	0,0295	-0,9370	0,0000	0,0196
	θ_1	1,8797	0,0000	0,0597	0,6723	0,0000	0,0140
	θ_2	0,9949	0,0000	0,0638	1,0000	0,0000	0,0173
(2,1,3)	φ_1	0,1531	0,3518	0,3066	-0,5355	0,0000	0,0206
	φ_2	0,6846	0,0126	0,2598	-0,9527	0,0000	0,0169
	θ_1	-0,1500	0,3546	0,3101	0,5102	0,0000	0,0534
	θ_2	-0,7329	0,0044	0,2429	0,9618	0,0000	0,0344
	θ_3	0,0091	0,3937	0,0576	-0,0662	0,1794	0,0524

Fonte: Elaborado pelo autor.

Para a seleção dos modelos utilizados no estudo que apresentaram parâmetros significativos, foram utilizados os critérios de Akaike, Erro Quadrático Médio de previsão (EQMp) e Erro Percentual Absoluto Médio (MAPE). O modelo escolhido será sempre aquele que apresentar o menor valor nos três testes. A série que obteve os melhores resultados foi da diferença de logaritmo de EGIE3 onde apresenta o menor valor nos três critérios, juntamente ao modelo com dois parâmetros Auto-Regressivo e dois parâmetro de Médias Móveis. Modelos com valores mais baixos no critério de Akaike sugerem equilíbrio eficaz entre a capacidade de se ajustar aos dados e a simplicidade inerente do modelo. Quanto ao Erro Quadrático Médio, avalia a média dos quadrados dos desvios entre as previsões do modelo e os valores reais, um EQM baixo indica que, em média, os erros individuais são pequenos, o que implica um ajuste consistente do modelo aos dados. Por sua vez, o Erro Médio Percentual Absoluto expressa os desvios médios como uma porcentagem dos valores reais, proporcionando uma interpretação intuitiva do desempenho do modelo em termos percentuais.

Considerando o modelo ARIMA (2,1,2), verificou-se por meio do teste de Lilliefors, apresen-

Tabela 4.3: Critério de Seleção do Modelo (EGIE3)

Série	Modelo	AIC	EQMp	MAPE
Diferença de EGIE3	(1,1,1)	1145,42	1,7904	0,0286
	(2,1,2)	1147,76	1,9454	0,0299
Diferença de Log	(2,1,2)	-1672,65	0,0013	0,0081

Fonte: Elaborado pelo autor.

tado na Tabela 4.4 que os resíduos apresentam distribuição normal, pois o valor-p apresentou-se maior do que 0,05. Este resultado é importante, pois corrobora para o fato de que este modelo, seja adequado.

Tabela 4.4: Teste para normalidade dos resíduos do modelo ARIMA (2,1,2)

Teste de Hipótese	Estatística	Valor-p
Teste de Lilliefors	0,0440	0,0505

Fonte: Elaborado pelo autor.

Pela Figura 4.6 observa-se que os resíduos seguem média zero e possuem homoscedasticidade na variação do tempo.

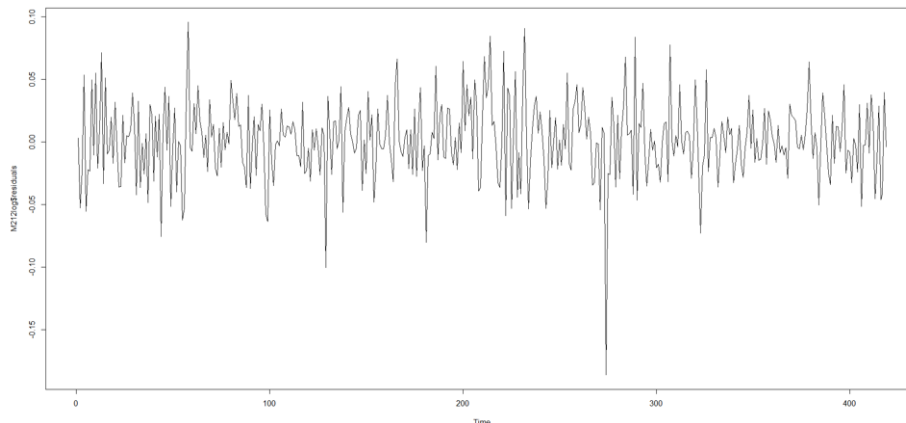


Figura 4.6: Gráfico de linha dos Resíduos do modelo ARIMA (2,1,2)

Fonte: Elaborado pelo autor.

Na Figura 4.7, o modelo ARIMA(2,1,2) foi testado por meio da análise do Espectrograma. Segundo Souza (1991) como o teste baseado no estudo do espectro acumulado mostra os valores dentro dos limites de confiança obtidos pelo teste de Smirnov-Kolmogorov, afirma-se que os resíduos são ruído branco.

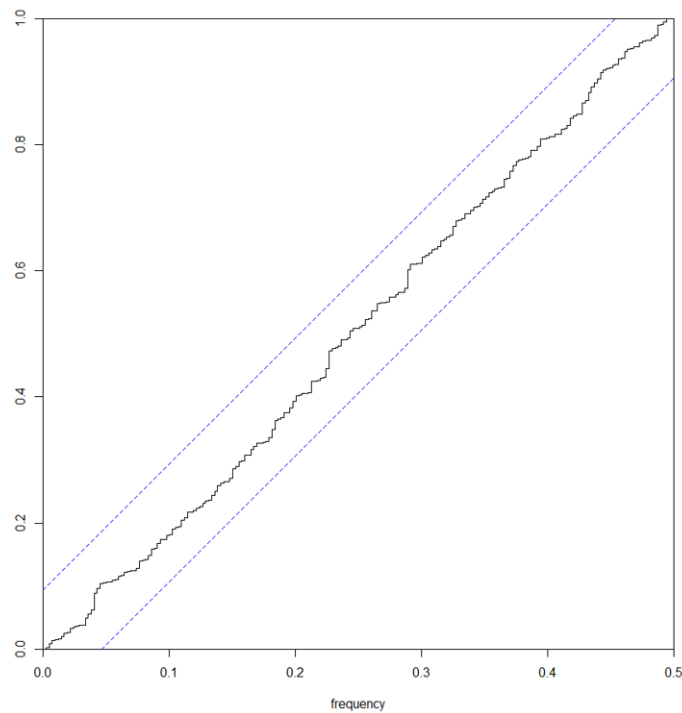


Figura 4.7: Teste do Espectrograma para o modelo ARIMA (2,1,2)
Fonte: Elaborado pelo autor.

Após as análises informais e os testes de hipóteses pode-se afirmar que os erros do modelo estimados estão normalmente distribuídos, não são autocorrelacionados e tem média zero com variação constante ao longo do tempo, ou seja, é ruído branco, concluindo então que o modelo foi bem ajustado.

Na Figura 4.8 o modelo ARIMA(2,1,2) é apresentado com a previsão do fechamento semanal da ação EGIE3 para as próximas 10 semanas em azul escuro, e em azul os intervalos de confiança para cada fechamento previsto.

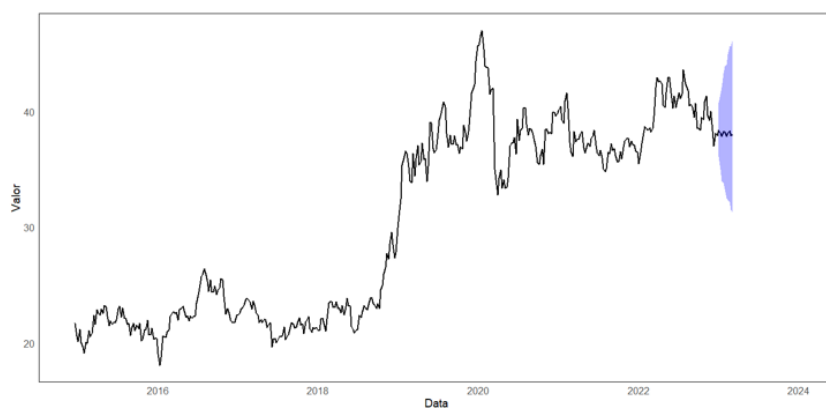


Figura 4.8: Previsão de fechamento de EGIE3 para as primeiras semanas de 2023
Fonte: Elaborado pelo autor.

Devido a análise dos dados ter sido realizada até o final de 2022, já temos os resultados de fechamento de EGIE3 para 2023, com isso podemos observar que nas primeiras 10 semanas de 2023 o resultado de fechamento de EGIE3 foi dentro dos limites e com valores coerentes ao

fechamento previsto, conforme podemos observar na Tabela 4.5 e na Figura 4.9.

Tabela 4.5: Fechamento EGIE3 realizado, Previsão do modelo, limites inferiores e superiores do modelo para as primeiras 10 semanas de 2023

Data	EGIE3 Fechamento	Prev. Fechamento	Lim. Inf.	Lim. Sup.
01/01/2023	36,87	38,37	36,13	40,75
08/01/2023	37,84	38,09	34,97	41,50
15/01/2023	38,16	37,82	33,98	42,09
22/01/2023	38,85	38,26	33,82	43,27
29/01/2023	37,94	38,23	33,33	43,85
05/02/2023	38,03	37,84	32,53	44,01
12/02/2023	39,39	38,12	32,37	44,89
19/02/2023	39,01	38,30	32,18	45,58
26/02/2023	39,26	37,91	31,51	45,62
05/03/2023	40,93	38,00	31,25	46,21

Fonte: Elaborado pelo autor.

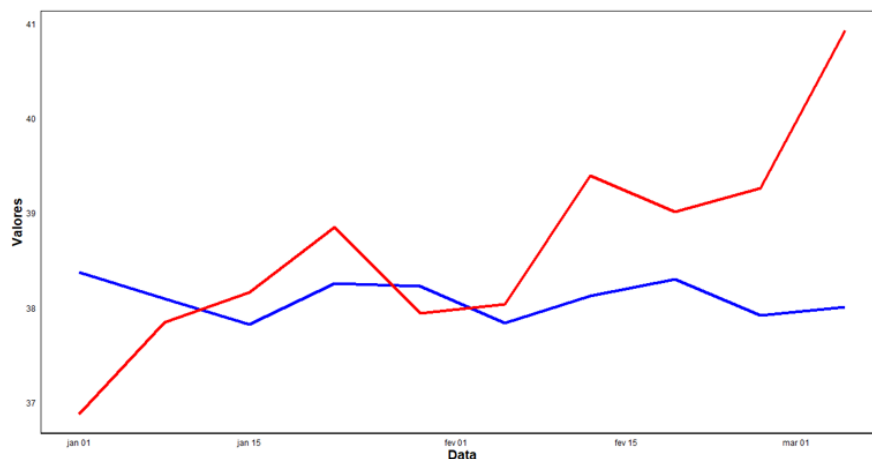


Figura 4.9: Fechamento EGIE3 realizado 2023 (em vermelho) versus previsão de fechamento EGIE3 (em azul)

Fonte: Elaborado pelo autor.

4.2 ENERGIAS DO BRASIL S.A. (ENBR3)

A Figura 4.10 exibe o gráfico da série histórica da ação ENBR3, revelando ausência de sazonalidade e ciclicidade nos dados, devido à presença de uma tendência na série, e portanto, a mesma é considerada uma série não estacionária. Conforme sugerido por Morettin & Tolo [28], para eliminar essa tendência, realizou-se a diferenciação da série, evidenciada na Figura 4.11, onde a primeira diferença revelou uma média constante com baixa variabilidade.

Com o propósito de reduzir a variabilidade, os dados foram submetidos a uma transformação logarítmica, resultando na série apresentada na Figura 4.12. A mesma transformação foi aplicada à diferença, ilustrada na Figura 4.13, revelando uma variabilidade mais uniforme em torno da média zero, sugerindo indícios de estacionariedade e eliminando a necessidade de realizar outra diferenciação.

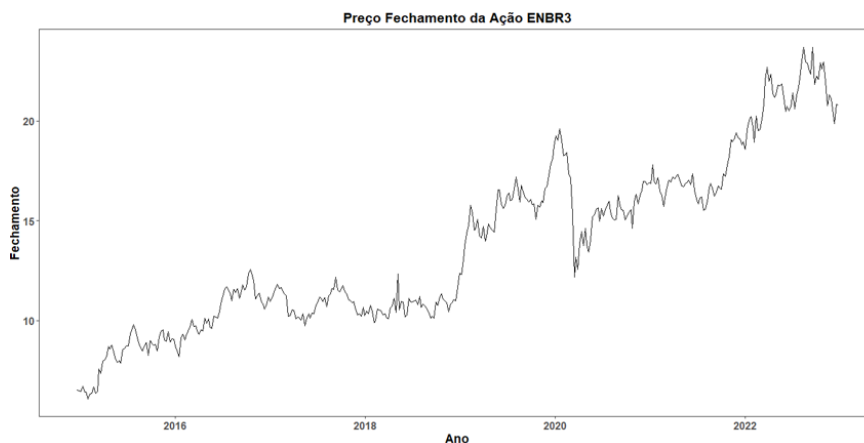


Figura 4.10: Gráfico de linha da série histórica da ação ENBR3
Fonte: Elaborado pelo autor.

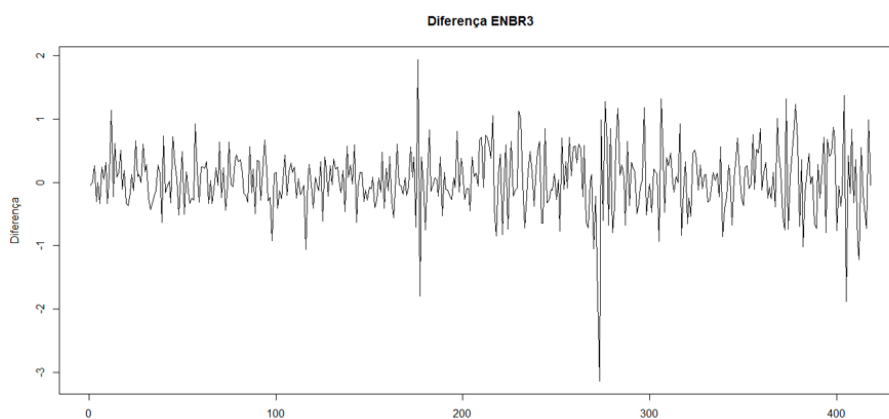


Figura 4.11: Gráfico de linha da diferença da série ENBR3
Fonte: Elaborado pelo autor.

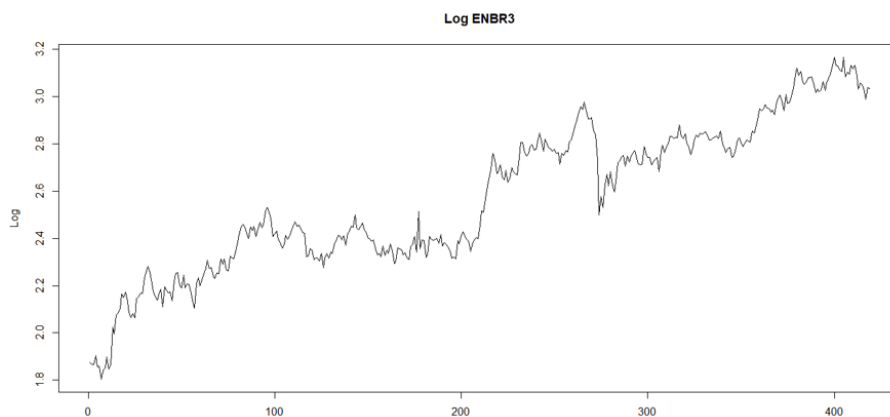


Figura 4.12: Gráfico de linha de log da série histórica ENBR3
Fonte: Elaborado pelo autor.

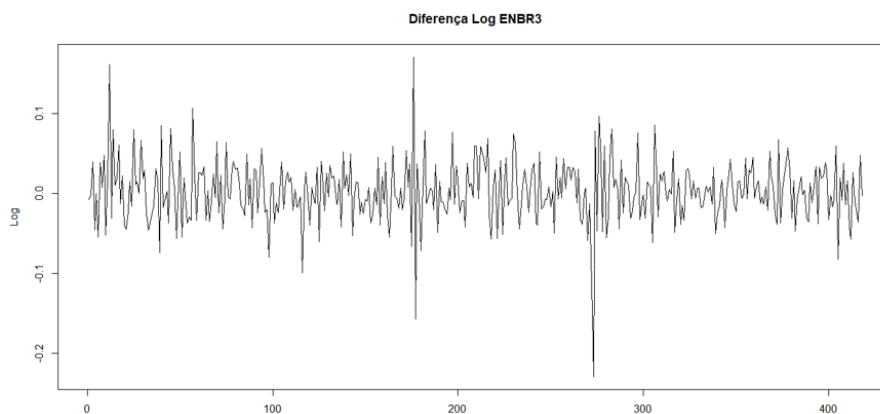


Figura 4.13: Gráfico de linha da diferença de log da série histórica ENBR3

Fonte: Elaborado pelo autor.

Na Tabela 4.6 o teste ADF (Augmented Dickey-Fuller), foi executado para testar a hipótese de que a série possui raiz unitária. Essa hipótese não foi rejeitada para a série de ENBR3, pois exibiu valor-p maior do que 0,05. Porém para a diferença da série, Log e para diferença de Log, podemos ver que a hipótese foi rejeitada, ou seja, são estacionárias.

Tabela 4.6: Testes ADF para estacionariedade (EGIE3)

Série Temporal	Estatística	Valor-p
ENBR3	-2,3782	0,4177
Diferença de ENBR	-8,0825	0,0100
Log	-4,0415	0,0100
Diferença de Log	-8,2493	0,0100

Fonte: Elaborado pelo autor.

Na análise das autocorrelações, constata-se na Figura 4.14(a) que o gráfico de autocorrelação para a diferença de ENBR3 os “lags” 1, 8, 10 e 21 ultrapassam os limites, indicando que possuem correlações significativas, o que pode sugerir a ordem da parte de Médias Móveis. Na Figura 4.14(b) os “lags” 1, 8, 10 e 21 são significativos, os resultados obtidos por meio da função de autocorrelação parcial, podem indicar a ordem da parte Auto-Regressiva. Para a diferença do logaritmo de ENBR3 na Figura 4.14(c) os “lags” 1 e 8 são significativos, indicando a ordem da parte de Médias Móveis, e na figura 4.14(d) os “lags” 1 e 8 são significativos, indicando a ordem da parte Auto-Regressiva. Em função destes resultados encontrados, utilizando o software R foi realizado ajustes dos modelos com até quatro parâmetros de Médias Móveis e quatro Auto-Regressivos.

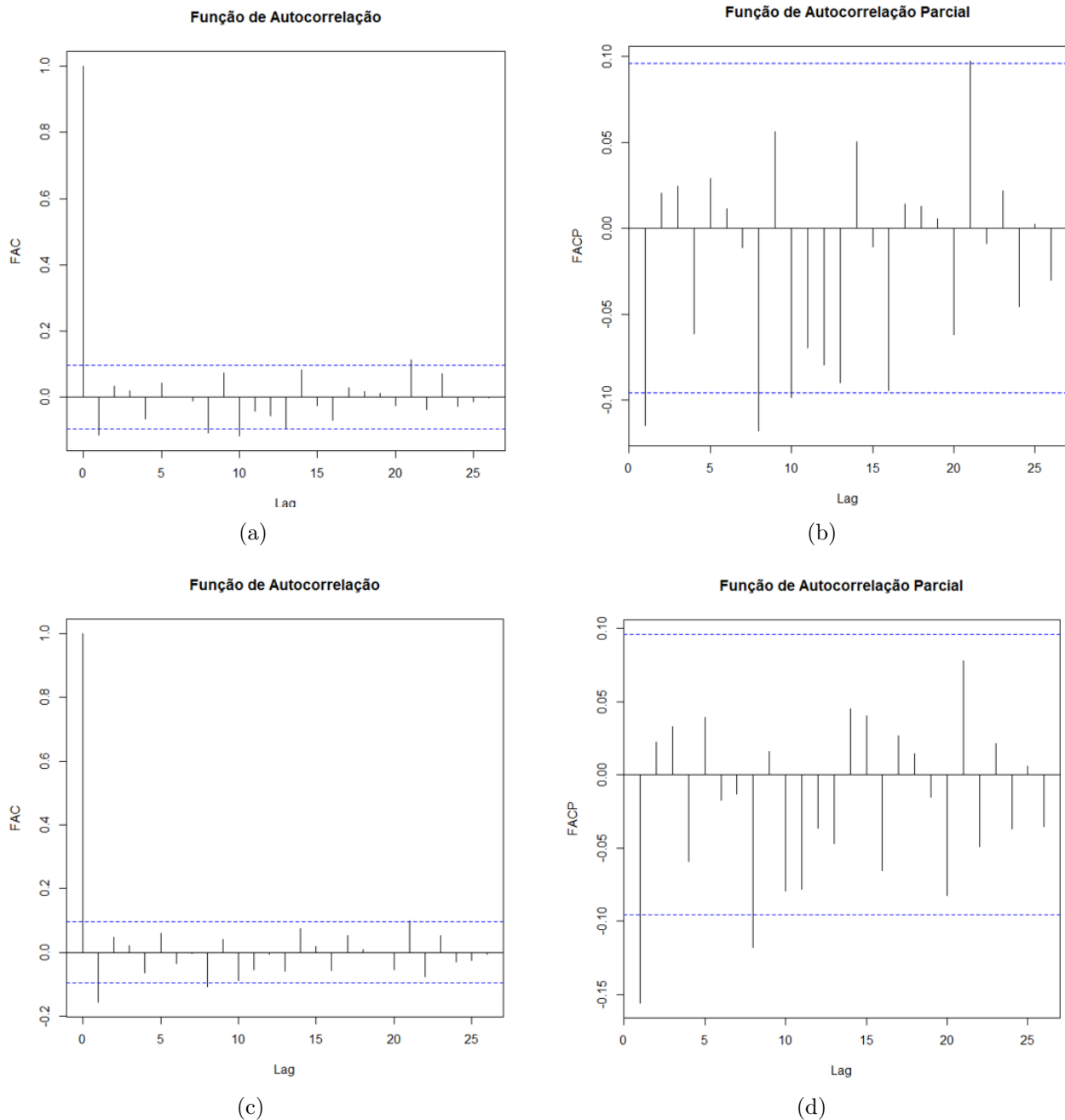


Figura 4.14: Função de Autocorrelação e Função de Autocorrelação Parcial, para Diferença de ENBR3 e Diferença de Log

Fonte: Elaborado pelo autor.

Levando em conta modelos com quatro parâmetros de Médias Móveis e quatro Auto-Regressivos, testou-se a significância das estimativas, obtidas por meio dos estimadores dos parâmetros a partir do teste t de Student, considerando um nível nominal de significância de 5% e considerando os dados da diferença da série e diferença de log da série.

Os modelos que tiveram a hipótese nula rejeitada, ou seja, que possuem todos os parâmetros significativos foram os modelos da primeira diferença da série ENBR3 ARIMA: (0,1,1), (1,1,0), (3,1,4) e (4,1,3); e da série com diferença de log ARIMA: (0,1,1), (1,1,0), (3,1,4) e (4,1,3).

Tabela 4.7: Teste t de Student para significância dos parâmetros nos modelos ajustados (ENBR3)

ARIMA	Parâmetro	Diferença em ENBR3			Diferença de Log		
		Estatística	Valor-p	Desv. Pad.	Estatística	Valor-p	Desv. Pad.
(0,1,1)	θ_1	-1,1024	0,0363	0,0467	-0,1370	0,0047	0,0458
(0,1,2)	θ_1	-0,1073	0,0355	0,0487	-0,1459	0,0047	0,0487
	θ_2	0,0449	0,2717	0,0513	0,0613	0,1896	0,0503
(0,1,3)	θ_1	-0,1063	0,0389	0,0492	-0,1449	0,0053	0,0491
	θ_2	0,0432	0,2844	0,0526	0,0591	0,2099	0,0522
	θ_3	0,0070	0,3948	0,0499	0,0078	0,3941	0,0515
(0,1,4)	θ_1	-0,0996	0,0521	0,0493	-0,1392	0,0077	0,0494
	θ_2	0,0407	0,2869	0,0502	0,0532	0,2298	0,0507
	θ_3	0,0087	0,3929	0,0507	0,0119	0,3885	0,0522
	θ_4	-0,061	0,2145	0,0548	-0,0474	0,2715	0,0541
(1,1,0)	φ_1	-0,1094	0,0319	0,0486	-0,1492	0,0035	0,0483
(1,1,1)	φ_1	-0,2369	0,2943	0,3043	-0,2598	0,2145	0,1128
	θ_1	0,1286	0,3655	0,3085	0,2334	0,3562	0,2377
(3,1,4)	φ_1	-0,7804	0,0000	0,1104	-0,9506	0,0000	0,1926
	φ_2	-0,7434	0,0000	0,1129	-0,8041	0,0004	0,2160
	φ_3	-0,7781	0,0000	0,0878	-0,6824	0,0000	0,1643
	θ_1	0,6921	0,0000	0,1139	0,8238	0,0000	0,1917
	θ_2	0,6874	0,0000	0,1132	0,7071	0,0007	0,1980
	θ_3	0,7241	0,0000	0,0914	0,6175	0,0000	0,1421
	θ_4	-0,1546	0,0041	0,0509	-0,1499	0,0125	0,0568
(4,1,3)	φ_1	-0,9258	0,0000	0,1190	0,6465	0,0000	0,0630
	φ_2	-0,8766	0,0000	0,1032	0,7664	0,0000	0,0865
	φ_3	-0,9275	0,0000	0,0900	-0,7409	0,0000	0,0644
	φ_4	-0,1647	0,0038	0,0537	-0,1847	0,0004	0,0497
	θ_1	0,8269	0,0000	0,1136	-0,8176	0,0000	0,0441
	θ_2	0,8106	0,0000	0,1016	-0,6548	0,0000	0,0741
	θ_3	0,8631	0,0000	0,0651	0,9410	0,0000	0,0434
(4,1,4)	φ_1	-0,8940	0,0000	0,1085	-0,2208	0,2371	0,2167
	φ_2	-0,1950	0,2161	0,1763	0,6580	0,0000	0,0705
	φ_3	0,6192	0,0008	0,1750	-0,3028	0,1052	0,1854
	φ_4	0,7783	0,0000	0,1041	-0,8416	0,0008	0,2383
	θ_1	0,8591	0,0000	0,0902	0,1029	0,3723	0,2783
	θ_2	0,1707	0,2019	0,1464	-0,6544	0,0000	0,1023
	θ_3	-0,6502	0,0000	0,1468	0,3953	0,0982	0,2360
	θ_4	-0,8713	0,0000	0,0886	0,8074	0,0219	0,3345

Fonte: Elaborado pelo autor.

Para a seleção dos modelos avaliados no estudo os quais apresentaram parâmetros significativos, foram utilizados os critérios de Akaike, Erro Quadrático Médio de previsão (EQMp) e Erro Percentual Absoluto Médio (MAPE). O modelo escolhido será sempre aquele que apresentar o menor valor nos três indicadores. A série que obteve os melhores resultados foi da diferença de logaritmo de ENBR3 onde apresenta o menor valor nos três critérios, juntamente ao modelo com três parâmetros Auto-Regressivo e quatro parâmetros de Médias Móveis. Modelos com menor valor de Akaike indicam um bom equilíbrio entre a capacidade de ajustar os dados e a

simplicidade do modelo, para Erro Quadrático Médio de previsão avalia a média dos quadrados dos erros entre os valores previstos e os valores reais, um EQM baixo indica que os erros individuais são pequenos em média, o que sugere que o modelo está bem ajustado aos dados e o critério de Erro Percentual Absoluto Médio expressa o erro médio como uma porcentagem dos valores reais, o MAPE fornece uma interpretação intuitiva em termos de porcentagem.

Tabela 4.8: Critério de Seleção do Modelo (ENBR3)

Série	Modelo	AIC	EQMp	MAPE
Diferença de ENBR3	(0,1,1)	625,67	3,2181	0,0757
	(1,1,0)	625,35	3,2183	0,0757
	(3,1,4)	626,38	3,1728	0,0749
	(4,1,3)	625,46	3,4109	0,0776
Diferença de Log	(0,1,1)	-1542,42	0,0098	0,0256
	(1,1,0)	-1542,66	0,0096	0,0256
	(3,1,4)	-1547,76	0,0072	0,0246
	(4,1,3)	-1545,81	0,0092	0,0277

Fonte: Elaborado pelo autor.

Considerando o modelo (3,1,4), verificou-se por meio do teste de Lilliefors, apresentado na Tabela 4.9 que os resíduos apresentam distribuição normal, pois o valor-p apresentou-se maior do que 0,05. Este desfecho é significativo, e portanto corrobora com a adequação desse modelo.

Tabela 4.9: Teste para normalidade dos resíduos do modelo ARIMA (3,1,4)

Teste de Hipótese	Estatística	Valor-p
Teste de Lilliefors	0,0420	0,0732

Fonte: Elaborado pelo autor.

Através da Figura 4.15, é visível que os resíduos mantêm uma média zero e exibem uma variabilidade constante ao longo do tempo.

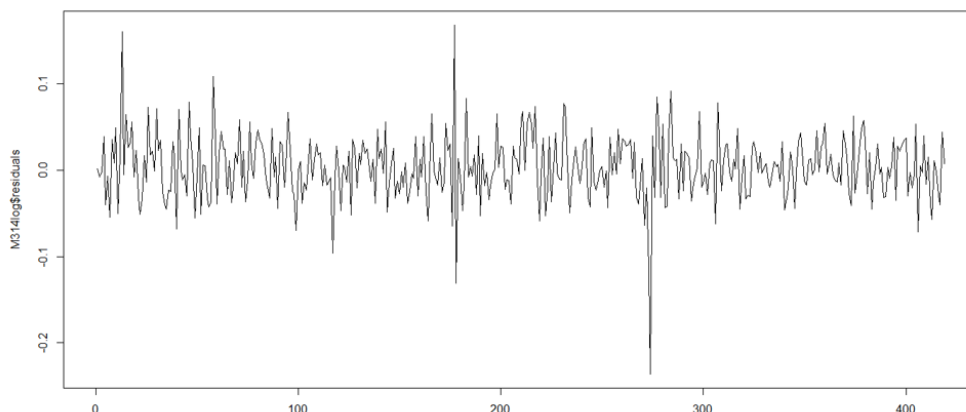


Figura 4.15: Gráfico de linha dos Resíduos do modelo ARIMA (3,1,4)

Fonte: Elaborado pelo autor.

Na Figura 4.16, o modelo ARIMA (3,1,4) foi examinado usando a análise do Espectrograma. Conforme Souza (1991), a análise do espectro acumulado exibe valores que se encontram nos li-

mites de confiança estabelecidos pelo teste de Smirnov-Kolmogorov, o que indica que os resíduos se comportam como ruído branco.

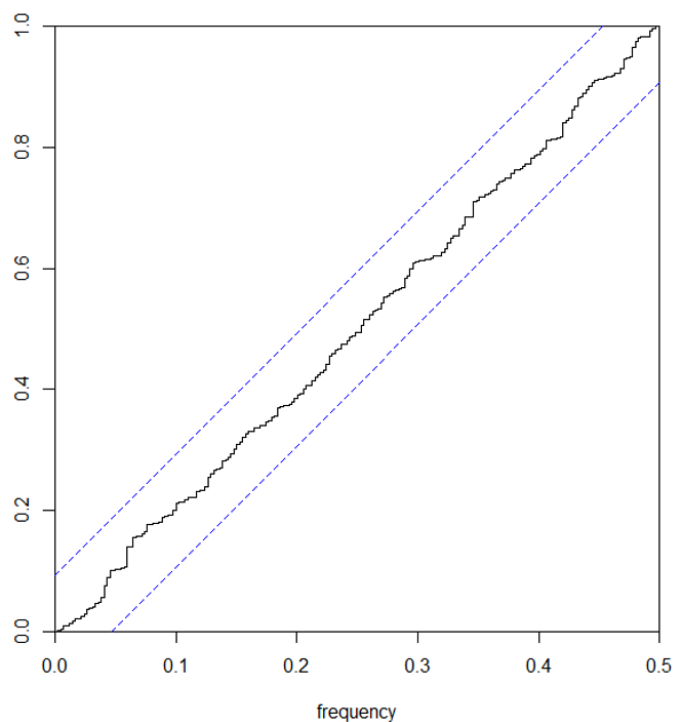


Figura 4.16: Teste do Espectrograma para o modelo ARIMA (3,1,4)
Fonte: Elaborado pelo autor.

Após as avaliações informais e os testes de hipóteses, pode-se concluir que os erros estimados do modelo exibem uma distribuição normal, não possuem autocorrelação, mantêm média zero e apresentam variação constante ao longo do tempo, caracterizando-se assim como ruído branco. Dessa forma, pode-se afirmar que o modelo foi ajustado de maneira satisfatória.

Na Figura 4.17 o modelo ARIMA (3,1,4) são exibidas as previsões do fechamento semanal da ação ENBR3 para as próximas 10 semanas, representadas por uma linha em azul escuro, junto aos intervalos de confiança, destacados em tons mais claros de azul, para cada previsão do fechamento.



Figura 4.17: Previsão de fechamento de ENBR3 para as primeiras semanas de 2023
Fonte: Elaborado pelo autor.

Após a análise dos dados até final de 2022, já dispomos dos resultados de fechamento de ENBR3 para o ano de 2023. Observamos que nas primeiras 10 semanas de 2023, os resultados de fechamento de ENBR3 mantiveram-se dentro dos limites, mas com divergências aos valores previstos, como evidenciado na Tabela 4.10 e Figura 4.18.

Tabela 4.10: Fechamento ENBR3 realizado, Previsão do modelo, limites inferiores e superiores do modelo para as primeiras 10 semanas de 2023

Data	ENBR3 Fechamento	Previsão Fechamento	Lim. Inf.	Lim. Sup.
01/01/2023	19,55	20,88	19,41	22,47
08/01/2023	19,74	20,94	19,00	23,08
15/01/2023	20,20	20,80	18,50	23,39
22/01/2023	20,17	20,80	18,18	23,81
29/01/2023	20,44	20,87	18,01	24,19
05/02/2023	19,95	20,90	17,79	24,55
12/02/2023	20,22	20,82	17,50	24,76
19/02/2023	19,64	20,82	17,30	25,07
26/02/2023	22,11	20,87	17,16	25,37
05/03/2023	22,00	20,88	16,99	25,64

Fonte: Elaborado pelo autor.

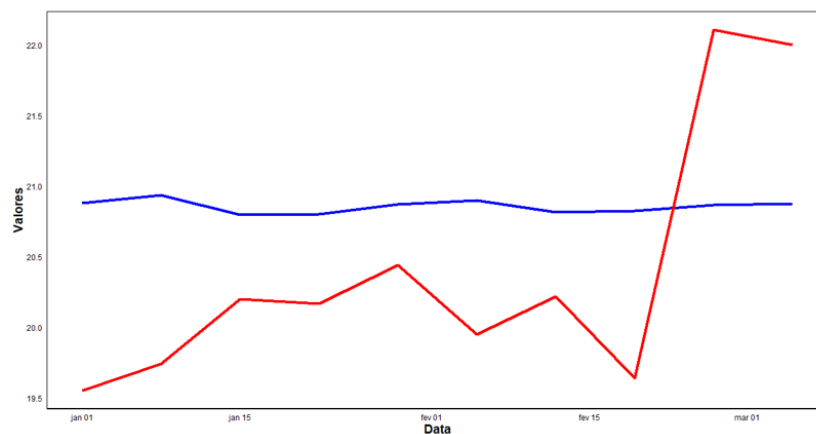


Figura 4.18: Fechamento ENBR3 realizado 2023 (linha vermelha) versus previsão de fechamento ENBR3 (linha azul)

Fonte: Elaborado pelo autor.

5. CONCLUSÕES

Nas condições em que foram analisados os dados, os modelos que apresentaram os melhores ajustes foram o modelo ARIMA (2,1,2) para o ativo EGIE3 e o modelo ARIMA (3,1,4) para o ativo ENBR3. No decorrer deste estudo, a aplicação da metodologia de séries temporais, conforme proposta por Box e Jenkins, revelou-se não apenas uma abordagem eficaz, mas também altamente benéfica tanto para as empresas do setor de energia elétrica quanto para os consumidores. Ao adotar o modelo ARIMA (*AutoRegressive Integrated Moving Average*) para análise dos ativos EGIE3 e ENBR3, constatamos que esta metodologia oferece *insights* valiosos e previsões precisas.

A execução de séries temporais nos ativos pode auxiliar investidores e instituições financeiras a preverem antecipadamente os preços das ações, possibilitando alocar recursos de forma mais planejada.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Abdi, H. e Molin, P.: *Lilliefors/Van Soest's test of normality*. Encyclopedia of Measurement and Statistics, 2007.
- [2] Action, P.: *Teste e Intervalo de Confiança para Parâmetros*, 2012. <http://www.portalaction.com.br/analise-de-regressao/14-testes-e-intervalos-de-confianca-para-os-parametros>.
- [3] Antunes, C.H. e Mattos, M.S.: *Análise da Relação entre os Mercados Financeiro e de Energia Elétrica no Brasil: Uma Abordagem com Co-integração*. Revista de Economia e Sociologia Rural, 2016.
- [4] B3: *Brasil, Bolsa, Balcão*, 2023. https://www.b3.com.br/pt_br/b3/institucional/quem-somos.
- [5] Bodie, Z., Kane, A. e Marcus, A. J.: *Investimentos*. Pearson, 2014.
- [6] Box, G. E. P. e Jenkins, G. M.: *Time Series Analysis*. Holden Day, 1970.
- [7] Brookwell, P. J. e Davis, R. A.: *Introduction to Time Series and Forecasting*. Springer, 2002.
- [8] Chatfield, C.: *A Análise de Séries Temporais: Uma Introdução*. Chapman and Hall, 2019.
- [9] Conceição Franco, G. da: *Apostila sobre modelagem Box e Jenkins*. Tese de Doutorado, 2019.
- [10] Core Team R, R. F. f. S. C.: *R: A language and environment for statistical computing*, 2022. <https://www.R-project.org/>.
- [11] Cowpertwait, P. S. P. e Metcalfe, A. V.: *Introductory Time Series with R*. Springer, 2009.
- [12] EGIE3: *Engie Brasil Energias S.A.*, 2005. <https://www.engie.com.br/investidores/>.
- [13] Ehlers, R. S.: *ANálise de Séries Temporais*. 2009.
- [14] ENBR3: *Energias do Brasil S.A.*, 2005. <https://ri.edp.com.br/pt-br/>.

- [15] Energia (Brasil), M. de Minas e: *Plano Decenal de Expansão de Energia 2027*, 2018. https://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/Documents/PDE%202027_aprovado_OFICIAL.pdf.
- [16] Faria, K. O. S.: *Estudo da série temporal do preço da arroba do boi gordo da BM&F para o estado de Goiás*. Trabalho de Conclusão de Curso, 2015. <https://repositorio.ufu.br/handle/123456789/17827>.
- [17] Fonseca, F., Moutinho, V. e Figueira, C.: *Regulação e Política Energética*. Elsevier, 2019.
- [18] Goldemberg, J., Lucchesi, P. R. e Simões, A. F.: *Energia no Brasil*. 2019.
- [19] Greene, W. H.: *Econometric Analysis*. Pearson, 2017.
- [20] Gutierrez, J. L. C.: *Monitoring of the Corumbá - i Dam Instrumentation Neural Networks and the Box Jenkins null Models*. 2003.
- [21] Hull, J. C.: *Opções, Futuros e Outros Derivativos*. Pearson, 2017.
- [22] Hyndman, R. J. e Athanasopoulos, G.: *Séries Temporais Econométricas Aplicadas*. 2018.
- [23] Jones, A. e Brown, M.: *Technological Innovations in the Electric Power Sector: Sustainability, Efficiency, and Resilience*. Renewable Energy Journal, 2020.
- [24] Madura, J.: *International Financial Management*. Cengage, 2018.
- [25] Makridakis, S., Wheelwright, S. e Hyndman, R.: *Forecasting: Methods and Applications*. New York: John Wiley & Sons, 3ª ed., 1998.
- [26] Mariani, E., Lourenço, A. L. e Santos, E. M.: *Políticas Públicas para o Setor Elétrico Brasileiro: Incentivos às Energias Renováveis e Eficiência Energética*. Economia e Energia: Estudos Multidisciplinares, 2019.
- [27] Minitab: *Suporte ao Minitab*, 2016.
- [28] Morettin, P. A. e Castro Toloí, C. M. de: *Modelos para Previsão de Séries Temporais*. 1981. https://impa.br/wp-content/uploads/2017/04/13_CBM_81_01.pdf.
- [29] Morettin, P. A. e Castro Toloí, C. M. de: *Análise de Séries Temporais*. Edgard Blucher, 2ª ed., 2006.
- [30] O'hara, M.: *Negociação de Alta Frequência: Novas Realidades para Traders, Mercados e Reguladores*. Revista de Mercados Financeiros, 2015.
- [31] Pindyck, R. S. e Rubinfeld, D.: *Econometria Modelos Previsões*. Elsevier, 4ª ed., 2004.
- [32] Priestley, M. B.: *Spectral analysis and time series*. New York: Academic Press, 6ª ed., 1989.

- [33] ProfitChart: *Nelogica*, 2023. <https://www.nelogica.com.br/>.
- [34] Santos, F.M. do: *Previsão de Médio Prazo de Demanda de Energia Elétrica do Setor Industrial na Região Sul e Sudeste do Brasil*. Trabalho de Conclusão de Curso, 2019. <http://repositorio.utfpr.edu.br/jspui/handle/1/12929>.
- [35] Shuumway, R. H. e Stoffer, D. S.: *Time Series Analysis and Its Applications: With R Examples*. Springer, 2017.
- [36] Siegel, J. J.: *Ações - Lucrando em Qualquer Mercado*. McGraw-Hill Education, 2014.
- [37] Smith, J.: *The Impact of Technology on Capital Markets: Operational Efficiency and Transparency*. Journal Of financial Technology, 2021.
- [38] Souza, A. M.: *Aplicação e performance da análise de intervenção em séries macroeconômicas brasileiras*. 1991.
- [39] Stevenson, W. J.: *Estatística aplicada à administração*. Harbra, 2001.
- [40] Sá, C. A.: *Gerenciamento de Riscos Financeiros em Instituições Financeiras*. Atlas, 2015.
- [41] Žitković, G.: *Introduction to Stochastic Processes*. 2010.