

**UNIVERSIDADE FEDERAL DE JUIZ DE FORA
FACULDADE DE ECONOMIA
CURSO DE CIÊNCIAS ECONÔMICAS**

Tufy Antônio Baracat Junior

Modelo de previsão de mercado para o consumo de gasolina no Brasil

Juiz de Fora

2024

Tufy Antônio Baracat Junior

Modelo de previsão de mercado para o consumo de gasolina no Brasil

Monografia apresentada ao Curso de Ciências Econômicas da Universidade Federal de Juiz de Fora, como pré-requisito para a obtenção do título de Bacharel em Ciências Econômicas.

Orientador: Alexandre Zanini

Juiz de Fora

2024

Ficha catalográfica elaborada através do programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

Baracat Junior, Tufy Antônio.

Modelo de previsão de mercado para o consumo de gasolina no Brasil / Tufy Antônio Baracat Junior. -- 2024.

50 p. : il.

Orientador: Alexandre Zanini

Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Juiz de Fora, Faculdade de Economia, 2024.

1. Gasolina. 2. Séries Temporais. 3. Amortecimento Exponencial. 4. Box & Jenkins. 5. Competição de métodos. I. Zanini, Alexandre, orient. II. Título.



**UNIVERSIDADE FEDERAL DE JUIZ DE FORA
REITORIA - FACECON - Depto. de Economia**

FACULDADE DE ECONOMIA / UFJF

ATA DE APROVAÇÃO DE MONOGRAFIA II (MONO B)

Na data de 02/07/2024, a Banca Examinadora, composta pelos professores

1 - Alexandre Zanini - orientador; e

2 - Rogério Silva de Mattos,

reuniu-se para avaliar a monografia do acadêmico TUFY ANTÔNIO BARACAT JÚNIOR, intitulada: MODELO DE PREVISÃO DE MERCADO PARA O CONSUMO DE GASOLINA NO BRASIL.

Após primeira avaliação, resolveu a Banca sugerir alterações ao texto apresentado, conforme relatório sintetizado pelo orientador. A Banca, delegando ao orientador a observância das alterações propostas, resolveu APROVAR a referida monografia.

ASSINATURA ELETRÔNICA DOS PROFESSORES AVALIADORES



Documento assinado eletronicamente por Alexandre Zanini, Professor(a), em 02/07/2024, às 18:51, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por Rogério Silva de Mattos, Professor(a), em 02/07/2024, às 19:05, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no Portal do SEI-Ufjf (www2.ufjf.br/SEI) através do ícone Conferência de Documentos, informando o código verificador 1846100 e o código CRC FDCDCC03.

Dedico este trabalho ao meu pai, Tufy Neto. Obrigado por sempre me incentivar e investir nos meus estudos.

RESUMO

O objetivo principal deste trabalho é prever e analisar o consumo de gasolina no Brasil. Devido à grande influência deste combustível para matriz energética brasileira, justifica-se a importância de um estudo voltado para a projeção de dados sobre esse derivado. Logo, é possível que as informações fornecidas sejam base para decisões de agentes econômicos. A base de dados é composta por dados mensais da média diária do consumo de gasolina, medidos em milhares de barris, entre os períodos de janeiro de 1979 e março de 2024. Os dados selecionados evidenciam o consumo aparente equivalente à produção industrial doméstica acrescida das importações e diminuída das exportações. Os métodos aplicados foram o Método de Amortecimento Exponencial (MAE) e o Método de Box & Jenkins (MBJ). Uma análise recursiva fora da amostra foi conduzida para selecionar o método mais eficaz, realizando uma competição entre os dois. Selecionado o método vencedor (MAE), utilizando um modelo Holt-Winters, teve-se como resultado as previsões de consumo de gasolina no Brasil até dezembro de 2025. Os resultados encontrados mostram a possível ocorrência de uma queda no consumo em 2025.

Palavras-chave: Gasolina, Séries temporais, Amortecimento Exponencial, Box & Jenkins, Competição de métodos.

ABSTRACT

The main objective of this work is to forecast and analyze gasoline consumption in Brazil. Due to the significant influence of this fuel on the Brazilian energy matrix, the importance of a study focused on data projection for this derivative is justified. Consequently, the information provided can serve as a basis for economic agents decisions. The database consists of monthly data on the daily average gasoline consumption, measured in thousands of barrels, from January 1979 to March 2024. The selected data highlight the apparent consumption equivalent to domestic industrial production plus imports and minus exports. The applied methods were the Exponential Smoothing Method (ESM) and the Box & Jenkins Method (BJM). An out-of-sample recursive analysis was conducted to select the most effective method, resulting in a competition between them. The winning method (ESM), using a Holt-Winters model, provided the gasoline consumption forecasts in Brazil until December 2025. The results indicate a possible decline in consumption in 2025.

Key words: Gasoline, Time series, Exponential Smoothing, Box & Jenkins, Method competition.

LISTA DE FIGURAS

| | |
|---|----|
| Figura 1 – Demanda mensal de Gasolina C..... | 10 |
| Figura 2 – Vendas de combustíveis líquidos – mil m ³ /dia | 11 |
| Figura 3 – Procedimento de out-of-sample rolling evaluation | 29 |
| Figura 4 – Consumo de gasolina no Brasil (em milhares de barris). | 33 |
| Figura 5 – Função de Autocorrelação dos Erros – MAE. | 36 |
| Figura 6 – Função de Autocorrelação dos Erros – MBJ. | 38 |
| Figura 7 – Consumo de gasolina (em milhares de barris) – período de 1979 a 2025. Valores reais, ajustado e previsões com intervalo de confiança de 95%. | 40 |

LISTA DE QUADROS

| | |
|--|----|
| Quadro 1 – Resumo das características teóricas da FAC e da FACP dos modelos AR(p), MA(q) e ARMA (p,q)..... | 27 |
|--|----|

LISTA DE TABELAS

| | |
|---|----|
| Tabela 1 – Produção de derivados de petróleo, energéticos e não energéticos – 2012 – 2023. | 9 |
| Tabela 2 – Parâmetro do Método de Amortecimento Exponencial | 34 |
| Tabela 3 – Fatores Sazonais..... | 34 |
| Tabela 4 – Estatística de desempenho preditivo – MAE | 35 |
| Tabela 5 – Parâmetros do Modelo de Box & Jenkins | 37 |
| Tabela 6 – Estatística dentro da amostra – MBJ | 37 |
| Tabela 7 – Competição entre os Métodos. | 38 |
| Tabela 8 – Valores da Previsão do Consumo de Gasolina (em milhares de barris) – Intervalo de Confiança de 95%. | 41 |

SUMÁRIO

| | |
|--|-----------|
| 1 INTRODUÇÃO | 7 |
| 2 MERCADO DE COMBUSTÍVEL..... | 9 |
| 3 REVISÃO DE LITERATURA..... | 13 |
| 4 METODOLOGIA | 15 |
| 4.1 Séries Temporais e Processos Estocásticos | 15 |
| 4.2 Método de Amortecimento Exponencial (MAE) | 16 |
| 4.3 Método Univariado de Box & Jenkins (MBJ)..... | 25 |
| 4.4 Competição de Métodos | 28 |
| 4.4.1 Desempenho Preditivo | 30 |
| 5 ANÁLISE DE DADOS E RESULTADOS | 33 |
| 5.1 Base da dados | 33 |
| 5.2 Equações de Previsão..... | 34 |
| 5.3 Competição de Métodos | 38 |
| 5.4 Previsão do consumo de Gasolina | 39 |
| 6 CONCLUSÃO | 42 |
| REFERÊNCIAS..... | 43 |

1 INTRODUÇÃO

Conforme dados da Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP), em 2024, o consumo de gasolina manteve-se estável, frente ao crescimento considerável da frota de veículos flex. Mesmo com a chegada de grandes inovações no setor automotivo, a gasolina ainda é consolidada como preferência para abastecimento de veículos leves. Isso ocorre apesar da popularização dos carros elétricos e híbridos, uma vez que os consumidores já possuem hábitos e costumes consolidados.

A gasolina representa cerca de 23% da produção de combustíveis. De 2022 para o ano de 2023, o aumento na produção foi de 8,3%. Um dos grandes responsáveis foi a crescente frota de veículos, que alavanca a demanda pelo combustível (ANP, 2024).

O etanol, maior concorrente da gasolina para uso em veículos leves, cresce a passos largos, atendendo à contínua demanda por energia. Esse fator coloca o petróleo e derivados como o álcool, em um patamar de grande importância para diversos setores da economia, principalmente no transporte e industrial (MME, 2008).

Com relação à oferta, o etanol sofre com evidentes barreiras à entrada neste setor, dada as desvantagens econômicas que possuem em relação ao mercado já consolidado. Corroborando com o argumento, a produção de gasolina esbanja maiores capacidades de armazenamento dos produtos, assim como melhor localização das bases, que são próximas das unidades produtoras. Esses fatores explicam como a demanda por gasolina C sofreu um aumento de 11% ao comparar a evolução entre os mesmos 12 meses de 2021 a 2022 e 2022 a 2023. (EPE, 2023).

Dada a importância da gasolina para matriz energética brasileira, assim como para economia do país, entende-se a necessidade de pesquisas na área.

Segundo Zanini:

A abertura da indústria nacional de petróleo e gás natural no Brasil fortalece a necessidade de gerar projeções confiáveis de demanda de combustíveis, em particular, os derivados de petróleo, gás natural e álcool. Acredita-se que uma melhor compreensão do funcionamento do mercado de petróleo e gás natural pode fornecer subsídios para estratégias de planejamento e gestão. (ZANINI, 2000, p.403).

Logo, este trabalho monográfico tem como principal objetivo fazer a predição do consumo de gasolina no país até o período de dezembro de 2025, com o uso de dados coletados do Instituto de Pesquisa Econômica Aplicada (IPEA). A base de dados deste trabalho possui dados mensais de 1979 até março de 2024.

Para atingir o objetivo do trabalho, foi realizado um estudo da metodologia de análise de séries temporais para aplicá-los na previsão do consumo de gasolina automotiva. Dessa forma, foram utilizados métodos autoprojativos, selecionando o melhor através de uma metodologia de competição.

Como embasamento deste trabalho monográfico, será analisado o comportamento do mercado de combustível, mais especificamente o da gasolina automotiva. Com uma introdução ao contexto histórico e seus impactos para economia, também foram conduzidas análises comparativas dos principais indicadores de produção e consumo, assim como projeções para o setor.

Além desta Introdução, o trabalho está estruturado em mais cinco capítulos. No capítulo 2 são apresentadas algumas considerações sobre o mercado de combustíveis, com foco no consumo de gasolina. No capítulo 3 são apresentados alguns trabalhos pesquisados e que fizeram uso de análise de séries de tempo na projeção do consumo de combustíveis. Já no capítulo 4 é apresentada a metodologia aplicada no presente trabalho monográfico. No capítulo 5 pode ser vista a análise de dados e principais resultados. No capítulo 6 são apresentadas as conclusões.

2 MERCADO DE COMBUSTÍVEL

Para o mercado de combustíveis no Brasil, a criação da Petrobras em 1953 consolidou o monopólio estatal sobre todo esse setor, desde a exploração do petróleo e derivados, até a distribuição. Com a descoberta de grandes reservas naturais na década de 1970, como a Bacia de Campos, e futuramente o pré sal, o Brasil tornou-se uma grande potência internacional para esse mercado (PETROBRAS, 2024).

Com a Lei nº 9.478, mais conhecida como a Lei do Petróleo de 1997, teve-se a criação da Agência Nacional do Petróleo. Logo, ocorreu a quebra do então monopólio estatal, e a chegada de empresas privadas. Dessa forma, ocorreram grandes mudanças, principalmente quanto a eficiência do setor, uma vez que a ANP é responsável por controlar a qualidade, gerando análises e relatórios úteis para tomadas de decisão, além de fiscalizar agentes e atribuir regulações (SOUZA; RODRIGUES; FURLAN, 2021).

A gasolina e outros derivados de petróleo possuem grande importância na economia do país, assim como na matriz energética. É observado na tabela 1 abaixo, que a produção desses derivados energéticos sofreu um aumento de aproximadamente 5% de 2022 para 2023.

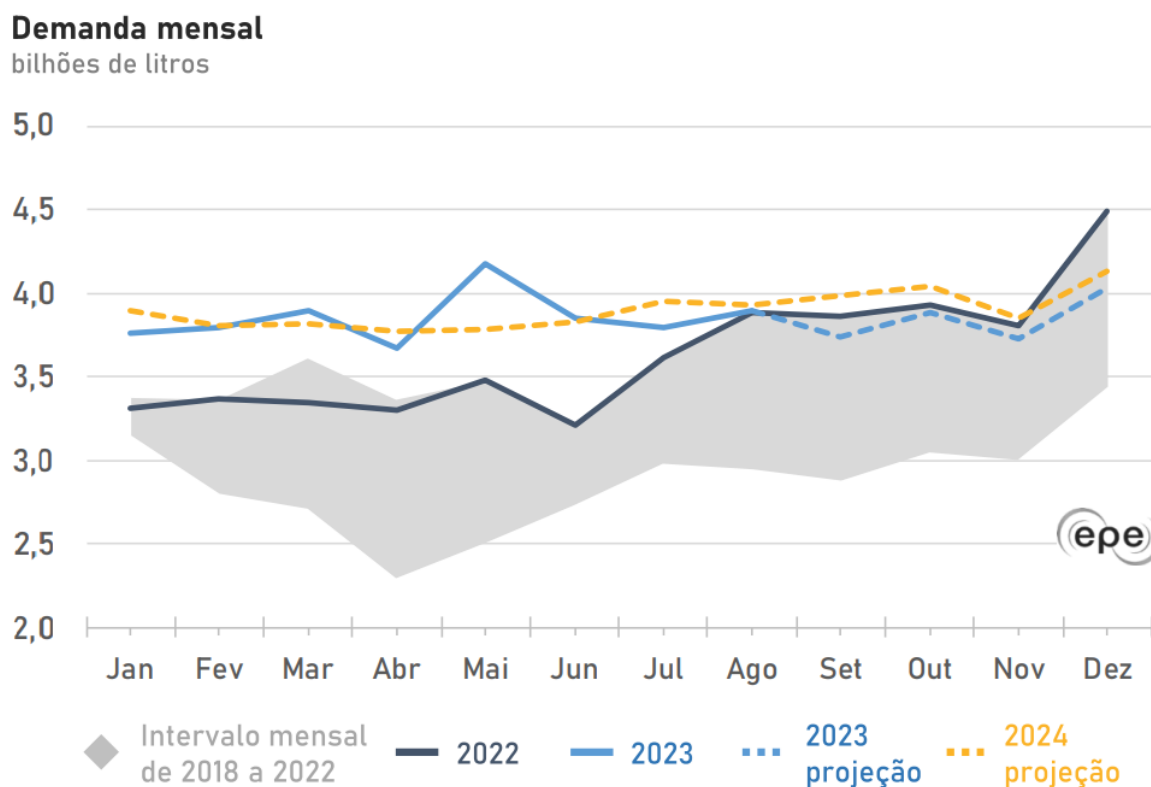
Tabela 1 – Produção de derivados de petróleo, energéticos e não energéticos – 2012 – 2023.

| DERIVADOS DE PETRÓLEO - ENERGÉTICOS | PRODUÇÃO (M ³) | | | | | | | | | | | | % / TOTAL |
|-------------------------------------|----------------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-----------|
| | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | |
| TOTAL DO ANO | 117.293.539 | 123.905.342 | 126.465.942 | 118.450.734 | 110.874.490 | 106.015.577 | 104.402.386 | 104.486.653 | 107.815.740 | 110.882.766 | 118.921.752 | 124.152.050 | 100% |
| ENERGÉTICOS | 99.656.897 | 106.771.525 | 109.038.517 | 103.051.898 | 96.691.753 | 92.386.705 | 90.383.727 | 90.165.945 | 91.843.286 | 96.572.388 | 102.843.340 | 107.895.479 | 86% |
| Gasolina A | 26.230.820 | 28.613.373 | 28.871.886 | 25.728.183 | 26.514.231 | 26.327.610 | 23.811.370 | 23.953.936 | 21.705.323 | 25.402.865 | 26.343.300 | 28.530.447 | 23% |
| Gasolina de Aviação | 77.606 | 93.685 | 93.762 | 72.486 | 53.902 | 59.662 | 46.220 | - | 2.216 | 55.899 | 54.651 | 39.130 | 0% |
| GLP | 8.320.384 | 7.946.127 | 7.578.584 | 7.426.409 | 7.329.727 | 7.445.175 | 7.435.631 | 7.275.015 | 7.343.999 | 7.228.154 | 7.687.672 | 8.137.146 | 7% |
| Óleo Combustível | 13.691.084 | 14.761.276 | 16.267.891 | 14.339.295 | 11.506.738 | 11.691.291 | 10.751.048 | 11.866.979 | 17.238.550 | 16.952.832 | 18.365.493 | 18.466.159 | 13% |
| Óleo Diesel | 45.504.004 | 49.539.186 | 49.675.057 | 49.457.609 | 45.369.807 | 40.625.830 | 41.957.390 | 40.998.507 | 42.215.122 | 42.852.980 | 45.529.431 | 47.277.211 | 39% |
| QAV | 5.422.769 | 5.554.391 | 6.079.114 | 5.656.859 | 5.789.278 | 6.168.600 | 6.376.333 | 6.066.674 | 3.332.770 | 4.074.103 | 4.854.283 | 5.438.445 | 5% |
| Querosene iluminante | 23.885 | 15.393 | 12.005 | 7.396 | 7.668 | 5.830 | 5.736 | 4.834 | 5.306 | 5.554 | 8.511 | 6.941 | 0% |
| Outros | 386.345 | 248.094 | 460.217 | 363.660 | 120.403 | 62.709 | - | - | - | - | - | - | 0% |
| NÃO ENERGÉTICOS | 17.636.642 | 17.133.817 | 17.427.426 | 15.398.837 | 14.182.737 | 13.628.871 | 14.018.659 | 14.320.708 | 15.972.454 | 14.310.378 | 16.078.412 | 16.256.571 | 14% |
| Asfalto | 2.569.635 | 2.653.348 | 3.248.853 | 2.015.366 | 2.152.075 | 1.955.427 | 1.899.816 | 1.721.537 | 2.318.275 | 1.857.380 | 2.399.521 | 2.660.040 | 2% |
| Coque | 4.452.350 | 4.810.510 | 4.748.864 | 4.958.620 | 5.076.586 | 4.928.529 | 4.468.572 | 4.411.889 | 4.332.411 | 4.367.408 | 4.502.608 | 4.558.219 | 4% |
| Lubrificante | 607.979 | 689.214 | 682.053 | 640.490 | 616.529 | 593.536 | 602.881 | 566.063 | 454.679 | 621.662 | 576.438 | 513.434 | 1% |
| Nafta | 6.405.849 | 5.354.014 | 5.072.548 | 4.608.816 | 3.175.436 | 3.086.305 | 4.046.483 | 4.480.133 | 6.194.691 | 4.623.048 | 5.809.086 | 5.671.965 | 4% |
| Parafina | 123.445 | 122.647 | 134.636 | 136.934 | 162.366 | 120.051 | 126.197 | 124.131 | 73.600 | 57.115 | 84.136 | 82.462 | 0% |
| Solvente | 285.547 | 449.080 | 378.620 | 354.022 | 331.453 | 321.954 | 324.721 | 307.653 | 313.185 | 407.962 | 371.938 | 463.301 | 0% |
| Outros | 3.191.837 | 3.055.004 | 3.161.852 | 2.684.589 | 2.668.293 | 2.623.069 | 2.549.988 | 2.709.302 | 2.285.612 | 2.375.803 | 2.334.684 | 2.307.151 | 2% |

Fonte: ANP (2024)

Dados da ANP (2024), trazem informações que a gasolina representa cerca de 23% da produção de combustíveis. Apenas de 2022 para 2023, a produção teve um aumento 8,3%. O aumento nos últimos anos, é muito devido ao crescimento da frota de veículos. Em 2023, o consumo de gasolina C, que é a mistura de gasolina A (pura) com etanol, sendo também a mais comum em circulação, atingiu 36 bilhões de litros, um aumento de 3,5% em comparação ao ano anterior. Esse crescimento é natural, uma vez que o país acaba de sair de uma recessão econômica causada pela pandemia, um período em que a população teve menos uso de transportes em geral. Mesmo com o aumento na comercialização de veículos flex, e com a prática de políticas de incentivos aos biocombustíveis, a gasolina manteve-se como insumo de maior preferência para veículos leves.

Figura 1 – Demanda mensal de Gasolina C.



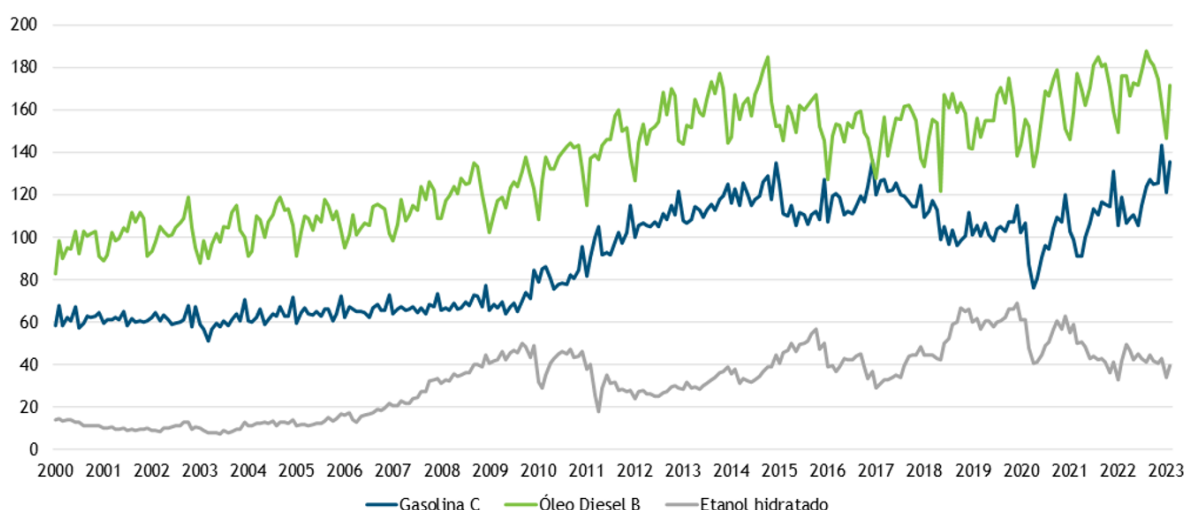
Fonte: EPE (2023a)

De acordo com a Empresa de Pesquisa Energética (EPE), a figura 1 evidencia que a demanda por gasolina C no Brasil sofreu um aumento de 11% ao comparar os últimos 12 meses (Setembro de 2022 a Agosto de 2023 vs. Setembro de 2021 a

Agosto de 2022). A pesquisa prevê um aumento de 1,3% para a demanda no ano de 2024, alcançando 46,8 bilhões de litros. (EPE, 2023).

É observado que a demanda por esse combustível tem um aumento significativo no mês de dezembro. Isso ocorre devido ao aumento das viagens e festividades de fim de ano, resultando em um maior uso dos veículos. Outro fator é o início das férias de muitos trabalhadores, assim como o recesso escolar. Segundo o IPEADATA, o consumo aparente de gasolina em dezembro de 2023 chegou a 610 mil barris, enquanto que em janeiro do mesmo ano o valor foi 557 mil barris, representando uma diferença de 9,5% entre os meses. (IPEA, 2024)

Figura 2 – Vendas de combustíveis líquidos – mil m³/dia



Fonte: Elaboração IBP, dados ANP.

Observando a figura 2 acima, tem-se a comparação entre as vendas dos derivados de petróleo energéticos mais comercializados no país. A variável “Vendas” pode ser utilizada como *proxy* para o consumo de combustíveis líquidos, pois reflete a quantidade de combustíveis adquirida pelos consumidores, sendo um indicativo confiável de consumo.

Segundo os dados da ANP (2024), no período de 2000 a 2022, as vendas de etanol hidratado cresceram 238%. Já a gasolina cresceu em 91%, seguida do óleo diesel (80%). Com a pandemia da Covid-19 causando impactos para as vendas de combustíveis em 2020, a gasolina mostrou maior estabilidade em comparação ao etanol, com uma queda de vendas em torno de 6% em relação a 2019, enquanto o de

etanol 14,5%. Os dados reforçam a estabilidade da gasolina como preferência do consumidor para veículos de passeio, que são atraídos pela estabilidade do produto no mercado, além da ampla rede de abastecimento e qualidade garantida.

Tendo em vista como a produção e o consumo de gasolina possuem uma grande importância para a economia do país, no próximo capítulo serão abordados alguns trabalhos que dedicaram ao estudo deste combustível ao longo dos anos, dando ênfase a trabalhos que elaboraram modelos eficientes para projeção de dados sobre o setor.

3 REVISÃO DE LITERATURA

Este capítulo abordará uma revisão da literatura que busca trazer estudos com diferentes abordagens sobre a demanda, consumo e a produção de derivados do petróleo, com foco na gasolina.

Zanini (2000) desenvolve e analisa um modelo para previsão da demanda de gasolina para o curto prazo no Brasil. A abordagem utilizada foi estratégia *bottom-up*. O primeiro passo foi gerar o modelo linear de Box & Jenkins. Em seguida foi gerado um modelo de regressão dinâmica, em que variáveis exógenas foram introduzidas, junto a um modelo de redes neurais que demonstrou melhoras significativas no desempenho preditivo da pesquisa.

Santiago (2009) desenvolve e analisa um modelo econométrico de séries temporais com um modelo de insumo-produto híbrido, com intuito de prever o cenário a longo prazo da demanda combustíveis no Brasil. Utilizou-se de séries históricas com dados ano a ano, assim como PIB, investimento, gastos do governo, consumo das famílias, importações e exportações. Dessa forma, verificou que as séries temporais além de não estacionárias, também são cointegradas. Feita as correções, aplicou um teste para comparar os modelos e suas capacidades preditivas, chegando à conclusão que o modelo com variáveis em log apresentou melhor desempenho.

Brafman (2009) elaborou um modelo para prever o consumo de ciclo de otto (gasolina, álcool e GNV) no Brasil, competindo seu modelo com um modelo MBJ ARMA. Com intuito de descobrir qual modelo seria o melhor em relação a venda de combustíveis. Escolhendo o método de regressão linear múltipla, não obteve êxito devido a uma correlação de duas variáveis, sendo uma exógena. Dessa forma, optou por uma regressão simples que foi capaz de ter um desempenho preditivo superior ao modelo MBJ ARMA.

Samohyl e Dantas (1998) estuda a demanda de gasolina no Brasil no período de 1955 a 1995, e apresenta um modelo com cointegração e fazendo uso das variáveis consumo de petróleo, o preço do barril em dólares e o Produto Interno Bruto. O estudo em questão visa calcular tanto as elasticidades preço quanto renda, além de criar cenários para as possíveis situações. Sendo assim, foi evidenciado pelo modelo, que se trata de uma demanda inelástica, em que a elasticidade é inferior a 1,

o que indica que variações no preço e na renda não influenciam de maneira abrupta o consumo.

A previsão do consumo de petróleo é um tema estudado por diversos países, principalmente por aqueles que estão no topo da rede de exportação global desse insumo. Sob esta luz, o trabalho de Keikha e Ziaei (2019) traz um modelo híbrido que utiliza Redes Neurais Artificiais (RNA) e Algoritmo Genético (AG) para prever o consumo de petróleo no Iran. Evitando métodos típicos como ARIMA ou regressão linear, os resultados obtidos tiveram um baixo erro médio percentual, o que elimina incertezas para os tomadores de decisão de diversos setores.

Fernandes e e Braga (2013), seu trabalho tem como principal objetivo analisar o grau de poder de mercado no setor de distribuição C, no Brasil, contemplando o período de 2002 a 2008. Sendo assim, foi baseado na Teoria da Nova Organização Industrial Empírica (NOIE), dando ênfase ao modelo proposto por Bresnahan (1982) e Lau (1982).

Apresentadas as leituras dos trabalhos auxiliares, o próximo capítulo será dedicado às metodologias que serão abordadas neste trabalho monográfico. Sendo assim, os dois métodos selecionados trazem diferentes abordagens preditivas, sendo necessário uma competição entre os modelos, para que seja selecionado o que trará os resultados mais fidedignos à realidade do consumo de gasolina no país.

4 METODOLOGIA

Observa-se que toda esta seção é baseada nos conceitos metodológicos do trabalho de Zanini (2023). Neste projeto, serão utilizados os métodos estatísticos de Amortecimento Exponencial (MAE) e de Box & Jenkins (MBJ) para selecionar o melhor modelo, que será utilizado para calcular as previsões de consumo de gasolina no Brasil.

4.1 Séries Temporais e Processos Estocásticos

Uma série temporal é um conjunto de observações de uma dada variável, ordenadas segundo o parâmetro tempo, geralmente em intervalos equidistantes. Se Z_t representa o valor da variável aleatória Z no instante t , denota-se a série temporal por Z_1, Z_2, \dots, Z_N onde N é o tamanho da série ou número de observações seriais da variável. Para séries temporais discretas as periodizações de coleta de dados usuais são dados diários, semanais, mensais e anuais.

Um processo estocástico caracteriza-se por uma sequência no tempo de variáveis aleatórias, ou seja, um modelo que descreve a estrutura de probabilidade de uma sequência de observações é chamado de processo estocástico. Desta forma, processos estocásticos são sistemas que evoluem no tempo e/ou no espaço de acordo com leis probabilísticas.

Um processo estocástico é uma família $Z=\{Z(t), t \in N\}$ tal que para cada $t \in R$, $Z(t)$ é uma variável aleatória. Se $N \equiv Z=\{1, \dots, t\}$, diz-se que o processo é de parâmetro discreto, denotando por Z_t . Se $N \in R$, diz-se que o processo é de parâmetro contínuo, denotando-se por $Z(t)$.

Segundo Souza & Camargo (1996), um processo estocástico está estatisticamente determinado quando se conhece suas funções de distribuição até a n -ésima ordem. Na prática, ocorrem situações problemáticas por não se conhecer todas as funções de distribuição até a n -ésima ordem e por ter-se apenas uma realização do processo estocástico, a partir da qual deseja-se inferir características do mecanismo gerador da série. Para superar essas dificuldades assumem-se duas restrições: Estacionaridade e Ergodicidade.

Se o processo estocástico que gerou a série de observações é invariante no tempo, diz-se que o processo é estacionário. Embora, a teoria mostre conveniência prática do uso de séries estacionárias, no mundo real, nem todas as séries encontradas podem ser classificadas como estacionárias. Logo, a estacionariedade é uma condição bastante restritiva imposta à série temporal. O conceito de processo não estacionário homogêneo surge quando se consegue um processo estacionário através de diferenças sucessivas do processo. Quanto à ergodicidade, um processo estocástico é dito ergódico se apenas uma realização é suficiente para obter todas as suas estatísticas. Logo, todo processo ergódico é também estacionário, já que apenas uma realização de um processo não estacionário não poderá conter todas as informações necessárias para a especificação do processo.

O objetivo deste trabalho é estruturar e analisar métodos de projeção de séries temporais com intuito de alcançar o modelo estocástico existente. Os modelos que serão apresentados são autoprojetivos (amortecimento exponencial e univariado de Box & Jenkins), que se baseiam de valores passados da amostra, e serão apresentados nos tópicos seguintes.

4.2 Método de Amortecimento Exponencial (MAE)

Conforme pôde ser visto na seção 1, uma série temporal pode ser definida como um conjunto de observações de uma dada variável, ordenadas segundo o parâmetro de tempo, geralmente em intervalos equidistantes, e que apresentam uma “dependência serial” (correlação) entre eles. O objetivo da análise, então, é estimar uma equação matemática que expresse a correlação dos dados históricos de forma que se possa projetá-los para um horizonte futuro.

Feitas estas considerações, salienta-se que uma breve descrição do método de amortecimento exponencial será feita a partir deste momento. Para este fim, imagine-se que o conjunto de observações Z_1, Z_2, \dots, Z_T seja uma série temporal de tamanho “T”.

Suponha-se agora que esta série represente um produto de determinada empresa cuja demanda mensal não apresenta uma variação significativa no seu nível ao longo do tempo, ou seja, não ocorrem mudanças no nível de venda com o tempo

ou, se ocorre, são variações pouco significativas. Então, para este produto, a equação de previsão pode ser representada por:

$$Z_t = a(T) + \varepsilon_t \quad (4.2.1)$$

Onde:

Z_t = venda no período t ($t = 1, 2, \dots, T$);

$a(T)$ = parâmetro representativo do nível médio das vendas no instante T ;

ε_t = erro de previsão e $\varepsilon_t \sim N(0, \sigma^2)$.

É importante observar que, dada as características deste produto (nível de venda mais ou menos constante), a estimativa ($\hat{a}(T)$) para o parâmetro “a” na equação (4.2.1) fornece exatamente uma previsão para o produto em questão. Sendo assim, salienta-se que este parâmetro pode ser estimado de diversas maneiras possíveis. Dentro de uma lógica “autoprojetiva”, poder-se-ia estimá-lo, por exemplo, através do modelo ingênuo (ou *naïve*) e que utiliza como previsor o último dado conforme indicado pela equação (4.2.2) abaixo:

$$\hat{Z}_T(\tau) = Z_T \quad (4.2.2)$$

Onde:

$\hat{Z}_T(\tau)$: previsão para Z_T , τ passos-à-frente (previsão de $Z_{T+\tau}$ feita no instante T);

Z_T : último dado disponível

τ : horizonte de previsão

Outras maneiras poderiam também ser utilizadas para estimar o parâmetro “a” na equação (4.2.1). Poderiam ser utilizadas uma média ou uma média móvel dos dados históricos (equações (4.2.3) e (4.2.4) respectivamente).

$$\hat{a}(T) = \bar{Z}_T = \frac{1}{T} \sum_{i=1}^T Z_i \quad (4.2.3)$$

$$\hat{a}(T) = MM(N) = M_T = \frac{Z_T + Z_{T-1} + Z_{T-N+1}}{N} \quad (4.2.4)^1$$

Onde:

MM(N): média móvel de tamanho N.

É importante observar que, tanto na média quanto na média móvel, existe uma desvantagem que diz respeito ao fato de que todos os dados entram com o mesmo “peso”, ou seja, $1/N$. O método de amortecimento exponencial vem suprir esta “deficiência”, ou seja, neste método, é possível dar pesos diferenciados de acordo com a “idade” da informação. A seguir é desenvolvida esta ideia.

O objetivo consiste em montar um “sistema” o qual possa reestimar os parâmetros do modelo a cada período de tempo incorporando a informação mais recente. Sabe-se que, ao final do período T, tem-se duas informações básicas disponíveis:

- 1 - A estimativa de “a” feita no final do período anterior $\hat{a}(T-1)$;
- 2 - O último dado disponível $Z(T)$;

Desta forma, o que se quer é utilizar estas informações para calcular uma estimativa atualizada do nível de venda: $\hat{a}(T)$. A proposta para a solução do sistema é fazer uma modificação na estimativa velha ($\hat{a}(T-1)$) do nível por uma fração do erro de previsão resultante do uso desta estimativa para o dado mais recente. Sabendo que o erro de previsão no último período pode ser definido como $\varepsilon(T) = Z(T) - \hat{a}(T-1)$, a formulação matemática para esta proposta está representada na equação abaixo:

$$\hat{a}(T) = \hat{a}(T-1) + \alpha * [Z(T) - \hat{a}(T-1)] \quad (4.2.5)$$

$$\hat{a}(T) = \alpha * Z(T) + (1 - \alpha) * \hat{a}(T-1) \quad (4.2.6)$$

Para simplificar a notação, define-se que $\hat{a}(T) \equiv S_T$. Desta forma, reescrevendo a equação (2.6) acima tem-se que:

$$S_T = \alpha * Z_T + (1 - \alpha) * S_{T-1} \quad (4.2.7)^2$$

¹ É fácil verificar que $M_T = M_{T-1} + \frac{Z_T - Z_{T-N}}{N}$.

² A equação (4.2.7) é também conhecida como Modelo de Brown.

Onde:

S_T = valor amortecido da série;

α = constante de amortecimento ou hiperparâmetro (número entre 0 e 1).

Ressalta-se que esta formulação (4.2.7) indica a idéia básica do método de amortecimento exponencial e indica que para se fazer uma atualização automática do parâmetro “a”, que representa o nível de vendas na equação (4.2.1), será feita uma combinação convexa³, onde é dado um peso α para o “presente” (último dado) e um peso $(1 - \alpha)$ para o “passado” (estimativa anterior para o nível e que pode ser obtida de várias formas como, por exemplo, uma média, uma média móvel, dentre outras)⁴

Em síntese, no método de amortecimento exponencial é possível dar pesos diferenciados para a “idade” da informação. Isto é, para séries mais “nervosas” (maior variância) pode se dar um peso maior para informações mais recentes, ao passo que para séries mais “comportadas” (menor variância) pode se ponderar de forma igual tanto dados presentes quanto dados passados. Uma extensão deste modelo pode ser feita quando se inclui parâmetros de tendência e sazonalidade para modelar o comportamento de determinada série temporal. Obviamente, são elaborados também procedimentos de atualização destes parâmetros mas sempre conservando a idéia de dar pesos diferenciados para “presente” e “passado”, ou seja, fazendo-se:

$$\alpha * \textit{Presente} + (1 - \alpha) * \textit{Passado}$$

Como dito anteriormente, a equação (4.2.1) pode ser utilizada para modelar o comportamento de uma série que apresente um comportamento mais ou menos constante, ou seja, sem grandes variações no nível. Entretanto, este modelo torna-se inadequado na presença de alterações do nível da série, ou seja, na presença de um componente de tendência.

³ A soma é igual a 1

⁴ Para mais detalhes, ver MONTGOMERY & JOHNSON (1990). Como um exemplo, imagine que se esteja trabalhando com dados mensais no período de janeiro de 2001 a dezembro de 2005. Neste caso, a equação (4.2.6) informa que a atualização do parâmetro de nível (portanto a previsão para o dado de janeiro de 2006, dado que o modelo para as vendas é constante) será feita dando-se um peso α para o dado de dezembro e um peso $(1 - \alpha)$ para a última estimativa feita para o dado de dezembro de 2005 (feita obviamente em novembro de 2005).

Para uma série que apresente oscilações no nível com o tempo, atenta-se que um modelo mais adequado é aquele representado na equação (4.2.8) a seguir⁵

$$Z_t = (a_1(T) + a_2(T) * t) + \varepsilon_t \quad (4.2.8)$$

Onde:

$a_1(T)$: parâmetro de nível no instante T;

$a_2(T)$: parâmetro de tendência no instante T;

t: variável tempo (t = 1, 2, ..., T sendo T é a quantidade de dados existentes);

ε_t é o erro de previsão e $\varepsilon_t \sim N(0, \sigma^2)$.

É importante salientar que um sistema de atualização paramétrica similar ao evidenciado na equação (4.2.7) será aplicado sobre os parâmetros da equação (4.2.8). Este modelo é conhecido como Modelo de Holt-2Parâmetros. A atualização dos parâmetros pode ser feita de acordo com as equações (4.2.9) e (4.2.10) a seguir:

$$\hat{a}_1(T) = \alpha * Z_T + (1 - \alpha) * \left[\hat{a}_1(T - 1) + \hat{a}_2(T - 1) \right] \quad (4.2.9)$$

$$\hat{a}_2(T) = \beta * \left[\hat{a}_1(T) - \hat{a}_1(T - 1) \right] + (1 - \beta) * \left[\hat{a}_2(T - 1) \right] \quad (4.2.10)$$

Pode-se observar que as equações acima contêm a mesma idéia de se ponderar “presente” e “passado” para se fazer a atualização dos parâmetros. Entretanto, vê-se que são usadas duas constantes de amortecimento (α e β), uma para o parâmetro de nível (a_1) e outra para o parâmetro de tendência (a_2).

Na equação (4.2.9), observa-se que a atualização do parâmetro de nível é feita dando um peso α para o dado real mais recente e um peso $(1 - \alpha)$ para a última estimativa feita para o nível que é composta por: $\hat{a}_1(T-1)$, estimativa feita para o nível no instante anterior (T-1), mais $\hat{a}_2(T-1)$, que é a estimativa feita para a tendência também no instante anterior (T-1). Ora, “nível mais tendência (taxa de crescimento)”

⁵ Verifique que, diferentemente da equação (4.2.1), existem agora dois parâmetros e por isto é feita a diferença entre a_1 e a_2 .

dá exatamente uma estimativa de um novo nível, sendo que (T-1) indica que este cálculo foi feito no instante anterior. Em resumo, na atualização do parâmetro de nível, dá-se um peso para o último dado (que fornece uma representação real e atualizada, portanto, “presente ou recente” para o nível) e um outro peso para um valor estimado para este nível quando se estava no momento anterior (portanto, “passado”).

Já na equação (4.2.10), atualização do parâmetro de tendência, vê-se que é dado um peso β para a diferença entre a nova estimativa do nível (calculada na equação 4.2.9) e a última estimativa do nível (feita no instante anterior T-1). Ora, variação de nível é exatamente o que caracteriza um componente de tendência ou taxa de crescimento. Se é dado um peso β para esta estimativa “presente” do parâmetro de tendência, é dado um peso $(1 - \beta)$ para a última estimativa da tendência feita no instante anterior (T-1).

Percebe-se que, como o objetivo proposto visa desenvolver um modelo autoprojeto ou univariado, é necessário que todos os “fatores” componentes de uma série sejam estimados ou “modelados”. Até o presente momento, foi possível abordar a estimativa de dois parâmetros (nível e tendência). Entretanto, imagine-se ainda que possa existir um certo comportamento periódico das vendas, ou seja, dependendo da época do ano, existe um incremento ou decréscimo nas vendas. Em séries temporais, este comportamento é o que se denomina exatamente por “sazonalidade”, ou seja, um movimento periódico (cíclico) da série no decorrer do tempo. O que se quer dizer é que a série de vendas do produto em questão pode ainda apresentar um comportamento sazonal, ou seja, uma “influência” provocada por determinados períodos do ano sobre seu nível (incluindo obviamente a variação deste nível). Neste caso, o modelo mais adequado pode⁶ ser o expresso na equação (4.2.11):

$$Z_t = (a_1(T) + a_2(T) * t) * \rho_t + \varepsilon_t \quad (4.2.11)$$

Onde:

$a_1(T)$: parâmetro de nível no instante T;

$a_2(T)$: parâmetro de tendência no instante T;

⁶ É dito “pode” pois a equação apresenta uma sazonalidade multiplicativa, entretanto, pode-se encontrar também uma sazonalidade aditiva.

t: variável tempo ($t = 1, 2, \dots, T$ sendo T é a quantidade de dados existentes);
 ρ_t : fator sazonal referente ao período t;
 ε_t é o erro de previsão e $\varepsilon_t \sim N(0, \sigma^2)$.

Este modelo é conhecido como Modelo de Holt-Winters. Pode-se observar que na equação (4.2.11) foi incluído um novo parâmetro para “captar” o efeito da sazonalidade sobre as vendas⁷. Atenta-se que um procedimento de atualização paramétrica similar ao expresso na equação (4.2.7) será utilizado para atualizar sequencialmente os parâmetros deste modelo. Este procedimento pode ser visualizado nas equações (4.2.12) a (4.2.14) a seguir:

$$\hat{a}_1(T) = \alpha * \frac{Z_T}{\hat{\rho}_t(T-1)} + (1 - \alpha) * \left[\hat{a}_1(T-1) + \hat{a}_2(T-1) \right] \quad (4.2.12)^8$$

$$\hat{a}_2(T) = \beta * \left[\hat{a}_1(T) - \hat{a}_1(T-1) \right] + (1 - \beta) * \left[\hat{a}_2(T-1) \right] \quad (4.2.13)^9$$

$$\hat{\rho}_t(T) = \gamma * \frac{Z_T}{\hat{a}_1(T)} + (1 - \gamma) * \left[\hat{\rho}_t(T-1) \right] \quad (4.2.14)$$

Pela equação (4.2.12) acima, vê-se que o nível estimado é função da última observação, do fator sazonal estimado no instante anterior (mesmo mês do ano anterior) e estimativas anteriores do nível e tendência. Vê-se que o termo $\frac{Z_T}{\hat{\rho}_t(T-1)}$ representa a observação no instante T dessazonalizada. Já pela equação (4.2.13), observa-se que a taxa de crescimento (tendência) estimada é função dos níveis estimados em T e T-1 e da taxa de crescimento estimada anteriormente¹⁰. Por fim,

⁷ Existem uma restrição de normalização que faz com que $\sum_{i=1}^L \rho_i = L$, onde L é o comprimento do período

sazonal. Isto é, caso se esteja trabalhando com dados mensais e um parâmetro de sazonalidade multiplicativo, a soma dos fatores sazonais precisa ser igual a 12 (obviamente pois existem 12 meses em um ano).

⁸ Onde $\hat{\rho}_t$ é o fator sazonal correspondente ao período (mês, trimestre, etc) t.

⁹ Trata-se da mesma equação (4.2.10).

¹⁰ Como visto anteriormente, a constante de amortecimento (β) é diferente daquela (α) usada para atualização do nível.

pela equação (4.2.14), vê-se novamente a ideia básica do método de amortecimento exponencial de atualizar os parâmetros do modelo atribuindo pesos diferenciados à “idade da informação”, ou seja, ponderando “presente” e “passado” de forma diferenciada. Esta equação indica que o fator sazonal correspondente ao período T é função do fator sazonal correspondente ao mesmo período no ano anterior e também da última observação. Logo, observa-se que o fator sazonal correspondente a um certo “mês”, por exemplo, só é atualizado uma vez por ano, ao se receber o dado referente àquele “mês”. Pode-se observar também que existem três constantes de amortecimento (α , β e γ), uma para cada parâmetro (nível, tendência e sazonalidade) a ser atualizado.

Em relação às constantes de amortecimento, é importante ressaltar que existem procedimentos de otimização implementados nos *softwares* que permitem a determinação dos valores destas. Importante registrar, também, que existe apenas um valor para cada uma das constantes de amortecimento e este valor é encontrado utilizando o histórico de dados sob análise. Portanto, a constante de amortecimento pode ser definida como uma quantidade fixa que é utilizada para fazer a atualização sequencial dos parâmetros. Este conceito caracteriza o modelo obtido através do método de amortecimento exponencial como um modelo com “validade local”, ou seja, a cada instante de tempo, a cada dado real que chega, é feita uma atualização dos parâmetros do modelo com base nas equações descritas anteriormente, sendo que os fatores responsáveis por esta atualização são exatamente as constantes de amortecimento ou hiperparâmetros.

Em determinadas situações, faz-se mister intervir nos valores projetados por um modelo do tipo Holt (equação 4.2.8). No modelo de Holt, vê-se que são estimados dois parâmetros, um para o nível e outro para a tendência. Dependendo da “magnitude” do parâmetro de tendência, pode-se gerar previsões “explosivas” (muito altas) no horizonte de previsão. Uma maneira de se corrigir este problema, é realizar o que se denomina de *damped trend*. Neste procedimento, inclui-se mais um hiperparâmetro no modelo conforme a equação 4.2.15 a seguir. Esta equação já representa a equação de previsão. Vê-se que a função deste hiperparâmetro é reduzir a tendência no horizonte de previsão.

$$Z_t = \hat{a}_1(T) + \sum_{j=1}^{\tau} \phi^{j-1} * \hat{a}_2(T) * \tau + \varepsilon_t \quad (4.2.15)$$

É importante entender que, na seção anterior, foram apresentados os modelos de amortecimento exponencial e as conseguintes equações de atualização paramétrica. Desta forma, é importante que se tenha em mente a forma da equação que gerará as previsões. Para cada um dos casos (vendas constantes, vendas com tendência e vendas com tendência e sazonalidade), a equação de previsão é apresentada a seguir:

$$\hat{Z}_t(\tau) = \hat{a}_1(T) + \varepsilon_t \quad (4.2.16)$$

$$\hat{Z}_t(\tau) = \hat{a}_1(T) + \hat{a}_2(T) * \tau + \varepsilon_t \quad (4.2.17)^{11}$$

$$\hat{Z}_t(\tau) = \left(\hat{a}_1(T) + \hat{a}_2(T) * \tau \right) * \hat{\rho}_{p(T+\tau)}^{(T)} + \varepsilon_t \quad (4.2.18)^{12}$$

Onde:

$\hat{a}_1(T)$ = estimativa do parâmetro de nível atualizado no instante T

$\hat{a}_2(T)$ = estimativa do parâmetro de tendência atualizado no instante T

$\hat{\rho}_{p(T+\tau)}^{(T)}$ = estimativa do parâmetro de sazonalidade referente ao mês T+ τ , atualizado até o instante T.

τ = horizonte de previsão

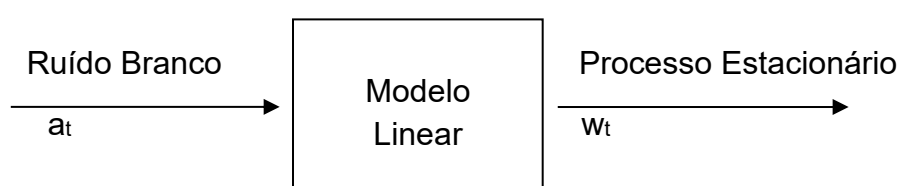
O procedimento de *damped trend* pode ser também aplicado ao modelo de Holt-Winters. Portanto, a introdução do parâmetro ϕ (equação 4.2.15) pode ser também feita na equação 4.2.18.

¹¹ Notação para quando há deslocamento de origem. Caso contrário faz-se (T+ τ).

¹² Idem.

4.3 Método Univariado de Box & Jenkins (MBJ)

O fundamento teórico do modelo de Box & Jenkins baseia-se na Teoria Geral de Sistemas Lineares que diz que a passagem de um ruído branco por um filtro linear de memória infinita gera um processo estacionário de segunda ordem¹³. Isto é, pela Teoria Geral de Sistemas:



Definindo um operador de atraso “B” (*backward shift operator*)

$$B^k Z_t = Z_{t-k}$$

Tem-se que:

$$w_t = a_t - \Psi_1 a_t B - \Psi_2 a_t B^2 - \dots$$

$$w_t = (1 - \Psi_1 B - \Psi_2 B^2 - \dots) a_t = \Psi(B) a_t$$

$$w_t = \Psi(B) a_t \Leftrightarrow a_t = \Psi(B)^{-1} w_t$$

Sendo que:

$$\Psi(B)^{-1} = \pi(B)$$

$$\text{onde } \pi(B) = 1 - \pi_1 B - \pi_2 B^2 \dots$$

Como $\Psi(B)$ [ou $\pi(B)$] possuem infinitos parâmetros, o que causa um problema, Box & Jenkins mencionam que, sob certas restrições, pode-se afirmar que todo polinômio pode ser aproximado por uma combinação finita de polinômios. Então:

$$\Psi(B) = \theta(B) / \phi(B), \text{ onde:}$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \Rightarrow \text{Polinômio MA (q)}$$

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \Rightarrow \text{Polinômio AR (p)}$$

¹³ Em termos gerais, processo estacionário de 2ª ordem é um processo estocástico com média e variância constantes.

Desta forma, surgem os denominados Modelos ARMA (p,q) que assumem a seguinte forma:

$$\phi(B)w_t = \theta(B)a_t \quad (4.3.1)$$

Entretanto, para se modelar séries não estacionárias na média, produz-se a estacionariedade através da diferenciação da série original, ou seja, produz-se uma série não estacionária homogênea¹⁴. Desta forma, se Z_t é uma série não estacionária, procurar-se-á transformá-la na série X_t não estacionária homogênea. Então:

$$X_t = Z_t - Z_{t-1} = Z_t - BZ_t = (1 - B) Z_t = \nabla Z_t$$

Onde: $\nabla = (1 - B) \Rightarrow$ operador de diferença

Isto significa que, aplica-se tantas diferenças quantas forem necessárias para produzir estacionariedade (na média) da série resultante. Genericamente, uma série w_t estacionária é obtida pela aplicação de “d” ($d = 0, 1, 2, \dots$)¹⁵ diferenças na série original: $w_t = \nabla^d Z_t$. Surgem então os chamados modelos ARIMA (p,d,q) que assumem a seguinte forma:

$$\phi(B)\nabla^d Z_t = \theta(B)a_t \quad (4.3.2)$$

A filosofia da modelagem Box & Jenkins fundamenta-se em duas ideias básicas: o princípio da parcimônia (escolher um modelo com o menor número de parâmetros possíveis para uma representação matemática adequada) e a construção de modelos através de um ciclo iterativo (estratégia de seleção de modelos até a obtenção de um modelo satisfatório). Desta forma, atenta-se que a modelagem através da metodologia de Box & Jenkins, abrange várias etapas de análise indo desde a identificação da estrutura do modelo, passando pela estimação paramétrica e por fim, fazendo vários testes de validação do modelo.

O primeiro passo fundamental da metodologia consiste em identificar a ordem de homogeneidade “d”, ou seja, caso seja necessário, trata-se de identificar o número de vezes que a série original deve ser diferenciada para se tornar uma série

¹⁴ A não estacionariedade homogênea exclui os processos de comportamento explosivos e/ou altamente não-lineares.

¹⁵ Na prática, normalmente o grau máximo de diferenciação é 2.

estacionária. Este procedimento pode ser feito através da observação do próprio gráfico da série ou da função de autocorrelação (FAC)¹⁶.

O passo seguinte na metodologia é a identificação do modelo, ou seja, da sua ordem (identificação de p e q). Para isso são utilizados os conceitos de função de autocorrelação (FAC) e autocorrelação parcial (FACP), ou seja, utiliza-se os correlogramas definidos na seção 1. De modo geral, para a identificação da ordem, observa-se os comportamentos da FAC e da FACP. No quadro 1 abaixo é feito um resumo das características destas funções para os modelos AR(p), MA(q) e ARMA (p,q).

Quadro 1 – Resumo das características teóricas da FAC e da FACP dos modelos AR(p), MA(q) e ARMA (p,q)

| Modelo | Função de Autocorrelação (ρ_k) | Função de Autocorrelação Parcial (ϕ_{kk}) |
|----------------|---|---|
| AR(p) | Infinita (Exponencial e/ou senóides amortecidas) | Finita (Corte após o <i>lag</i> “ p ”) |
| MA (q) | Finita (Corte após o <i>lag</i> “ q ”) | Infinita (Exponencial e/ou senóides amortecidas) |
| ARMA (p,q) | Infinita (Exponencial e/ou senóides amortecidas após o <i>lag</i> “ $q-p$ ”) | Infinita (Exponencial e/ou senóides amortecidas após o <i>lag</i> “ $p-q$ ”) |

Fonte: Zanini, 2023.

Através da máxima verossimilhança, após a ordem do modelo ser identificada, é necessário estimar os parâmetros. Logo, é utilizado testes para aderência como de resíduos, sobrefixação entre outros, para verificar a adequação final do modelo.

Em caso da ocorrência de sazonalidades, tem-se o modelo SARIMA(p,d,q)x(P,D,Q), seguindo as mesmas orientações para obtenção de um modelo ARIMA não sazonal.

Nesse caso a modelagem segue a equação:

¹⁶ A Função de Autocorrelação para uma série não estacionária apresenta um lento decréscimo.

$$\phi(B)\Phi(B^S)\nabla_S^D\nabla^d Z_t = \theta(B)\Theta(B^S)a_t \quad (4.3.3)$$

onde:

- $\phi(B)$: operador não sazonal auto-regressivo
- ϕ_i : parâmetros auto-regressivo não-sazonais
- $\nabla^d = (1-B)^d$: operador diferença não sazonal de ordem d
- $\Phi(B^s)$: operador sazonal auto-regressivo
- Φ_i : parâmetros auto-regressivo sazonais
- $\nabla^{D_s} = (1-B^s)^D$: operador diferença sazonal de ordem D
- $\theta(B)$: operador não sazonal de médias móveis
- θ_i : parâmetros de médias móveis não sazonais
- $\Theta(B^s)$: operador sazonal de médias móveis
- Θ_i : parâmetros de médias móveis sazonais

4.4 Competição de Métodos

Para seleção de um melhor método de previsão, pode ser utilizado um procedimento de análise *in-sample* e *out-of-sample*. Isto significa que parte dos dados são selecionados (por exemplo os últimos 12 meses) para validar o poder de previsão dos modelos ajustados com os dados restantes (ou seja, avalia-se o poder de previsão dos modelos dentro e fora do período amostral utilizado).

O procedimento pode envolver ainda o que se denomina de *out-of-sample rolling evaluation* (análise recursiva fora da amostra), ou seja, com os mesmos parâmetros estimados para os dados passados, move-se a origem da previsão no período *out-of-sample*, fazendo-se previsões para cada origem. Caso sejam retirados os 12 últimos dados¹⁷, isto significa que serão feitas no período *out-of-sample* um total de 78 previsões (12 previsões para um passo (mês) à frente, 11 previsões para 2

¹⁷ Obviamente retirar os últimos dados é apenas um procedimento de análise para seleção de modelos. Para se obter as previsões desejadas, depois de feita análise *in-sample* e *out-of-sample*, os dados retirados são novamente incorporados e os parâmetros atualizados. Foi dado como exemplo a retirada dos últimos 12 dados o que geraria um total de 78 previsões fora da amostra. Entretanto, este “período de corte” pode ser definido de acordo com cada problema. Como exemplo, se são retirados os últimos 6 dados, geram-se 21 previsões fora da amostra.

passos à frente, 10 previsões para 3 passos à frente e assim sucessivamente). Pode-se definir como critério, por exemplo, que o método vencedor a ser selecionado será aquele que minimizar o MAD (*Mean Absolute Deviation*) acumulado ou erro médio absoluto acumulado fora da amostra. Para entender melhor o procedimento de rolling evaluation observar a figura 03 a seguir.

Figura 3 – Procedimento de *out-of-sample rolling evaluation*

| | | HORIZONTE | | | | | | | | | | | |
|--------|-----|-----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | | jan | fev | mar | abr | mai | jun | jul | ago | set | out | nov | dez |
| ORIGEM | dez | A | A | A | A | A | A | A | A | A | A | A | A |
| | Jan | F | F | F | F | F | F | F | F | F | F | F | F |
| | Fev | A | A | F | F | F | F | F | F | F | F | F | F |
| | mar | A | A | A | F | F | F | F | F | F | F | F | F |
| | abr | A | A | A | A | F | F | F | F | F | F | F | F |
| | mai | A | A | A | A | A | F | F | F | F | F | F | F |
| | Jun | A | A | A | A | A | A | F | F | F | F | F | F |
| | Jul | A | A | A | A | A | A | A | F | F | F | F | F |
| | ago | A | A | A | A | A | A | A | A | F | F | F | F |
| | set | A | A | A | A | A | A | A | A | A | F | F | F |
| | out | A | A | A | A | A | A | A | A | A | A | F | F |
| | nov | A | A | A | A | A | A | A | A | A | A | A | F |

Fonte Zanini (2023) Legenda: F = Forecast, A = Actual.

Obtidas as previsões out-of-sample, várias métricas de desempenho podem ser calculadas como, por exemplo, o MAPE e o MAD para cada horizonte de previsão e estas mesmas estatísticas acumuladas. Neste estudo, utiliza-se a métrica para 1 passo à frente, e no acumulado para todos até 12 passos à frente. Além destas estatísticas, pode-se também calcular o GMRAE (*Geometric Mean Relative Absolute Error*) que compara o erro do modelo selecionado com o erro do modelo ingênuo (que usa como previsão o último dado disponível). O GMRAE é a média geométrica da razão (quociente) entre o erro absoluto (portanto, em módulo) do modelo estimado e

o erro absoluto do método ingênuo. Portanto, é desejável que o GMRAE seja igual ou menor do que 1.

É importante ressaltar que, selecionado o melhor método de acordo com algum critério previamente estabelecido, incorpora-se novamente os dados retirados à amostra, atualiza-se os parâmetros e faz-se as projeções. Em síntese, a análise *in-sample* e *out-of-sample* é apenas um procedimento que visa definir o melhor dentre dois ou mais métodos de previsão.

4.4.1 Desempenho Preditivo

A partir da comparação dos valores reais e “ajustados”, várias métricas são calculadas para medir o desempenho do modelo. Sendo assim, pode-se observar a seguir algumas métricas que serão utilizadas na pesquisa.

1) MAPE (*Mean Absolute Percentual Erro*)

O MAPE (erro médio absoluto percentual) é calculado através da diferença entre valores estimados e reais e equivale às previsões um passo-à-frente (por exemplo, para o mês seguinte). Veja a equação (4.4.1) seguir:

$$MAPE = \frac{\sum_{t=1}^N \frac{|Y(t) - \hat{Y}(t)|}{Y(t)} \times 100}{N} \quad (4.4.1)$$

Onde:

$Y(t)$ = valor da série temporal no período (t) ;

$\hat{Y}(t)$ = valor ajustado da série temporal para o período (t);

N = total de dados utilizados (total de observações).

2) MAD (*Mean Absolute Deviation*)

O MAD (erro médio absoluto) é também calculado através da diferença entre valores estimados e reais para as previsões um passo-à-frente (equação (4.4.2) a seguir). Entretanto, ao contrário do MAPE, não pode ser lido em termos percentuais, mas sim na unidade de medida da variável sob interesse.

$$MAD = \frac{\sum_{t=1}^N |Y(t) - \hat{Y}(t)|}{N} \quad (4.4.2)$$

Onde:

$Y(t)$ = valor da série temporal no período (t)

$\hat{Y}(t)$ = valor ajustado da série temporal para o período (t)

N = total de dados utilizados (total de observações)

3) GMRAE (*Geometric Mean Relative Absolute Error*)

O GMRAE (*Geometric Mean Relative Absolute Error*) compara o erro do modelo selecionado com o erro do modelo ingênuo (que usa como previsão o último dado disponível). O GMRAE (equação 4.4.3) é a média geométrica da razão (quociente) entre o erro absoluto (portanto, em módulo) do modelo estimado e o erro absoluto do método ingênuo. Portanto, é desejável que o GMRAE seja igual ou menor do que 1.

$$GMRAE = \sqrt[N]{\prod_{t=1}^N \left(\frac{|Y(t) - \hat{Y}(t)|}{|Y(t) - Y(t-1)|} \right)} \quad (4.4.3)$$

Onde $Y(t)$ é o valor da série temporal no período (t);

$\hat{Y}(t)$ é a previsão da série temporal para o período (t);

$Y(t-1)$ é o valor da série temporal no período (t-1);

N é o número de observações.

4) Coeficiente de Explicação Ajustado (R^2 ajustado)

O coeficiente de explicação ajustado (R^2 ajustado) tem interpretação similar ao coeficiente de explicação, ou seja, indica o quanto da variação total dos dados é explicada pelo modelo. Entretanto, faz-se uma correção tendo em vista a quantidade de parâmetros no modelo. Veja a equação (4.4.4) a seguir:

$$R^2_{ajust} = \left(1 - \frac{\sum_{t=1}^N \frac{(Y(t) - \hat{Y}(t))^2}{N - k}}{\sum_{t=1}^N \frac{(Y(t) - \bar{Y})^2}{N - 1}} \right) \times 100 \quad (4.4.4)$$

Onde:

$Y(t)$ = valor da série temporal no período (t);

$\hat{Y}(t)$ = previsão da série temporal para o período (t);

\bar{Y} = média das observações (média da série temporal) ;

N = total de dados utilizados (total de observações);

k = número de parâmetros do modelo.

Já apresentada a metodologia de análise, no próximo capítulo podem ser vistas a análise dos dados e resultados do trabalho.

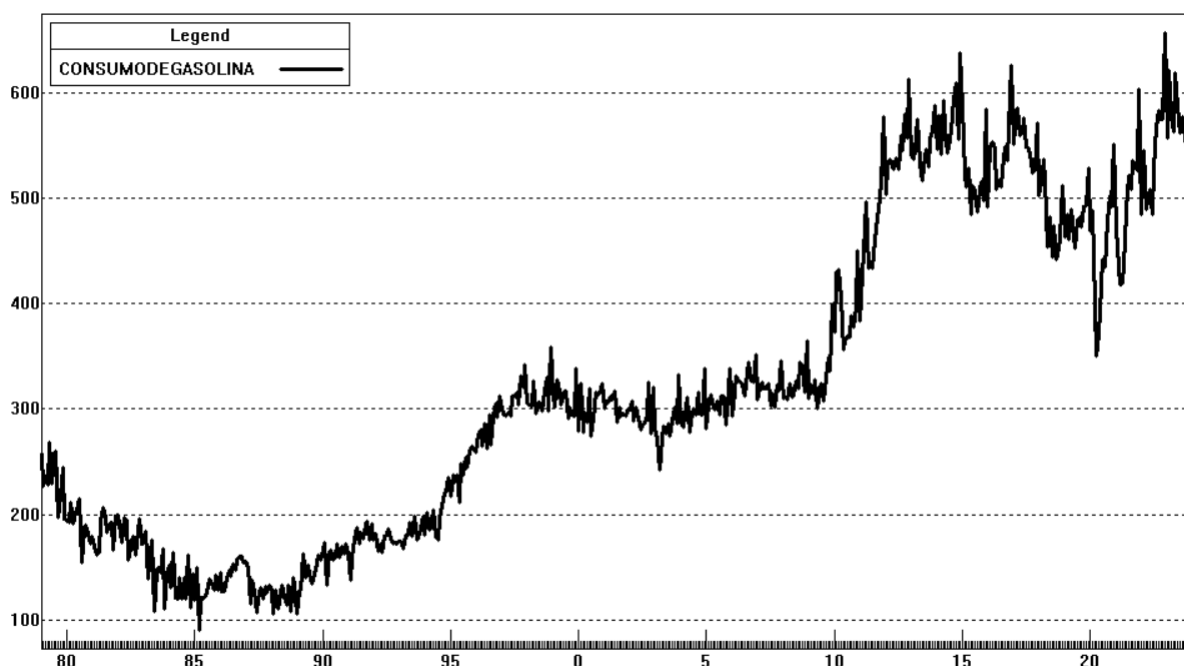
5 ANÁLISE DE DADOS E RESULTADOS

5.1 Base da dados

Para a previsão do modelo de consumo de gasolina no Brasil, os dados utilizados foram coletados do Instituto de Pesquisa Econômica Aplicada (IPEA)¹⁸. A previsão foi realizada através do *Software Forecast Pro for Windows* (FPW).

As informações representam o consumo aparente que corresponde à produção industrial doméstica acrescida das importações e diminuída das exportações. Sendo assim, os dados são mensais, e representam a média diária do consumo de gasolina, medidos em milhares de barris entre os períodos de janeiro de 1979 e março de 2024. Na figura 4 a seguir, é possível observar o gráfico com a evolução do consumo de gasolina no período analisado:

Figura 4 – Consumo de gasolina no Brasil (em milhares de barris).



Fonte: Elaboração própria.

A seguir, serão apresentados os resultados obtidos pelos métodos apresentados no capítulo 4 assim como a competição entre eles, com intuito de

¹⁸ Os dados primários foram publicados através da Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP) e disponibilizados no Instituto de Pesquisa Econômica Aplicada (IPEA), <https://www.ipeadata.gov.br/>.

determinar qual mais eficiente. O período de previsão do consumo de gasolina foi entre abril de 2024 a dezembro de 2025.

5.2 Equações de Previsão

Utilizando o MAE, é observado que o modelo que melhor explica a evolução do consumo de gasolina é o Modelo de Holt-Winters apresentado no capítulo 4 (equação 4.2.11). A seguir, tem-se os parâmetros estimados e os hiperparâmetros apresentados na tabela 2. Na tabela 3 estão os valores dos fatores sazonais estimados:

Tabela 2 – Parâmetro do Método de Amortecimento Exponencial

| Componente | Parâmetros | Hiperparâmetros |
|-------------------|-------------------|------------------------|
| Nível | 543,98 | 0,53975 |
| Tendência | 0,3934 | 0,00113 |
| Sazonalidade | * | 0,21477 |

Fonte: Elaboração própria. Nota: * Fatores sazonais apresentados na Tabela 3.

De acordo com o Modelo de Holt-Winters, com os hiperparâmetros gerados pela estimação, é possível observar o peso que é designado ao presente e ao passado para os parâmetros de tendência, nível e sazonalidade. Logo, os hiperparâmetros de nível observados na tabela 2 mostram que o peso atribuído para o presente (0,53975) é ligeiramente maior que o peso atribuído para o passado (0,46025). Já para a tendência e sazonalidade observa-se o inverso, ou seja, nos cálculos de atualização paramétrica é dado um peso menor para o presente do que para o passado.

Tabela 3 – Fatores Sazonais

| Meses | Fator | Meses | Fator |
|--------------|--------------|--------------|--------------|
| Janeiro | 0,97307 | Julho | 0,98347 |
| Fevereiro | 1,00808 | Agosto | 0,99426 |
| Março | 0,97300 | Setembro | 1,00707 |
| Abril | 0,97673 | Outubro | 1,01325 |
| Maió | 0,97288 | Novembro | 1,01110 |
| Junho | 0,97793 | Dezembro | 1,11757 |

Fonte: Elaboração própria.

Com base nos dados obtidos, a estrutura do modelo de Holt-Winters apresenta a seguinte equação:

$$Z_t = (\alpha_1 + \alpha_2 * \tau) * \rho_t$$

$$Z_t = (543,98 + 0,3934) * \rho_t \quad (5.2.1)$$

Onde:

Z_t = Quantidade de gasolina consumida em um certo período da amostra;

ρ_t = Parâmetro de sazonalidade sazonal;

α_1 = Parâmetro de nível;

α_2 = Parâmetro de tendência;

τ = variável “tempo”.

Sendo assim, tem-se que o nível atual do consumo de gasolina é de 543,98 mil barris. O nível do próximo período ($t + 1$) sofrerá com acréscimos ou decréscimos de acordo com os parâmetros sazonais estimados de cada mês (tabela 3).

Vale a pena ressaltar o parâmetro estimado no mês de dezembro, com um fator de 1,11757, que representa um acréscimo de 11,7% no consumo aparente de gasolina para esse mês. Isso ocorre devido a fatores sazonais específicos desse período do ano, como recessos e eventos festivos, já evidenciados no capítulo 2 sobre o Mercado de combustível.

Tabela 4 – Estatística de desempenho preditivo – MAE

| Componentes | Resultados |
|-------------------------|-------------------|
| R ² Ajustado | 98,23% |
| MAPE | 5,60% |

Fonte: Elaboração própria.

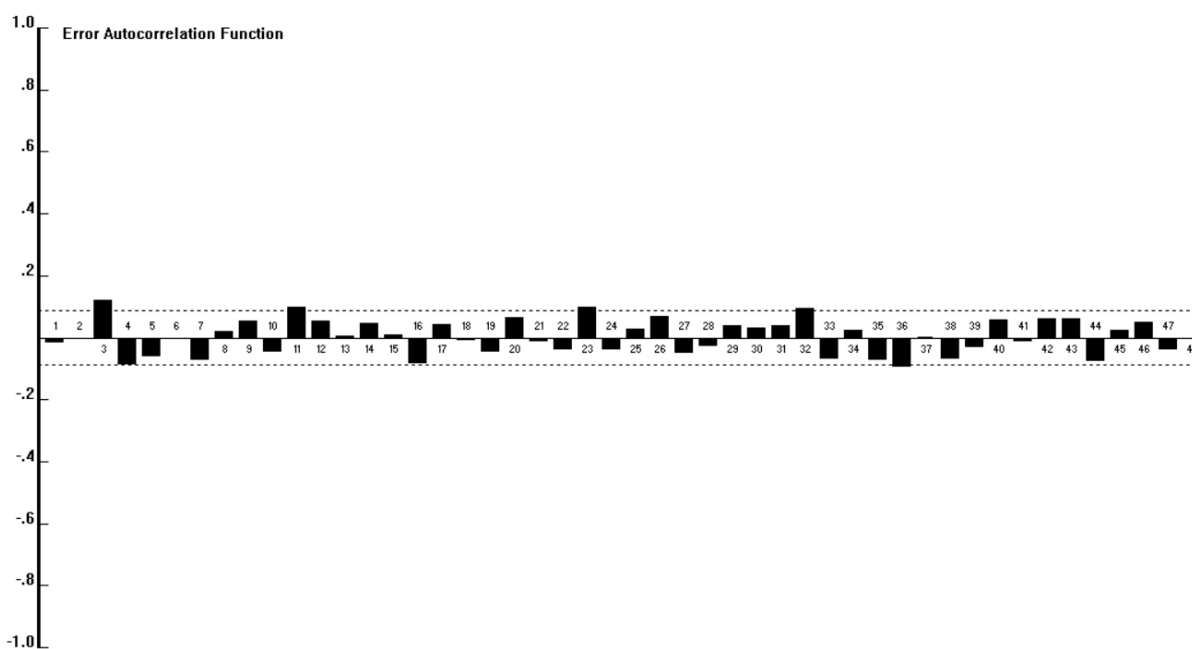
Neste estudo, foi avaliado o desempenho preditivo do modelo, considerando duas métricas principais: o R² ajustado, que avalia a capacidade do modelo em explicar a variação no consumo de gasolina, e o MAPE, que indica o erro médio absoluto percentual das previsões para o próximo período. Dessa forma, na tabela 4 vê-se que o modelo Holt-Winters demonstrou uma capacidade explicativa de 98,23% em relação ao consumo no período analisado, com uma margem de erro médio de 5,6% para cima ou para baixo ao projetar o próximo período.

A seguir, foi efetuado o teste de Ljung-Box e pode-se ver na figura 5 o gráfico ACF Erro, com intuito de verificar a fidedignidade do modelo¹⁹. Com o teste de Ljung-Box, testa-se a hipótese de que as "k" primeiras autocorrelações são nulas, isto é:

$$H_0: \rho_1 = \rho_2 = \dots = \rho_k = 0$$

Para a estatística amostral, caso o p-valor seja menor que o nível de confiança (0,95), é aceita a hipótese nula de erros descorrelatados (H_0). Como o p-valor estimado foi 0,9951, seria rejeitada a hipótese nula no teste de Ljung-Box. Porém, observando o gráfico da autocorrelação dos erros abaixo, pode-se considerar que os resíduos são na verdade descorrelatados, pois vê-se que as autocorrelações encontram-se na faixa de -0,2 a 0,2.

Figura 5 – Função de Autocorrelação dos Erros – MAE.



Fonte: Elaboração própria.

Em relação ao modelo estimado utilizando o método Univariado de Box & Jenkins (MBJ), com intuito de projetar o consumo de gasolina, teve-se como resultado um modelo ARIMA(0,1,2)*(1,0,1) com transformação de raiz quadrada.

¹⁹ Ver Zanini (2023).

Tabela 5 – Parâmetros do Modelo de Box & Jenkins

| Termo | Coefficiente | P-valor |
|--------------|---------------------|----------------|
| θ_1 | 0,5820 | 1,0000 |
| θ_2 | -0,0924 | 0,9746 |
| Φ | 0,9981 | 1,0000 |
| Θ | 0,8814 | 1,0000 |

Fonte: Elaboração própria.

Após análise de significância dos parâmetros estimados, é observado (tabela 5) que todos são significantes ao nível de significância de 5%, uma vez que o p-valor de cada parâmetro está acima de 0,95. Logo, a equação 5.2.2 abaixo representa o modelo ARIMA(0,1,2)*(1,0,1), já com os valores encontrados para cada parâmetro.

$$(1 - \Phi_1 B^{12})(1 - B)\sqrt{Z_t} = (1 - \Theta_1 B^{12})(1 - \theta_1 B - \theta_2 B^2)a_t \quad (5.2.2)$$

$$(1 - 0,9981B^{12})(1 - B)\sqrt{Z_t} = (1 - 0,8814B^{12})(1 - 0,5820B + 0,0924B^2)a_t \quad (5.2.3)$$

Tabela 6 – Estatística dentro da amostra – MBJ

| Componentes | Resultados |
|-------------------------|-------------------|
| R ² Ajustado | 98,07% |
| MAPE | 5,55% |

Fonte: Elaboração própria.

O R² ajustado do modelo de Box & Jenkins demonstrou uma capacidade explicativa de 98,07% em relação ao consumo no período analisado, com uma margem de erro médio (MAPE) de 5,55% para cima ou para baixo ao projetar o próximo mês (tabela 6). Já quanto ao teste de Ljung-Box²⁰, como o p-valor estimado foi 0,961, seria rejeitada a hipótese nula de erros descorrelatados. Porém, observando o gráfico da autocorrelação dos erros (figura 6), pode-se considerar que os resíduos são na verdade descorrelatados, pois vê-se que as autocorrelações encontram-se na faixa de -0,2 a 0,2.

²⁰ Para a estatística amostral, caso o p-valor seja menor que o nível de confiança, é aceita a hipótese nula de erros descorrelatados (H₀).

Figura 6 – Função de Autocorrelação dos Erros – MBJ.



Fonte: Elaboração Própria.

5.3 Competição de Métodos

A tabela 7 abaixo organiza alguns resultados da análise recursiva fora da amostra apresentada no capítulo 4.

Tabela 7 – Competição entre os Métodos.

| Dentro da Amostra (<i>In sample</i>) | | | Fora da Amostra ²¹ (<i>Out of Sample</i>) | | | |
|---|-----------------------|----------|---|------------------|----------------------|---------------|
| Modelo | R ² adj(%) | MAPE (%) | MAPE (%) H=1 N=12 | MAPE ACUM (%) | GMRAE H=1 N=12 | GMRAE ACUM |
| MAE | 98,21 | 5,64 | 4,5 | 7,2 | 7,5 | 1,22 |
| MBJ | 98,01 | 5,58 | 4,4 | 8,3 | 7,4 | 1,50 |

Fonte: Elaboração própria.

²¹ H representa o Horizonte de Previsão e o termo “ACUM” representa o acumulado, sendo referente às 78 previsões.

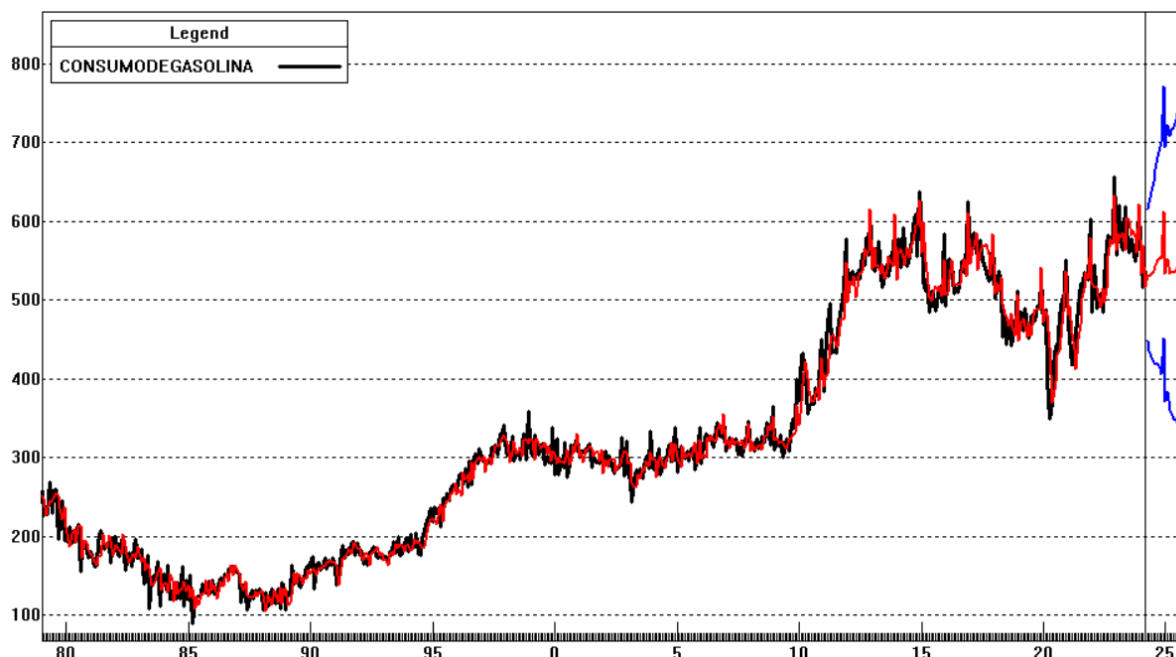
Analisando a tabela 7, vê-se que os dois métodos têm bom poder de generalização, ou seja, prevêem melhor fora da amostra do que dentro. No desempenho preditivo dentro da amostra, os resultados são muito similares, sendo que o MAE apresenta um poder de explicação um pouco maior. Fora da amostra, o MBJ possui o MAPE ligeiramente menor para as previsões 1 passo à frente, mas no acumulado das 78 previsões o MAE tem erro menor do que o MBJ. Desta forma, seleciona-se o MAE como método vencedor. Quanto ao GMRAE, nas previsões 1 passo à frente, ambos os métodos ganharam do método *naive*, e ambos perdem deste no acumulado das 78 previsões.

Feitas estas considerações e selecionado o Método de Amortecimento Exponencial como método vencedor, os dados anteriormente retirados são reintegrados à amostra, fazendo-se a atualização paramétrica. Desta forma, chega-se na equação 5.2.1, apresentada anteriormente e que será utilizada para calcular as previsões 21 meses à frente, ou seja, até dezembro de 2025.

5.4 Previsão do consumo de Gasolina

Já selecionado o método vencedor (MAE), utilizando um modelo Holt-Winters, tem-se os valores projetados exibidos na figura 7 que mostra a evolução do consumo de gasolina no Brasil, além do intervalo de confiança de 95%.

Figura 7 – Consumo de gasolina (em milhares de barris) – período de 1979 a 2025. Valores reais, ajustado e previsões com intervalo de confiança de 95%.



Fonte: Elaboração própria. Nota: — Valores reais — Valores projetados — Intervalo de Confiança de 95%

Na tabela 8 a seguir estão as previsões para o período de abril 2024 a dezembro de 2025, assim como os valores calculados para o intervalo de confiança de 95%. A tabela 08 evidencia uma das vantagens de se trabalhar com métodos estatísticos de análise. Vê-se que a previsão é obtida de forma probabilística, ou seja, não se tem apenas um número, a previsão, mas o intervalo de confiança que possui determinada probabilidade (neste caso de 95%) de conter o verdadeiro valor do consumo. Sendo assim, estes limites poderiam perfeitamente serem usados para gerar “cenários”. O limite inferior poderia configurar um “cenário pessimista” e o limite superior um “cenário otimista” com a previsão sendo considerada o “cenário esperado”.

Tabela 8 – Valores da Previsão do Consumo de Gasolina (em milhares de barris) – Intervalo de Confiança de 95%.

| Mês | Limite Inferior | Previsão | Limite Superior |
|------------|------------------------|-----------------|------------------------|
| 2024-04 | 448 | 531 | 614 |
| 2024-05 | 435 | 529 | 624 |
| 2024-06 | 428 | 533 | 637 |
| 2024-07 | 422 | 536 | 650 |
| 2024-08 | 419 | 542 | 666 |
| 2024-09 | 417 | 550 | 682 |
| 2024-10 | 413 | 553 | 694 |
| 2024-11 | 405 | 553 | 701 |
| 2024-12 | 451 | 611 | 772 |
| 2025-01 | 371 | 533 | 694 |
| 2025-02 | 383 | 552 | 722 |
| 2025-03 | 359 | 533 | 707 |
| 2025-04 | 356 | 536 | 716 |
| 2025-05 | 349 | 534 | 719 |
| 2025-06 | 346 | 537 | 728 |
| 2025-07 | 344 | 541 | 737 |
| 2025-08 | 345 | 547 | 749 |
| 2025-09 | 347 | 554 | 762 |
| 2025-10 | 346 | 558 | 771 |
| 2025-11 | 340 | 557 | 775 |
| 2025-12 | 391 | 617 | 843 |

Fonte: Elaboração própria.

6 CONCLUSÃO

O trabalho apresentado teve como objetivo principal estimar um modelo de consumo de gasolina no Brasil. Devido à grande influência da gasolina para economia brasileira, justifica-se a importância de um estudo voltado para a projeção de dados sobre esse derivado. Logo, é possível que as informações fornecidas sejam base para decisões de agentes econômicos.

A priori, foi apresentado o comportamento do mercado de combustível, mais especificamente, da gasolina. Com uma introdução ao contexto histórico e seus impactos para economia, também foram realizadas análises comparativas dos principais indicadores de produção e consumo, assim como projeções para o setor.

Buscando o melhor modelo para a previsão dos dados, foram utilizados dois métodos, univariados e autoprojetivos. O Método de Amortecimento Exponencial (MAE) e o Método de Box & Jenkins (MBJ). Dessa forma, foram estudadas as metodologias abordadas por cada método, assim como as diferentes aplicações utilizadas nos trabalhos evidenciados pela revisão de literatura.

Logo, o método vencedor foi escolhido após a competição dos modelos. Com a comparação do desempenho preditivo, o MAE mostrou-se com uma melhor capacidade de explicação, assim como menor erro acumulado nas 78 previsões à frente. O modelo de Holt-Winters foi capaz de captar as tendências, e calcular o consumo mês a mês. Os fatores sazonais mostraram, por exemplo, como eventos sociais podem influenciar diretamente o mercado, aumentando o consumo em meses específicos, principalmente em dezembro.

As previsões para o consumo de gasolina foram apresentadas na forma probabilística, o que é uma das vantagens de se trabalhar com modelos estatísticos aplicados à análise de séries de tempo. Dessa forma, não se gera apenas um número, mas um intervalo de confiança que possibilita a configuração de cenários de análise.

Com intuito de trazer sugestões para trabalhos futuros, atenta-se que a utilização de Modelos como os de Regressão Dinâmica podem trazer diferentes resultados. A inclusão de outras variáveis na análise permite trazer diferentes perspectivas para o futuro do mercado de gasolina no Brasil.

REFERÊNCIAS

AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS.

Dados estatísticos do setor de combustíveis. Disponível em:

<<http://www.anp.gov.br>>. Acesso em: 07 mar. 2024

BOX, G. E. P., JENKINS, G. M. Time Series Analysis, Forecasting and Control, San Francisco, Holden-Day, 1994.

BRAFMAN, Ilana. **Modelo econométrico para a projeção do consumo aparente de combustível no Brasil - Otto e Diesel**. Rio de Janeiro: IBMEC, 2009.

EMPRESA DE PESQUISA ENERGÉTICA. PDE 2031: Oferta de Derivados de Petróleo. Disponível em: <https://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/Documents/PDE%202031_RevisaoPosCP_rvFinal_v2.pdf>. Acesso em: 03 jun. 2024.

Empresa de Pesquisa Energética (EPE). (2023). Perspectivas para o Mercado Brasileiro de Combustíveis no Curto Prazo.

FERNANDES, R. A; BRAGA, José. **Poder de mercado no segmento de distribuição de gasolina C: uma análise a partir da Nova Organização Industrial Empírica**. Disponível em: <https://www.scielo.br/j/ee/a/WHf7BSTkzLbjvvvnTtPCJ7z/?lang=pt#>>.

GOODRICH, R.L. Applied Statistical Forecasting, Belmont, Business Forecast Systems, 1989.

INSTITUTO BRASILEIRO DE PETRÓLEO E GÁS (IBP). Evolução das vendas de combustíveis líquidos. Disponível em: <https://www.ibp.org.br/observatorio-do-setor/snapshots/evolucao-das-vendas-de-combustiveis-liquidos/>. Acesso em: 6 jun. 2024.

INSTITUTO DE PESQUISA ECONÔMICA APLICADA – **Ipeadata**. Dados macroeconômicos e regionais. Disponível em: <<http://www.ipeadata.gov.br>>. Acesso em 25 mar. 2024.

KEIKHA, M. R.; ZIAEI, M. **Modeling and forecasting petroleum consumption in Iran using hybrid intelligent algorithm**. Energy Reports, v. 5, p. 182-189, 2019.

MINISTÉRIO DE MINAS E ENERGIA. Política energética brasileira. Disponível em: <<http://www.mme.gov.br>>. Acesso em: 07 mar. 2024.

PETROBRAS. História da Petrobras. Disponível em: <<http://www.petrobras.com.br/pt/quem-somos/historia/>>. Acesso em: 5 jun. 2024.

PETROBRAS. Relatório de produção e capacidade de refino. Disponível em: <<http://www.petrobras.com.br>>. Acesso em: 07 mar. 2024.

SAMOHYL, R. W.; DANTAS, A. de B. (1998). Previsão com cointegração e modelo a correção de erro: aplicação empírica no consumo de gasolina no Brasil.

SANTIAGO, F. S. Um modelo econométrico + insumo-produto para a previsão de longo prazo da demanda de combustíveis no Brasil. Universidade Federal de Juiz de Fora, Juiz de Fora, 2009.

SOUZA, R. C., CAMARGO, M. E., Análise e Previsão de Séries Temporais: Os Modelos ARIMA. SEDIGRAF, 1996.

SOUZA, Y. M. de; RODRIGUES, L.; FURLAN AMARAL, D. ANÁLISE ESPACIAL DA DEMANDA POR DIESEL NOS ESTADOS BRASILEIROS. Revista de Economia e Agronegócio, [S. l.], v. 19, n. 1, p. 1–26, 2021. DOI: 10.25070/rea.v19i1.11671. Disponível em: <https://periodicos.ufv.br/rea/article/view/11671>. Acesso em: 26 mar. 2024.

MONTGOMERY, D.C., JOHNSON, L.A. **Forecasting and Time Series Analysis**, New York, McGraw-Hill Book Co., 1990.

ZANINI, A. Modelos de Previsão para Séries Temporais. Material Didático. Universidade Federal de Juiz de Fora, 2023.

ZANINI, A. **Redes Neurais e Regressão Dinâmica: um modelo híbrido para previsão de curto prazo da demanda de gasolina automotiva no Brasil**. 85p. Dissertação - Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2000.