

PROJETANDO VENDAS DE ÓLEO DIESEL NO RIO GRANDE DO SUL

Matheus Vizzotto dos Santos - 00290240

RESUMO

A venda de óleo diesel no Rio Grande do Sul tem representado um importante componente da economia tanto no setor industrial quanto no agrícola. Em ambos a comercialização do derivado de petróleo tem como um dos principais destinos o abastecimento de caminhões de transporte para equipamentos, matérias primas e produtos agropecuários, além de servir como insumo para as máquinas agrícolas na colheita e no plantio. Com o objetivo de estimar as vendas mensais de óleo diesel no estado em 2022, este trabalho procura atestar a validade dos componentes da série temporal em questão. Utilizando a metodologia de Box & Jenkins, apurou-se o melhor modelo para fazê-lo: chegando à projeção de 3,86 bilhões de litros de óleo diesel vendidos no período de estimativa, o estado tende a apresentar um aumento de 23,4 milhões de litros vendidos em relação ao ano anterior.

PALAVRAS-CHAVE

Óleo diesel; projeção; Box & Jenkins.

1. INTRODUÇÃO

O óleo diesel é um dos insumos mais representativos para a atuação nos setores da indústria, da agricultura e da pecuária no que diz respeito ao transporte. Em 1990, a comercialização em nível nacional foi de 27,2 bilhões de litros, chegando a 57,4 bilhões em 2020 com um máximo histórico de 60 bilhões em 2014.

Com um crescimento de 110,8% entre 1990 e 2020, torna-se relevante a estimativa das vendas de combustíveis tanto para avaliações setoriais quanto para projeções de empresas, que podem basear a evolução de suas vendas no *market share* corrente. No entanto, estando intimamente ligada ao segmento agrícola da economia, é preciso notar que a venda de óleo diesel será

influenciada pelos ciclos naturais das principais culturas do estado: soja, milho e trigo. A primeira e a segunda terão suas colheitas no primeiro semestre do ano, concentrando-se nos meses de fevereiro a abril; a terceira, no segundo semestre do ano, com foco entre setembro e novembro. Além disso, o plantio da soja e do milho também ocorrem no final do ano (entre outubro e dezembro). É razoável levar em conta, pois, que a venda do combustível sofrerá efeitos de sazonalidade nesses períodos e que precisam ser levados em conta na projeção mensal de sua comercialização. Sendo também relacionada ao nível de atividade econômica da região (especialmente ao PIB), a série histórica da comercialização de óleo diesel apresenta tendência significativa – outro componente que necessita de tratamento para a projeção de vendas. Ambos estes aspectos da série temporal devem receber atenção, uma vez que afetam a análise de autocorrelação e de autocorrelação parcial e podem levar a estimativas mensais errôneas.

O objetivo deste estudo é contribuir para a elaboração de modelos de séries temporais que projetem vendas de óleo diesel no Rio Grande do Sul com base nos dados disponíveis de 1990 a 2021. Fazendo-o, é possível auxiliar o setor de forma agregada ou ao nível da firma na sua organização interna.

2. REVISÃO DE LITERATURA

Para modelar séries temporais, é possível trabalhar com modelos estatísticos tradicionais, como média móvel, alisamento exponencial e ARIMA. A modelagem de séries temporais para projeção é usada principalmente para prever demanda. Kurawarwala e Matsuo, por exemplo, calcularam a variação sazonal de demanda usando dados históricos e validaram os modelos por meio da performance das projeções. Um modelo ARIMA (p , d , q) é aquele no qual p é o número de termos autoregressivos; q , o número de termos de médias móveis; d , o número de diferenciações. Para o primeiro processo (AR), os modelos assumem que Y_t é uma função linear dos seus valores passados e expressam o componente pela equação I:

$$Y_t = \alpha_1 Y_{t-1} + \varepsilon_t$$

O valor corrente de um processo de médias móveis (MA), por sua vez, é uma combinação linear da perturbação corrente com uma ou mais perturbações passadas:

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

O processo de integração (I), diz respeito ao comportamento de séries temporais afetadas por efeitos cumulativos, as quais precisam passar por uma transformação à estacionariedade antes de serem modeladas para projeção.

No entanto, existe a possibilidade de que algumas séries apresentem componente sazonais a serem incluídos na modelagem; a partir disso, chega-se ao modelo SARIMA (p, d, q) (P, D, Q)_s, onde *P* representa o número de coeficientes sazonais autorregressivos multiplicativos; *Q*, o número de coeficientes sazonais de médias móveis multiplicativos; *D*, o número de diferenciações sazonais; e *s*, o período sazonal. Um modelo SARIMA (0,1,1) (0,1,1)₁₂, por exemplo, é dado pela seguinte equação:

$$\Delta_{12}\Delta Y_t = \mu + \phi_Q(L^{12}) \theta(L) \varepsilon_t$$

Onde $\phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p$ e $\theta(L) = 1 + \theta_1 L + \dots + \theta_q L^q$, sendo *L* o operador de defasagem $LY_t = Y_{t-1}$ e ε_t o termo de erro assumido como variável aleatória independente identicamente distribuída e retirada de uma amostra com distribuição de média zero e variância constante.

A metodologia de Box & Jenkins foi fundada nas contribuições de Yule e Wold para desenvolver uma abordagem prática na determinação de modelos ARIMA. Inicialmente, é necessário ver que, se uma série é obtida de um modelo ARIMA, ela deveria ter algumas propriedades teóricas de autocorrelação que, unidas aos padrões empíricos observados, irão possibilitar a identificação de um ou mais modelos potenciais para a série em questão. Box & Jenkins propuseram o uso da função de autocorrelação (FAC) e da função de autocorrelação parcial (FACP) da amostra de dados como as ferramentas básicas para identificar as ordens de um modelo ARIMA. Para esse objetivo, é necessário partir de uma série temporal estacionária: com a diferenciação, por exemplo, é possível remover a tendência e estabilizar a variância antes que o modelo seja estimado. Depois disso, passa-se à etapa de checagem de diagnóstico (verificar se o modelo escolhido se ajusta razoavelmente bem), onde se observa se os resíduos são ruído branco. É possível fazer isso com a análise das FAC e FACP, e também com os testes de raiz unitária e de Durbin-Watson.

4. ESTIMAÇÃO E RESULTADOS

A projeção das vendas de óleo diesel no Rio Grande do Sul foi estimada com base em dados mensais da Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP) desde

1990 até setembro de 2021. A partir da metodologia de Box & Jenkins, procurou-se alcançar o melhor modelo na determinação de vendas futuras do derivado do petróleo submetendo modelos às etapas de identificação, estimação e verificação. Inicialmente, no entanto, dividiu-se a série em duas partições de treino e teste a fim de que seja possível assegurar, dentro de certos limites, a eficácia do modelo em dados não observados. A segunda, de teste, corresponde aos 12 últimos meses; a primeira, de treino, ao restante da série (janeiro de 1990 a agosto de 2020).

No processo de identificação, inicia-se com análises de autocorrelação da série em nível e com testes de raiz unitária para verificar o nível de diferenciação necessário para tornar a série estacionária, para então escolher possíveis valores de p e q . Além disso, fez-se a transformação logarítmica da série para que ela se aproximasse de uma variância constante. Com os gráficos da figura I, pode-se indicar a presença de tendência e, portanto, a necessidade de diferenciação. O teste de estacionariedade utilizado no presente artigo foi o ADF (*Augmented Dickey-Fuller*), que tem como hipótese nula (H_0) a presença de raiz unitária na série e como hipótese alternativa (H_1) a estacionariedade. O resultado pode ser observado na tabela I, e corrobora as indicações dos gráficos de FAC da série em nível: com um p-valor superior a 10%, não é possível negar a presença de raiz unitária nos dados e é necessário calcular a diferença.

Figura I: gráficos da FAC e da FACP da série em nível e em logaritmo

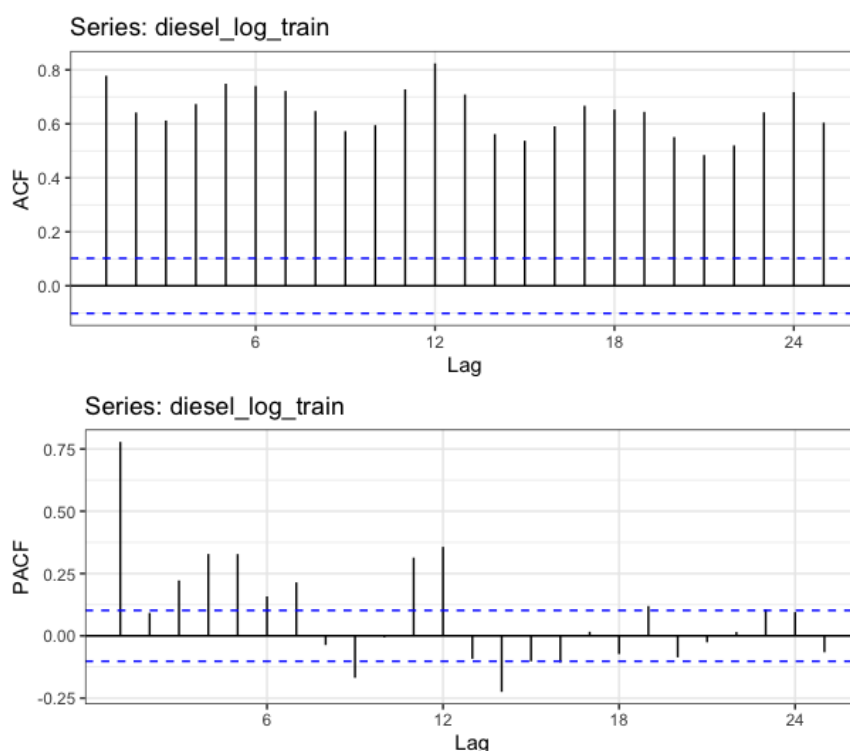


Tabela I. Resultados do teste ADF com a série em nível

| | |
|---------------|---------|
| Dickey-Fuller | -2,9055 |
| Lags | 7 |
| p-valor | 0,1948 |

Na segunda etapa, os modelos foram estimados pelo algoritmo de máxima verossimilhança. Os escolhidos foram estimações ARIMA(6, 1, 3), SARIMA (6, 1, 3) (0, 0, 1)₁₂ e SARIMA(5, 1, 2)(1, 1, 1)₁₂. Para serem comparados, observaram-se métricas como os critérios de informação Akaike (AIC) e Bayesiano (BIC), o erro médio (ME), a raiz do erro quadrático médio (RMSE), o erro médio absoluto (MAE) e o erro médio absoluto percentual (MAPE), que podem ser observados na tabela II.

Tabela II - Comparação dos modelos estimados

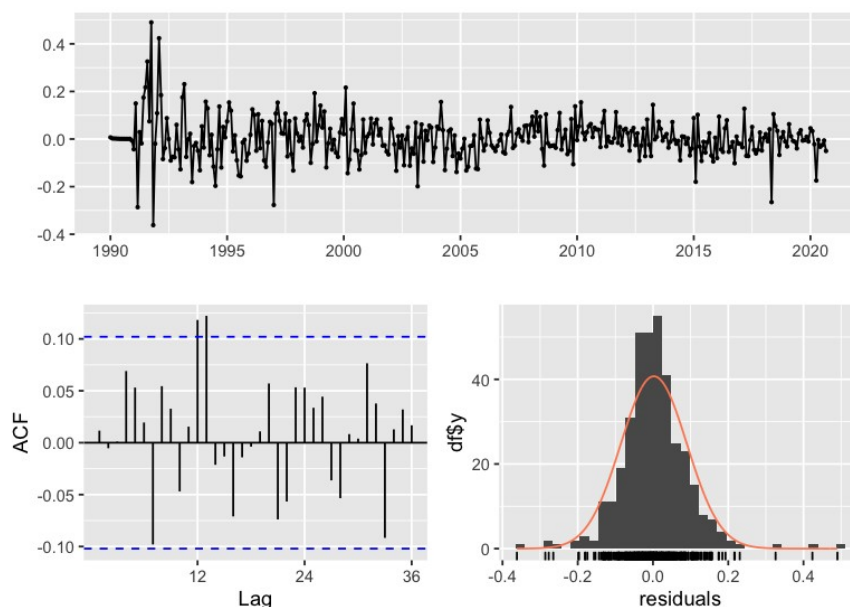
| | ARIMA(6, 1, 3) | SARIMA(6, 1, 3)(0, 0, 1)[12] | SARIMA(5, 1, 1)(1, 1, 1)[12] |
|------|----------------|------------------------------|------------------------------|
| AIC | -529,80 | -574,97 | -673,15 |
| BIC | -490,72 | -531,98 | -634,41 |
| ME | 0,0666 | 0,0726 | 0,0456 |
| RMSE | 0,1449 | 0,1408 | 0,0796 |
| MAE | 0,1189 | 0,1132 | 0,0610 |
| MAPE | 0,9345 | 0,8894 | 0,4794 |

O modelo que apresenta melhor desempenho quando comparado com a partição de teste é o SARIMA (5, 1, 1)(1, 1, 1)₁₂, e seus coeficientes estimados podem ser vistos na tabela III, junto com o diagnóstico dos seus resíduos, na figura II. Embora o modelo em questão apresente as melhores métricas, o gráfico ACF dos seus resíduos indica a presença de autocorrelação nos lags 11 e 13, o que poderia ser explorado em modelos mais complexos. Isso é confirmado pelo teste de Ljung-Box, cuja hipótese nula é a de ausência de autocorrelação nas defasagens e retornou um p-valor de 1,3% (portanto, significativo). A comparação dos valores observados com os estimados e da partição de treinamento com a de teste pode ser visualizada na seção de Anexo do artigo.

Tabela III - Estimação dos coeficientes

| ar1 | ar2 | ar3 | ar4 | ar5 | ma1 | ma2 | sar1 | sma1 |
|---------|--------|--------|--------|---------|---------|---------|--------|---------|
| -0,5829 | 0,1486 | 0,1374 | 0,0172 | -0,0606 | -0,2603 | -0,5702 | 0,0631 | -0,8415 |

Figura II - Diagnóstico dos resíduos do modelo



Depois de definir o modelo mais apropriado para estimar as vendas de óleo diesel, extraiu-se o modelo para, então, aplicá-lo à série completa (janeiro de 1990 a setembro de 2021). Com isso, foi possível chegar aos valores disponibilizados na tabela IV: nota-se o característico aumento das vendas de óleo diesel nos meses de março, abril, outubro e novembro (relacionados com os ciclos das culturas de soja, milho e trigo para colheita e plantio). Chegando à projeção de vendas de 3,866 bilhões de litros, o estado do Rio Grande do Sul alcançará, de acordo com o modelo, 934 milhões de litros vendidos ainda no primeiro trimestre de 2022.

Tabela IV - Projeção de vendas de óleo diesel para 2022

| Data | 01/01/22 | 01/02/22 | 01/03/22 | 01/04/22 | 01/05/22 | 01/06/22 | 01/07/22 | 01/08/22 | 01/09/22 | 01/10/22 | 01/11/22 | 01/12/22 |
|----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| Projeção | 282.308,8 | 274.439,9 | 377.292,4 | 366.523,5 | 310.228,9 | 300.139,1 | 305.743,2 | 330.838,4 | 313.343,7 | 356.816,1 | 346.037,6 | 302.785,9 |

5. CONCLUSÃO

A projeção de vendas de óleo diesel é um importante indicador dos setores de combustíveis, da indústria e da agricultura, tanto no nível macroeconômico quanto no nível da empresa. Neste último, o trabalho contribui com a construção cenários de crescimento mais assertivos para, por exemplo, empresas de TRR (Transportador-Revendedor-Retalhista), cujo ramo se baseia no atendimento das propriedades rurais com abastecimento de óleo diesel. A

especificação do modelo nos dados de treinamento proporciona também uma maior capacidade de generalização, já que é possível testar sua eficiência em dados *fora de amostra*. Seguindo a metodologia de Box & Jenkins, o estudo foi conduzido por métricas largamente utilizadas na literatura, como o Critério de Informação Akaike (AIC), o Critério de Informação Bayesiano (BIC) e as medidas de erro (ME, RMSE, MAE, MAPE). O modelo que apresenta os melhores desses indicadores foi o SARIMA (5, 1, 1)(1, 1, 1)₁₂, o que torna acessível a utilização para projetar as vendas de óleo diesel no estado. Em trabalhos futuros, pretende-se eliminar o problema de autocorrelação dos resíduos nas defasagens 11 e 13 por meio de metodologias mais complexas, como um modelo SARIMAX (*Seasonal Autoregressive Moving Average with Exogenous Variables*), que permite a adição de variáveis exógenas ao modelo, ou o uso de algoritmos como os de Redes Neurais Artificiais (ANN).

6. BIBLIOGRAFIA

ANP (2021) *Anp – dados abertos*. Disponível em <https://www.gov.br/anp/pt-br/centrais-de-conteudo/dados-abertos/vendas-de-derivados-de-petroleo-e-biocombustiveis>. Acesso em: 20 Out 2021

Box, G.; Jenkins, G. (2008) *Time Series Analysis: Forecasting and control*. 4th ed. New York: John Wiley & Sons

Bueno, Rodrigo (2018) *Econometria de Séries Temporais*. 2a edição. São Paulo: Cengage Learning

Enders, Walter (2009) *Applied econometric time series*. 3rd ed. New York: Wiley

Kurawarwala, A. A. and Matsuo H (1998) *Product growth models for medium-term forecasting of short life cycle products*. Tech- nol Forecast Soc Chang

7. ANEXO

