UNIVERSIDADE FEDERAL DE JUIZ DE FORA
FACULDADE DE ECONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

Junior Gabriel Nardino Fumagali

Three essays about the fuel prices at the pump in Brazil (gasoline, ethanol, diesel, and Liquefied Petroleum Gas): (univariate and multivariate) forecasting and volatility

Juiz de Fora
2026

Junior Gabriel Nardino Fumagali

Three essays about the fuel prices at the pump in Brazil (gasoline, ethanol, diesel, and Liquefied Petroleum Gas): (univariate and multivariate) forecasting and volatility

Tese apresentada ao Programa de Pós-Graduação em Economia da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Doutor em Economia. Área de concentração: Economia Regional e Macroeconomia.

Orientador: Prof.Dr. Wilson Luiz Rotatori Corrêa

Coorientador: Prof. Dr. Rafael Morais de Souza

Juiz de Fora

2026

Ficha catalográfica elaborada através do Modelo Latex do CDC da UFJF
com os dados fornecidos pelo(a) autor(a)

Sobrenome, Nome do autor.
Three essays about the fuel prices at the pump in Brazil (gasoline, ethanol, diesel, and Liquefied Petroleum Gas): (univariate and multivariate) forecasting and volatility / Junior Gabriel Nardino Fumagali. – 2026.
135 f. : il.

Orientador: Wilson Luiz Rotatori Corrêa
Coorientador: Rafael Morais de Souza
Tese (Doutorado) – Universidade Federal de Juiz de Fora, Faculdade de Economia. Programa de Pós-Graduação em Economia, 2026.

1. Palavra-chave. 2. Palavra-chave. 3. Palavra-chave. I. Sobrenome, Nome do orientador, orient. II. Título.

Junior Gabriel Nardino Fumagali

Three essays about the fuel prices at the pump in Brazil (gasoline, ethanol, diesel, and Liquefied Petroleum Gas): (univariate and multivariate) forecasting and volatility

Tese apresentada ao Programa de Pós-graduação em Economia da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Doutor em Economia. Área de concentração: Economia.

Aprovada em 15 de junho de 2026.

BANCA EXAMINADORA

Dr. Wilson Luiz Rotatori Corrêa - Orientador

Universidade Federal de Juiz de Fora

Dr. Rafael Morais de Souza - Coorientador

Universidade Federal de Juiz de Fora

Dr. André Suriane

Universidade Federal de Juiz de Fora

Dr. Cláudio Roberto Foffano Vasconcelos

Universidade Federal de Juiz de Fora

Dr. Guilherme Armando de Almeida Pereira

Universidade Federal do Espírito Santo

Dr. Victor Eduardo de Mello Valerio

Universidade Federal de Itajubá

Juiz de Fora, 08/06/2026.

Documento assinado eletronicamente por **Wilson Luiz Rotatori Correa**, **Professor(a)**, em 18/06/2026, às 11:49, 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 **ANDRÉ SURIANE DA SILVA**, **Usuário Externo**, em 18/06/2026, às 13:05, 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 **Rafael Morais de Souza**, **Professor(a)**, em 18/06/2026, às 14:15, 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 **Claudio Roberto Foffano Vasconcelos**, **Professor(a)**, em 18/06/2026, às 19:01, 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 **Guilherme Armando de Almeida Pereira**, **Usuário Externo**, em 19/06/2026, às 09:32, 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 **Victor Eduardo de Mello Valerio**, **Usuário Externo**, em 23/06/2026, às 11:00, 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 **3017735** e o código CRC **3E31A3DA**.

Para os meus pais e amigos.

ACKNOWLEDGMENTS

Bom, foi uma longa jornada desde fevereiro de 2022 até junho de 2026, em que eu fiz muitos amigos, conheci pessoas incríveis e diferentes, e essa pós me permitiu ir a lugares extraordinários em quatro anos.

Primeiramente, o apoio dos meus pais, Elias e Lizete, que me apoiaram a ir para Juiz de Fora e começar esse doutorado. E me mantiveram firme quando as coisas desandaram e me ajudaram a pôr nos trilhos. Além de suportarem quando eu achei que o mundo tinha desmoronado, e ele desmoronou algumas vezes.

Eu cheguei ao PPGE-UFJF sem expectativas e sem saber muita coisa, tinha passado por experiências não muito boas em termos acadêmicos, nas etapas anteriores, então estava de coração aberto para Juiz de Fora. Esse primeiro e importante passo abriu portas para que eu conhecesse diversos servidores dessa incrível universidade, que foi o meu lar por 4 anos. Assim, quero agradecer a cada um dos servidores, em especial às pessoas incríveis: Jerusa Furtado Vital, Adriana Guedes, Julia, Rose e Leni, que sempre me ajudaram em todas as etapas da minha trajetória.

Em termos de coordenação, o apoio da Nicole e do Antônio foi fundamental para eu concluir a minha pós. Os meus orientadores, Wilson Luiz Rotatori Corrêa e Rafael Morais de Souza, foram excepcionais do começo ao fim, sobretudo o Rafael, que me deu a minha primeira oportunidade em julho de 2022, ao apresentar o SSA, e o resto foi história. O prof. Wilson, que me acolheu logo na primeira semana, dada a nossa intrínseca familiaridade com modelos VAR (que me acompanhou e me levou a uma jornada nunca antes vista!).

Durante a jornada desse doutorado, é importante ressaltar o apoio que recebi do Laboratório de Análises Territoriais e Setoriais (LATES), por ceder sua infraestrutura para uso na minha pesquisa.

Ainda em termos acadêmicos, agradecimentos especiais ao prof. Marcelo Medeiros da UIUC por receber em um sanduíche que eu achei que estava perdido após uma negativa da Austrália na última hora. Como também ao apoio fundamental do professor Yaser Samadi, da SIU, que gentilmente me entregou seu modelo para que eu pudesse adaptá-lo ao meu problema de pesquisa e respondeu prontamente a todas as minhas dúvidas. Ainda sobre os EUA, um destaque especial a Elis Artz e família por todo apoio nos EUA.

Por último, agradecer ao meu amigo Maurício Beirão da ROcha que tem sido desde 2022 a pessoa que mais conversa comigo e me apoiou em vários momentos difíceis. Ao José Carlos Grizendi, por toda a parceria desde 2022 e por ter me ajudado quando precisei. Ao Gabriel Arcanjo, que sempre se dispôs a me ajudar no que precisei durante meu tempo no PPGE. Por fim, aos meninos do futebol que sempre estiveram presentes, como também ao demais alunos que sempre conviveram na salinha comigo.

"Só sei que nada sei."

Sócrates

ABSTRACT

This thesis consists of three essays that investigate retail fuel prices in Brazil, focusing on forecasting and volatility dynamics of gasoline and diesel prices. Given the importance of fuel prices for inflation, economic activity, and public policy, the essays employ different econometric and statistical approaches to provide new evidence on the behavior of the Brazilian fuel market.

The first essay evaluates the forecasting performance of the non-parametric Singular Spectrum Analysis (SSA) method for retail gasoline and diesel prices across Brazil's 27 states. The predictive accuracy of SSA is compared with ARIMA, Exponential Smoothing, and Neural Networks using monthly data from July 2001 to August 2023. The results show that SSA achieved the best in-sample performance, whereas ARIMA outperformed the benchmark models in the out-of-sample analysis. However, SSA provided complementary forecasting information for long-term horizons.

The second essay examines the transmission of volatility from international crude oil prices to domestic gasoline and diesel prices using a multivariate GARCH model with the BEKK specification. The results indicate that gasoline prices are more responsive to international oil price shocks than diesel prices. Moreover, volatility spillovers intensified during the Import Parity Pricing (IPP) policy, when domestic fuel prices became more closely linked to international oil markets.

The third essay introduces the Envelope Matrix Autoregressive (EMAR) model as an alternative approach for forecasting retail fuel prices in Brazil. By incorporating envelope dimension reduction, EMAR jointly models multiple fuel types and states while estimating fewer parameters than conventional multivariate models. Its forecasting performance is comparable to that of the Matrix Autoregressive (MAR) model, providing a more parsimonious specification.

Overall, the thesis contributes to the literature by advancing forecasting methodologies and improving the understanding of retail fuel price dynamics and volatility transmission in Brazil, offering insights for researchers, policymakers, and market participants.

Keywords: Fuel Prices; Forecasting; Volatility Spillovers

JEL CODE: C32; C53; Q41

RESUMO

Esta tese é composta por três ensaios que investigam os preços dos combustíveis ao consumidor no Brasil, com foco na previsão e na dinâmica da volatilidade dos preços da gasolina e do diesel. Considerando a importância dos preços dos combustíveis para a inflação, a atividade econômica e a formulação de políticas públicas, os ensaios empregam diferentes abordagens econométricas e estatísticas para fornecer novas evidências sobre o comportamento do mercado brasileiro de combustíveis.

O primeiro ensaio avalia o desempenho preditivo do método não paramétrico Singular Spectrum Analysis (SSA) na previsão dos preços da gasolina e do diesel ao consumidor nos 27 estados brasileiros. A capacidade preditiva do SSA é comparada aos modelos ARIMA, Suavização Exponencial e Redes Neurais, utilizando dados mensais de julho de 2001 a agosto de 2023. Os resultados mostram que o SSA apresentou o melhor desempenho na análise in-sample, enquanto o modelo ARIMA superou os demais modelos na análise out-of-sample. Entretanto, o SSA forneceu informações preditivas complementares para horizontes de previsão de longo prazo.

O segundo ensaio analisa a transmissão da volatilidade dos preços internacionais do petróleo bruto para os preços domésticos da gasolina e do diesel por meio de um modelo GARCH multivariado na especificação BEKK. Os resultados indicam que os preços da gasolina são mais sensíveis aos choques nos preços internacionais do petróleo do que os preços do diesel. Além disso, o transbordamento de volatilidade tornou-se mais intenso durante a vigência da política de Preço de Paridade de Importação (PPI), período em que os preços domésticos passaram a refletir de forma mais estreita as condições do mercado internacional.

O terceiro ensaio apresenta o Envelope Matrix Autoregressive Model (EMAR) como uma abordagem alternativa para a previsão dos preços dos combustíveis ao consumidor no Brasil. Ao incorporar a redução de dimensão via envelope, o EMAR modela conjuntamente diferentes tipos de combustíveis e estados brasileiros, estimando um número menor de parâmetros em comparação aos modelos multivariados convencionais. Seu desempenho preditivo mostrou-se comparável ao do Matrix Autoregressive Model (MAR), oferecendo uma especificação mais parcimoniosa.

De forma geral, a tese contribui para a literatura ao avançar as metodologias de previsão e ampliar a compreensão da dinâmica dos preços e da transmissão de volatilidade dos combustíveis no Brasil, fornecendo evidências empíricas relevantes para pesquisadores, formuladores de políticas públicas e participantes do mercado.

Palavras-Chave: Preço de Combustíveis; Previsão; Transbordamento de Volatilidade.

Código JEL: C32; C53; Q41

LIST OF FIGURES

Figure 0.1 – Freight Transport Matrix for Countries with Territorial Dimensions Similar to Brazil. 16
Figure 1.1 – Gasoline time series prices in Brazil 37
Figure 1.2 – Diesel time series prices in Brazil 38
Figure 1.3 – ARIMA encompassing test for gasoline 49
Figure 1.4 – ES encompassing test for gasoline 50
Figure 1.5 – SSA encompassing test for gasoline 51
Figure 1.6 – ARIMA encompassing test for diesel 52
Figure 1.7 – ES encompassing test for diesel 53
Figure 2.1 – Growth Rate of Fuel Prices 68
Figure 2.2 – Conditional Correlations between Gasoline, Diesel and Oil Prices . . . 73
Figure 2.3 – Conditional Variance of Gasoline 76
Figure 2.4 – Conditional Variance of Diesel 77
Figure 2.5 – Gasoline price historical 78
Figure 2.6 – Diesel price historical 79
Figure 2.7 – Conditional Correlations between Gasoline, Diesel and Oil Prices in the IPP Period 82
Figure 2.8 – Conditional variance of gasoline during the IPP 84
Figure 2.9 – Conditional variance of diesel during the IPP 84
Figure 3.1 – Gasoline price in Brazilian states 108
Figure 3.2 – Ethanol price in Brazilian states 109
Figure 3.3 – Diesel price in Brazilian states 110
Figure 3.4 – LGP price in Brazilian states 111

LIST OF TABLES

Table 1.1 – Literature Review 26
Table 1.2 – Descriptive statistics for gasoline across Brazilian states 39
Table 1.3 – Descriptive statistics for Diesel across Brazilian states 40
Table 1.4 – RMSE for the estimated models in the in-sample period for gasoline and diesel 41
Table 1.5 – RMSE values for the models estimated based on the out-of-sample data for gasoline 43
Table 1.6 – RMSE values for the estimated models based on the out-of-sample period for diesel 45
Table 2.1 – Summary of the Literature Review 61
Table 2.2 – Prior distributions 66
Table 2.3 – Description of the variables 67
Table 2.4 – Constant variance components associated with matrix C for gasoline and diesel 69
Table 2.5 – Coefficients of the A matrix for gasoline and diesel 70
Table 2.6 – Coefficients of the B matrix for gasoline and diesel 71
Table 2.7 – Constant variance components associated with matrix C for gasoline and diesel during the IPP 80
Table 2.8 – Coefficients of the A matrix for gasoline and diesel during the IPP . . . 81
Table 2.9 – Coefficients of the B matrix for gasoline and diesel during the IPP . . . 81
Table 3.1 – Literature Review 91
Table 3.2 – Descriptive statistics of the first log-differences gasoline prices by state . 112
Table 3.3 – Descriptive statistics of the first log-differences ethanol prices by state . 113
Table 3.4 – Descriptive statistics of the first log-differences diesel prices by state . . 114
Table 3.5 – Descriptive statistics of the first log differences of LPG prices by state . 115
Table 3.6 – Forecasting performance for the models considering four fuels and twenty-five Brazilian states 117
Table 3.7 – Forecasting performance for the models considering four fuels and the southeast region 118
Table 3.8 – Forecasting performance for the models considering four fuels and the Midwest region 118
Table 3.9 – Forecasting performance for the models considering four fuels and the Northeast region 118
Table 3.10–Forecasting performance for the models considering four fuels and the North region 119
Table A.1–Parameters of the Gasoline Forecasting Models 131
Table A.2–Parameters of the diesel Forecasting Models 132

Table B.1 – Matrix C Coefficients for Imported Gasoline 133
Table B.2 – Matrix A Coefficients for Imported Gasoline 133
Table B.3 – Matrix B Coefficients for Imported Gasoline 133
Table B.4 – Forecasting performance for the models considering four fuels and twenty-five Brazilian states, $\hat{u}_1 = 2$ and $\hat{u}_2 = 3$ 133
Table B.5 – Forecasting performance for the models considering four fuels and twenty-five Brazilian states, $\hat{u}_1 = 2$ and $\hat{u}_2 = 4$ 133
Table B.6 – Forecasting performance for the models considering four fuels and twenty-five Brazilian states, $\hat{u}_1 = 2$ and $\hat{u}_2 = 5$ 134
Table C.1 – Forecasting performance for the models considering four fuels and twenty-five Brazilian states, $\hat{u}_1 = 2$ and $\hat{u}_2 = 3$ 135
Table C.2 – Forecasting performance for the models considering four fuels and twenty-five Brazilian states, $\hat{u}_1 = 2$ and $\hat{u}_2 = 4$ 135
Table C.3 – Forecasting performance for the models considering four fuels and twenty-five Brazilian states, $\hat{u}_1 = 2$ and $\hat{u}_2 = 5$ 135

LIST OF ABBREVIATIONS AND ACRONYMS

ANP	National Agency of Petroleum, Natural Gas and Biofuels
ARIMA	Autoregressive Integrated Moving Average
ANTF	National Association of Railway Transport Operators
ANFAVEA	National Association of Motor Vehicle Manufacturers
BEKK	Baba–Engle–Kraft–Kroner Model
CCC	Constant Conditional Correlation
CNT	National Confederation of Transport
DCC	Dynamic Conditional Correlation
EMAR	Envelope Matrix Autoregressive
ES	Exponential Smoothing
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
IPCA	Extended National Consumer Price Index
IPP	Import Parity Price
LPG	Liquefied Petroleum Gas
MAR	Matrix Autoregressive
MGARCH	Multivariate Generalized Autoregressive Conditional Heteroskedasticity
MSE	Mean Squared Error
PROALCOOL	National Alcohol Program
PNPB	National Program for the Production and Use of Biodiesel
PETROBRAS	Petróleo Brasileiro
RMSE	Root Mean Squared Error
SSA	Singular Spectrum Analysis
VAR	Vector Autoregression

CONTENTS

List of abbreviations **12**

General Introduction **15**

I The historical use of fuels in Brazil 15

II Research Objectives and Main Results 20

1 Essay 1 - Forecasting Gasoline and Diesel At-the-Pump Prices Across Brazilian States Using a Nonparametric Approach **22**

1.1 Introduction 23

1.2 Literature Review 25

1.3 Empiric Strategy 29

1.3.1 Singular Spectrum Analysis 30

1.3.1.1 Short Description about the SSA 31

1.3.2 Selection of the Window Length (L) and Eigentriples 32

1.3.3 SSA Forecasting 33

1.3.4 Accuracy Test 35

1.4 Data Base 36

1.5 Forecasting Results 41

1.5.1 Comparison of the accuracy of the forecasts 41

1.5.2 Encompassing Test 47

1.6 Policy Implications 54

1.7 Conclusion 55

2 Essay 2 - Volatility spillovers between oil and gasoline/diesel in Brazil: A Bayesian approach **57**

2.1 Introduction 57

2.2 Fuel Volatility and the Oil Market 59

2.3 Empirical Strategy 64

2.3.1 Multivariate GARCH Model 64

2.4 Database 67

2.5 Results 68

2.5.1 Multivariate Model Estimation 68

2.5.2 Estimations for Gasoline and Diesel during the IPP Period by Petrobras 78

2.6 Conclusions 84

3 Essay 3 - Fuel Price Forecasting in Brazil under Cross-State Dependence: An EMAR Framework **87**

3.1 Introduction 87

3.2 Literature Review on fuel forecasting 90

3.3 Empirical Strategy 98

3.3.1 Matrix Autoregressive Models 99
3.3.1.1 One-Sided Matrix Autoregressive Models 101
3.3.2 Vectorized Matrix-Variate Time Series 102
3.3.3 Envelope-Based Matrix Autoregressive Analysis 102
3.3.3.1 Envelope Matrix Autoregressive (EMAR) Models . . . 103
3.3.3.2 Envelope Formulation for EMAR models 104
3.3.4 EMAR Model estimation 105
3.3.5 Accuracy test 107
3.4 Data and Summary Statistics 107
3.5 Results 115
3.5.1 Estimation for Brazil 115
3.5.2 Estimation for Brazilian regions 117
3.6 Conclusion 119
Final Consideration 121
REFERENCES 123
APPENDIX A – Tables A 131
1.1 Appendix 131
APPENDIX B – Tables B 133
APPENDIX C – Table C 135

General Introduction

I The historical use of fuels in Brazil

Fuel consumption in Brazil plays a significant role globally. In terms of oil consumption, the country ranks 10th, with 235 thousand barrels per day. For gasoline, Brazil consumes 1,004 thousand barrels per day, making it the third largest gasoline consumer in the world (The Global Economy, 2022b; The Global Economy, 2022a). Finally, Brazil is the eighth largest oil consumer globally (Statista, 2024). Thus, these numbers indicate that fuels are highly relevant in Brazil, with many different uses.

In Brazil, there are four important fuels: gasoline, ethanol, diesel, and LPG. Gasoline, diesel, and LPG are derived from oil, and ethanol is a renewable fuel from sugar cane or corn.

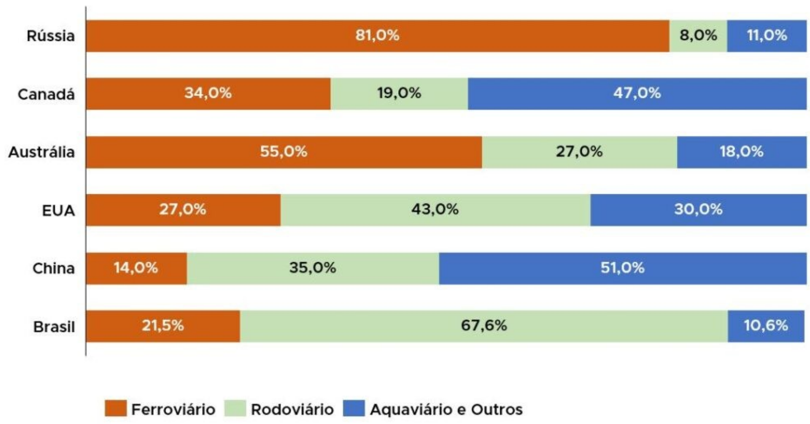
Ethanol rose in Brazil as an answer to the first oil shock in 1973. From this year, the Brazilian government created the PROALCOOL. Launched on November 14, 1975, the program aimed primarily to develop combustion engine technology that did not rely on petroleum as the main fuel, but rather on sugarcane and its by-product, alcohol (later known as ethanol) (National Association of Automobile Manufacturers, 2024b),. Although the PROALCOOL program faced challenges typical of new technologies, particularly regarding cold-start performance, the first alcohol-powered engines had significant difficulties starting in low temperatures. By the late 1970s, the first results of this new technology emerged, with the introduction of the Fiat 147 running on alcohol, the world's first ethanol-powered car (National Association of Automobile Manufacturers, 2024a).

Following the creation of PROALCOOL, the Brazilian sugarcane–ethanol market faced challenges due to the stabilization of oil prices during the 1980s and 1990s, coupled with mechanical issues in alcohol-powered vehicles. However, the Brazilian market experienced a surge of innovation in the 2000s, with the launch of the first large-scale flex-fuel vehicle in 2003 (National Association of Automobile Manufacturers, 2024a). As a result, Brazilian consumers were able to choose which fuel to use in their vehicles, bringing greater dynamism to the energy sector.

In addition to the introduction of flex-fuel engines in the vehicle fleet as an alternative to reduce dependence on gasoline, the Brazilian government has promoted an increase in the proportion of ethanol blended into gasoline. Currently, ethanol accounts for 27% of the gasoline mix in the country (Civil House, 2015). This blending policy aims not only to mitigate the impact of oil price shocks on gasoline prices but also to reduce CO_2 emissions from motor vehicles.

Regarding diesel, there are some features that must be explained. First, most of the matrix transportation in Brazil is by highway (70%), as shown in Figure 0.1; second, Brazil has a fleet of 129 million vehicles (Ministry of transport, 2025); which is responsible for transporting goods, services, and people throughout the country. This number is a record of the Brazilian fleet and reflects the strength of vehicles in people's daily lives. Furthermore, it is important to highlight that there are 3 million trucks in Brazil.

Figure 0.1 – Freight Transport Matrix for Countries with Territorial Dimensions Similar to Brazil.



Source: National Association of Rail Transporters, where the orange color is railroad, the grey color is highway, and the blue color is waterway, and others.

Thus, as shown in Figure 0.1, among countries with large territorial dimensions, Brazil relies primarily on road transport as its main freight mode. Consequently, fluctuations in fuel prices directly affect the nation's logistical capacity. Therefore, sharp changes in the prices of diesel, ethanol, and gasoline have a direct impact on the prices of a wide range of goods.

Furthermore, according to the Business Profile Survey conducted by National Transport Confederation (2021), Brazil has 266,000 freight transport companies, 847,000 independent freight carriers, and 519 road freight transport cooperatives, which together account for approximately 2.5 million freight vehicles. This figure is 70% higher than that recorded fifteen years ago.

Moreover, according to the National Transport Confederation (2021), 99.89% of freight transport companies in Brazil use diesel as fuel for their trucks. Consequently, a constant concern within the sector is managing the costs arising from fluctuations in diesel prices. As an alternative, many companies choose to operate their own refueling facilities in order to negotiate large volumes of diesel and ensure fuel quality. This strategy is adopted by 55% of freight transport companies in the country.

Being a possible solution, Brazil conducted studies on the feasible application of biodiesel to transform the country's energy matrix. However, it was only in 2005, with the enactment of Law No. 11,097, that a 5% biodiesel blend in diesel was officially mandated.

This law was also part of the PNPB, which aims to restructure the diesel energy sector in the country. By 2025, the biodiesel share in Brazilian diesel increased to 15% of the blend (Ministry of Mines and Energy, 2025a).

In addition to blending ethanol into Brazilian gasoline to mitigate potential price increases from oil shocks, the Brazilian government implemented other measures to stabilize gasoline prices. In 2010, as the main shareholder of Petrobras, the government initiated a series of market interventions. These interventions included price freezes, particularly on gasoline, diesel and LPG (Moraes; Bacchi, 2014; Almeida; Oliveira; Losekann, 2015). According to Almeida, Oliveira and Losekann (2015), since 2011, the Brazilian government has noted that domestic fuel demand increased, so there is a need to import more fuel from overseas. The government solution was to decouple international fuel prices from domestic prices. So, the government imported fuels at international prices and sold them only at national prices between 2011 and 2014. Therefore, until 2014, the loss from this fuel policy cost 21 billion reais.

Due to Petrobras (Petroleo Brasileiro) not passing on the gasoline price adjustments to consumers, the government at the time decided to implement a price increase at the beginning of 2013, thereby removing the artificially maintained fuel price. One of the reasons for the adjustment was the company's severely strained financial and operational conditions resulting from the lack of price updates. However, following the gasoline price adjustment in early 2013, a significant increase in inflation was observed (Cagnin et al., 2013).

The government's control over gasoline prices also reflects the weight that gasoline holds within the Broad National Consumer Price Index (IPCA), from acronym in Portuguese, where gasoline accounts for 5.82% of the total, according to the Brazilian Institute of Geography and Statistics (IBGE), from acronym in Portuguese, (Brazilian Institute of Geography and Statistics, 2024). Furthermore, considering the nine sectors that make up the IPCA, the transportation sector carries a weight of 20.70%, making it the second-largest component of the index's basket of goods. This explains the government's concern in controlling or mitigating adverse effects that may impact fuel prices.

In case of LPG, this fuel has several applications for domestic and industrial use. According to Sindicato Nacional das Empresas Distribuidoras de Gás Liquefeito de Petróleo (2023), Brazilian industry uses LPG in the following sectors: food and beverages, paper and pulp, asphalt, ceramics, packaging, pharmaceuticals, plastics, steel, glass, chemicals and metallurgy, cosmetics, sanitizing products, and aerosols. While LPG is mainly used in industry, fuels such as gasoline, diesel, and ethanol are used for the transport of goods and services throughout Brazil. Almeida, Oliveira and Losekann (2015) highlights that from 2011 to 2014, the Brazilian government avoided adjusting the price to keep inflation low, since LPG is used to cook food and other important goods in the IPCA index.

Regarding the average price of diesel, Petrobras also did not pass on the necessary price increases to final consumers. Diesel is primarily used in cargo trucks, and between 2011 and 2014, Brazil experienced strong growth in freight transport, particularly in the agricultural sector, resulting in high demand for trucks. During this period, the government-maintained diesel prices through successive freezes in order to meet the transportation needs of the agricultural sector (Almeida; Oliveira; Losekann, 2015).

With respect to diesel, this practice of price freezes and abrupt adjustments has a clear rationale. As Petrobras is the leading company in the sector and is partly government-owned, when international diesel prices surge, the government instructs the company not to pass these increases on to domestic prices. This directly affects private diesel refiners and importers, who, in order to avoid losses, reduce their production. To prevent domestic diesel shortages, Petrobras imports diesel at international prices from the private sector and sells it domestically at artificially lower prices. This situation creates two main issues: a divergence between domestic and international diesel prices and a deterioration of Petrobras' cash flow (Gauto; Delgado; Couto, 2021).

In the case of gasoline, Brazil imports a portion of gasoline from the international market. This is necessary because the crude oil extracted by Petrobras domestically does not possess the full characteristics required for refining in Brazilian refineries. Consequently, it is necessary to import a certain quantity of international gasoline to achieve the proper blend for refining (Delgado; Gauto, 2021). Therefore, when Petrobras imported gasoline at international prices, the pricing policy between 2011 and 2014 created a lag in passing these international costs on to domestic consumers.

Thus, even though Brazil has substantial oil reserves (pre-salt), the country still needs to import diesel and gasoline from foreign suppliers, with Brent crude serving as the benchmark for oil import prices (Delgado; Gauto, 2021). Finally, Brazilian oil can be refined into fuels.

Between 2010 and 2016, fuel prices in Brazil experienced significant fluctuations. Once that, between 2010 and 2014, fuel prices remained stable; the main reason was that the government decided not to pass along international increases in oil and derivatives prices. For example, gasoline, diesel, and ethanol started at BRL 2,58, 2,00, and 1,89, respectively, in 2010. At the end of 2014, the prices of gasoline, diesel, and ethanol were BRL 3,03, 2,56, and 2,04, respectively (National Agency of Petroleum, Natural Gas and Biofuels, 2024). In 2015, the average price of diesel started the year at BRL 2.61 per liter and ended at BRL 2.98 per liter, following the fuel price survey (National Agency of Petroleum, Natural Gas and Biofuels, 2024). Throughout the year, there were multiple truckers' protests and strikes aimed at freezing diesel prices, among other demands from the sector. These price freezes resulted in a loss of BRL 34.8 billion for Petrobras during the year (Economática, 2015).

For the year 2016, political changes occurred in Brazil, including a presidential impeachment, which led to the installation of a new government. The new administration implemented a new pricing policy at Petrobras, known as the Import Parity Pricing (IPP). This policy lasted from October 2016 to May 2023. The objective of this policy was to align domestic fuel prices with international prices, thereby preventing price distortions and relieving pressure on the company's cash flow. Given the necessity to import certain quantities of crude oil to create the appropriate blend and refine fuels domestically, any fluctuations in international oil prices were directly reflected in domestic fuel prices.

Nevertheless, this policy of aligning domestic fuel prices with international prices led to constant fluctuations in fuel costs. This created a flow of volatility that was ultimately passed on to consumers, whether in the form of price increases or decreases. According to the Fuel Price Survey Series compiled by National Agency of Petroleum, Natural Gas and Biofuels (2024), during 2017, as domestic prices followed international levels, the average price of regular gasoline in Brazil decreased from BRL 3.76 in January to BRL 3.53 in July, before rising again to BRL 3.78 in August. This sustained volatility had a significant impact on domestic consumers, raising concerns about the future prices of gasoline and other petroleum-based fuels.

Concerns over the frequent fluctuations in fuel prices, particularly diesel, have led to two major strikes by transport companies. Known as the "truckers' strikes", which occurred in 2015 and 2018, these actions were primarily mobilized by independent truck drivers, who are more sensitive to changes in fuel prices. Although the strikes were of short duration, they significantly impacted the country, due to the disruption of supplies of various goods, particularly in areas distant from major urban centers (Alves et al., 2018; Barra; Silva; Silveira, 2020; Silva; Chaves, 2020).

In particular, during the 2018 strike, the stoppage lasted 11 days, from May 20 to May 30. According to the Central Bank of Brazil (2018), this period resulted in an economic loss of 15 billion BRL. In addition, there was a rapid acceleration of food prices in the country, with vegetables and milk experiencing particularly significant increases. Vegetable prices rose by 45% over the course of the strike. Furthermore, the GDP was affected: prior to the strike, GDP growth was expected at 2.5%, while after the strike, the expected GDP growth was 1.6% (Central Bank of Brazil, 2018).

In order to reduce or smooth the frequent increases in fuel prices, particularly diesel, in 2018 the government issued Provisional Measure No. 838/2018, allocating BRL 9.5 billion in subsidies to diesel producers and importers in the country. The purpose of this subsidy was to prevent abrupt changes in diesel prices (National Agency of Petroleum, Natural Gas and Biofuels, 2018).

During the COVID-19 pandemic, industrial activities were largely halted, leading to price increases across various goods, including fuels. In response, the Brazilian government, between 2020 and 2022, implemented a series of tax adjustments to mitigate rising petroleum product prices. In March 2021, the government reduced the PIS/PASEP contribution to zero for diesel and cooking gas in order to contain prices, through Decree No. 10,638 (Official Gazette of the Union, 2021). In 2022, the government established a maximum cap for the ICMS tax (between 17% and 18%) on fuels, aiming to reduce gasoline and diesel prices nationwide, through Complementary Law 194/2022 (Official Gazette of the Union, 2022).

These measures resulted in a reduction in fuel prices, particularly gasoline. According to the Fuel Price Survey held by National Agency of Petroleum, Natural Gas and Biofuels (2024), between June and July 2022, the average price of gasoline dropped significantly from BRL 7.25 in June to BRL 6.05 in July. This demonstrates that the ICMS price cap for gasoline effectively contributed to lowering fuel prices in the country.

Finally, in 2023, Brazil came under a new government, which decided to end Petrobras' Import Parity Pricing (IPP) policy in May 2023. The termination of the IPP aimed to mitigate the impact of international oil prices on domestic fuel prices, allowing for fewer price adjustments and promoting greater price stability in Brazil.

Therefore, from the historical perspective of the Brazilian fuel sector, there has been a consistent pattern of government intervention aimed at stabilizing and/or controlling fuel prices for various objectives, primarily to manage inflation. Price controls have also sought to prevent potential future strikes, such as the one that occurred in May 2018 in response to rising diesel prices.

II Research Objectives and Main Results

Thus, based on the historical use of fuels in Brazil, this study has two main goals: first, to forecast the fuel prices considering univariate (first essay) and multivariate (third essay), and the second is to examine how the international oil price volatility affects domestic fuels (gasoline and diesel).

The first essay had the main goal of predicting retail prices at the pump for gasoline and diesel in each Brazilian state. Therefore, the forecast provided a broad overview of the future of gasoline and diesel prices across Brazilian states. Moreover, the first essay ensured the predictability of fuel prices.

In the second essay, the main goal is to examine how oil price volatility will affect national domestic fuel prices (gasoline and diesel). Therefore, the second essay sought to understand how oil price shocks can affect national retail prices for gasoline and diesel.

Finally, in the third essay, the main goal is to forecast fuel prices for each Brazilian state, accounting for a multivariate context. So, the difference between the first and third essays lies in their approach: while the first essay predicts gasoline and diesel prices using a univariate approach, the third uses all fuels (gasoline, diesel, ethanol, and LPG) to forecast their prices. Therefore, in the third essay, the main idea is to show how gasoline prices in the state can help to forecast diesel prices in other states, for example.

1 Essay 1 - Forecasting Gasoline and Diesel At-the-Pump Prices Across Brazilian States Using a Nonparametric Approach

Abstract

The present study aims to forecast gasoline and diesel at-the-pump prices across Brazil's 27 states. Forecasting the prices of these fuels provides a broad overview of the price dynamics that may occur across Brazil, given the country's vast territory. To offer such an overview of fuel price dynamics, a non-parametric approach—Singular Spectrum Analysis (SSA)—was applied, with the objective of assessing whether this method can achieve better predictive performance compared to traditional forecasting methodologies, such as ARIMA, Exponential Smoothing, and Neural Networks, which were used as benchmarks in this study. The analysis covers the period from July 2001 to August 2023. The results present some conclusions: 1) In-sample, the SSA was the best predictor model to forecast gasoline and diesel, which was the best in 60% of Brazilian states for gasoline and 57.70% for diesel; 2) Out-of-Sample, the ARIMA model was the best predictor model in 96.30% of the cases for gasoline and 99.03% for diesel. However, according to the Modified Diebold-Mariano test, only the 1-month-ahead forecasts were statistically different. Finally, the Encompassing test showed that, for a 12-month horizon, the SSA model can be useful for forecasting fuel prices, and in this horizon, the forecast was the only one for which SSA outperformed the other models (Roraima state to gasoline).

Resumo

O presente estudo tem como objetivo prever os preços da gasolina e do diesel ao consumidor nos 27 estados brasileiros. A previsão desses combustíveis permite compreender a dinâmica dos preços em todo o território nacional, considerando a dimensão territorial do Brasil. Para isso, foi aplicada a abordagem não paramétrica Singular Spectrum Analysis (SSA), avaliando seu desempenho em comparação com métodos tradicionais de previsão, como ARIMA, Suavização Exponencial e Redes Neurais. A análise compreende o período de julho de 2001 a agosto de 2023.

Os resultados mostram que, na amostra (in-sample), o SSA apresentou o melhor desempenho em 60% dos estados para gasolina e em 57,70% para diesel. Fora da amostra (out-of-sample), o ARIMA foi superior em 96,30% dos casos para gasolina e em 99,03% para diesel. Entretanto, segundo o teste Modified Diebold-Mariano, diferenças estatisticamente significativas foram observadas apenas para previsões de um mês à frente. Por fim, o teste de Encompassing indicou que, para previsões de 12 meses, o SSA pode fornecer informações complementares, superando os demais modelos apenas na previsão do preço da gasolina em Roraima.

1.1 Introduction

According to Instituto Brasileiro de Geografia e Estatística (2025), Brazil has 8.5 million square kilometers and relies mostly on road transport. This confidence in road transport has a main concern: fluctuations in fuel prices, particularly oil prices. Therefore, slight changes in oil and its derivatives may have a considerable impact on Brazilian lives.

Moreover, crude oil is the most traded commodity in the world, according to the Intercontinental Exchange (2024), indicates that a certain degree of uncertainty exists regarding the fuel prices to be observed in the future. Consequently, price predictability is a central issue to ensure greater transparency within the petroleum and derivatives sector.

Price predictability is also present in contracts established between producers and distributors of petroleum and its derivatives. The National Agency of Petroleum, Natural Gas and Biofuels (2018) issued a technical note to ensure that information regarding future prices of products traded under these contracts is available to the contracting parties. This measure aims to reduce uncertainty in petroleum product transactions, thereby contributing to price stability in the present and future.

The predictability requirements imposed by the ANP on market participants can also be extended to end consumers. In this context, predictability for the end consumer is achieved through predictive models. The use of these models to forecast future fuel prices in Brazil allows consumers to be informed about the prices that will be charged in the future.

According to the literature on fuel prices, the following studies are highlighted:García-Martos, Rodríguez and Sánchez (2013), which predict the oil, natural gas, and coal in Spain; Baumeister, Kilian and Lee (2017) predict the gasoline price in the USA, considering extensive methods of forecasting to obtain the best forecasting. In Brazil, Silva (2019) predicted ethanol prices in the state of São Paulo for 2002-2017, and Queiroz et al. (2022) predicted diesel prices nationwide. Finally, oil price forecasting has been extensively analyzed in several studies, for instance: Abramson (1991); Ye, Zyren and Shore (2005); Safari and Davallou (2018); Rizvi, Syed and Qureshi (2022); Chen, Lu and Ma (2022) and Anjos et al. (2023).

Therefore, the objective of this study is to forecast gasoline and diesel prices across Brazilian states. To achieve this goal, three forecasting models were employed: Singular Spectrum Analysis (SSA), ARIMA, and Exponential Smoothing (ES). At the end of the forecasting exercise, the study aims to determine which method provides the most accurate predictions of fuel prices across Brazil's states.

The choice of these methods is justified by the following reasons: 1) ARIMA and Exponential Smoothing (ES) are widely accepted in the forecasting literature, particularly in the context of petroleum prices, and have been applied in a wide range of scenarios, as documented by García-Martos, Rodríguez and Sánchez (2013), Alves et al. (2018), Safari and Davallou (2018), Chen, Lu and Ma (2022), Anjos et al. (2023), among others; 2) The SSA model was selected due to its intrinsic characteristics: it is a nonparametric model capable of decomposing a time series into subseries, such as trend, seasonality, and noise.

These features of SSA have been shown in the literature to provide better forecasting performance compared to models such as ARIMA and ES, for example, in Hassani et al. (2015) and Hassani et al. (2017), the forecasting context was the tourism sector. In the first, the country analyzed was the USA from 1996 to 2012, and the authors used ARIMA, Exponential Smoothing, Neural Network and SSA to forecast the international tourist arrivals. The results indicated that SSA was 50% more accurate in forecasting than other methods. The second study considered the international tourist arrivals in 9 countries of Europe (Austria, Cyprus, Germany, Greece, the Netherlands, Portugal, Spain, Sweden, and the United Kingdom) from 2000 to 2013. The following methods were used: ARIMA, Neural Networks, Exponential Smoothing, ARFIMA, TBATS, Recurrent, and Vector SSA. The results indicated that no single model stood out in terms of predictive performance. However, SSA, ARIMA, and TBATS exhibited the best forecasting results when evaluated using RMSE.

Additionally, Hassani, Heravi and Zhigljavsky (2009) applied the method to forecast production in eight industrial sectors across three European countries. Considering these three SSA studies, it was observed that the method produced forecasting results that were equal to or superior to those of the comparison methods in most cases, indicating that SSA may be highly valuable for forecasting in the Brazilian fuel sector.

This study contributes to the Brazilian literature on fuel price forecasting in three ways. First, it employs a nonparametric methodology (SSA) to forecast future fuel prices. Second, the study contributes by producing forecasts for the monthly prices of the two most important fuels in the country (gasoline and diesel), taking into account end-consumer prices (i.e., at-the-pump prices). Finally, the study provides forecasts for all twenty-seven states of Brazil, offering a comprehensive overview of the expected monthly behavior of fuel prices. This overview highlights the potential differences in gasoline and diesel prices across regions, from north to south of the country.

This paper is structured into seven sections. In addition to this introduction, Section 2 provides a literature review on the forecasting of gasoline, diesel, and crude oil prices, covering both the international and domestic contexts. Section 3 presents the predictive methods used to forecast the monthly prices of gasoline and diesel across Brazilian states, as well as the accuracy tests employed in this study. Section 4 describes the data set used, along with the procedures applied for data processing. Section 5 presents the state-level results obtained from the predictive methods for gasoline and diesel. Section 6 discusses the results and their potential implications for public policies regarding gasoline and diesel forecasting. Finally, Section 7 outlines the main conclusions of the study, as well as its limitations and contributions.

1.2 Literature Review

Table 1 present the studies on fuel price forecasting, which are divided in two branches: (i) forecasting in the global petroleum and fuel sector; (ii) forecasting in the Brazilian petroleum and fuel sector.

Considering the first branch, the forecasting of oil and its derivatives is a theme of global interest, being addressed in different countries and under various forecasting scenarios. Thus, oil price forecasting has been studied in numerous studies across different countries over the years.

Regarding the U.S. market, Abramson (1991) forecasted oil prices between 1989 and 1990, while Ye, Zyren and Shore (2005) analyzed the period from 1992 to 2003. More recently, Horák and Jannová (2023) predicted Brent crude oil prices from 2020 to 2022. Oil price forecasting was also addressed in the Spanish market, as in the study by García-Martos, Rodríguez and Sánchez (2013) between 2009 and 2011. Considering OPEC member countries (Organization of the Petroleum Exporting Countries), (Safari; Davallou, 2018) forecasted prices from 2003 to 2016. Additionally, Chen, Lu and Ma (2022) conducted oil price forecasts considering ports in three countries (the Netherlands, Singapore, and the United Arab Emirates) between 2005 and 2021.

Still within the first branch of analysis, some studies focused on fuels derived from petroleum. Tipyan and Leenawong (2009) analyzed gasoline and diesel price forecasting in Thailand between 2002 and 2018. Asuamah and Ohene-Manu (2015) forecasted premium fuel prices in Ghana from 2000 to 2011. Finally, Baumeister, Kilian and Lee (2017) conducted an extensive forecasting analysis using various methods to obtain the best predictions for gasoline prices in the United States.

The second branch of forecasting analysis focuses on Brazil, with studies examining both oil and fuel prices. Anjos et al. (2023) analyzed WTI crude oil prices from 1986 to 2022, employing ARIMA and neural networks as predictive methods. Silva (2019) examined ethanol prices in the state of São Paulo, using ARIMA between 2002 and 2017. Finally, Queiroz et al. (2022) analyzed diesel prices using a fuzzy forecasting approach from 2013 to 2020.

Table 1.1 – Literature Review

Study	Time	Variables	Countries	Methodologies	Main results
Forecasting in the oil and fuels sector around the world					
Abramson (1991)	1989–1990 (Quarterly)	Over 30 variables, mainly regarding to oil.	United States.	Use of the Atlantic Richfield Company (ARCO), which is a belief-based network.	The model was able to accurately forecast the approximate price of oil in 60% of the scenarios.
Ye, Zyren & Shore (2005)	January/1992–April/2003 (Monthly)	Oil WTI, Industrial Activity of the OECD.	United States	Naive, RSTK, MALT model.	The RSTK model provided best forecasts of WTI prices at 1-, 2-, 3-, and 6-month horizons.
Tipyan & Leenawong (2009)	01/2002–12/2008 (Monthly).	Gasoline and diesel.	Thailand.	Naive, Exponencial Smoothing, Multiple Regression, Forecasting Combination and Decomposition method.	The forecasting combination method achieved the best predictive performance, according to the RMSE.
García-Martos, Rodríguez & Sánchez (2013)	06/March/2009–29/April/2011 (Daily)	Oil, natural gas and coal.	Spain	ARIMA and VARMA.	For forecast horizons up to 15 days, the VARMA model outperformed ARIMA on average in predicting prices.
Asumah & Ohene-Manu (2015)	January/2000–December/2011 (Monthly)	Premium fuel (High Octane rating).	Ghana	ARIMA (1,1,0), ARIMA (0,1,1) and ARIMA (1,1,1).	The ARIMA (1,1,1) model achieved the best forecasts according to the RMSE. However, the model tended to underestimate prices compared to actual values.

Continues on next page

Study	Time	Variables	Countries	Methodologies	Main results
Baumeister, Kilian & Lee (2017)	January/1992–March/2014 (Monthly)	Gasoline	United States	ARMA (1,1); IMA (1); ARIMA (1,1,1); AR (12); Bayesian AR (12); AR (AIC); Bayesian AR (AIC); Exponential Smoothing, UC-SV and no-change forecast.	The UC-SV model outperformed the comparison model (no-change forecast) at forecast horizons of 3, 6, 9, 12, 15, 18, 21, and 24 months.
Safari & Davallou (2018)	January/2003–September/2016 (Monthly)	Oil	Countries members of the OPEC (Organization of the Petroleum Exporting Countries).	ARIMA, Neural Networks, and Exponential Smoothing, with the use of hybrid models based on these three methods.	The best forecasting method was a hybrid model (PHM), according to the RMSE. The model incorporates time-varying weights and was applied for a one-month-ahead forecast only.
Rizvi, Syed & Quresh (2022)	13/11/2018 – 28/11/2018/ daily	Oil	-	Causal-effect neural network, ARIMA, long short-term memory (LSTM) neural network, and Gated Recurrent Unit (GRU) neural network.	The causal-effect neural network demonstrated superior forecasting performance compared to the other methods.
Chen, Lu & Ma (2022)	01/01/2005 – 31/12/2021 daily	Oil and Bunker fuel	Rotterdam Harbor (The Netherlands); Singapore Harbor and Fujairah Harbor (United Arab Emirates).	ARMA, ARMAX, VAR and VEC.	It was found that the VEC model achieved the best predictive performance compared to the other models, particularly at 1- and 4-day-ahead horizons, according to the RMSE.

Continues on next page

Study	Time	Variables	Countries	Methodologies	Main results
Forecasting for fuels and oil sector in Brazil					
Silva (2019)	11/2002 – 12/2017 monthly.	Hydrated Ethanol	São Paulo State/ Brazil	ARMA (1,1); ARMA (1,2); ARMA (2,1); ARMA (2,2); ARMA (1,3) and ARMA (3,1).	The ARMA (1,1) model achieved the best forecasting performance among the models used, exhibiting the lowest RMSE.
Queiroz, Alves, Oliveira & Aguiar (2022)	12/2012 – 05/2020 monthly	Diesel S10 and S500	Brazil	Fuzzy models to forecasting.	Among the proposed models, VS-ePL-KRLS achieved the lowest RMSE values and was the best predictive method for diesel prices in Brazil.
Dos Anjos, Besarria, Medeiros, De Jesus & De Araújo (2023)	01/1986 – 01/2022	Oil WTI	-	Neural Network (LSTM), ARIMA and hybrid models between both.	Using a hybrid model combining neural networks and ARIMA proved to be the best predictive method for forecasting oil prices, according to the RMSE.

Source: Elaborated by the authors.

Considering fuel forecasting for Brazil, this paper presents some differences: first, it considers all Brazilian states in the analysis, so the forecasting was performed for each state rather than for a single state or specific region, as in Silva (2019) and Queiroz et al. (2022). Second, the paper used a nonparametric approach to forecast fuel prices in Brazil; this technique does not require the assumptions made by parametric techniques (such as stationarity, functional form, and others). The third and last paper predicted the prices at the pump for gasoline and diesel, the country's main fuels. Therefore, this paper gathered various methods to forecast gasoline and diesel prices for each Brazilian state, offering a new panorama for forecasting Brazilian fuels.

Furthermore, a consensus emerged in the literature regarding the methods employed for forecasting. It was observed that the ARIMA model was used in 9 out of the 16 studies reviewed, making it the most frequently applied method. This indicates that ARIMA is a reliable and widely accepted predictive tool. Neural networks, in their various forms, also stood out, appearing in 8 studies[1]. Finally, exponential smoothing was employed in 6 studies. These three methods are therefore the most commonly used for time series forecasting.

Finally, the RMSE was used in 12 of the 16 studies reviewed, underscoring its prevalence as a criterion for evaluating various modeling methods. This widespread application contributed to its designation as the principal measure of accuracy.

1.3 Empiric Strategy

For this study, the forecasting methods considered were ARIMA, Holt's Exponential Smoothing, and Singular Spectrum Analysis (SSA). The selection of these methods was based on two reasons: i) ARIMA and Exponential Smoothing were used as a benchmark to forecast fuels, according to literature review; ii) the studies that used SSA (Hassani, Heravi and Zhigljavsky (2009), Hassani et al. (2015), Hassani et al. (2017)) used ARIMA and Exponential Smoothing as a benchmark. In those studies, SSA demonstrated superior forecasting performance compared to the other models; therefore, it was decided to test whether SSA exhibits superior predictive capacity relative to these methods when applied to fuel price data in Brazil. Moreover, the use of these models is supported by the literature on fuel price forecasting, as presented in Table 1 of the literature review section.

The ARIMA model was initially proposed by Box and Jenkins (1970) with the objective of integrating autoregressive (AR) and moving average (MA) models within

[1] In this study, a neural network was estimated using the R package (nnet); however, the package failed to capture the nonlinearities in the time series, and the results were poor. n addition, the in-sample Once that, normally Neural Networks involve a large number of parameters, particularly when the number of neurons or layers increases, which can lead to overfitting in small samples and poor out-of-sample forecasting performance (ZHANG, 2003; HYNDMAN; ATHANASOPOULOS, 2018).

a single forecasting framework. This model was designed for the forecasting of non-stationary time series, given the inability of ARMA models to handle series with a unit root. Consequently, the ARIMA model allows the analysis of time series that do not exhibit explosive behavior, while maintaining some homogeneity in their non-stationary behavior (Morettin; Toloi, 2006).

A base ARIMA model is expressed as in Equation (1) below:

$$\phi(B)(1-B)^d y_t = \theta(B)\varepsilon_t \tag{1.1}$$

where: y_t represents the value of the time series at time t; e_t is the random error; B is the backshift operator; $\phi(B)$ represents the autoregressive (AR) component of the series; $\theta(B)$ represents the moving average (MA) component of the series; and $(1-B)^d$ denotes the d-fold differencing used to make the series stationary.

Holt's Exponential Smoothing model was proposed in 1957 and is particularly suitable for time series exhibiting a linear trend, which applies to all series considered in this study, including both gasoline and diesel prices. According to Morettin and Toloi (2006), the basic principle of this model consists of two equations: the first smooths the level of the series using a smoothing constant, while the second models the trend present in the series through a second smoothing constant. Accordingly, Equation (2) smooths the level of the series, and Equation (3) smooths the trend of the series.

$$\bar{y}_t = \alpha y_t + (1-\alpha)\left(\bar{y}_{t-1} + \hat{T}_{t-1}\right) \tag{1.2}$$

$$\hat{T}_t = \beta\left(\bar{y}_t - \bar{y}_{t-1}\right) + (1-\beta)\hat{T}_{t-1} \tag{1.3}$$

where, α and β are constants models, which it between $0 < \alpha < 1$ and $0 < \beta < 1$, e $t = 2, \ldots, N$.

Regarding the ARIMA model, in order to determine the values of p and q, the combination of (p,q) that minimized the Bayesian Information Criterion (BIC) was chosen. Furthermore, the (p, q) values free of autocorrelation were identified using the Ljung-Box test.

1.3.1 Singular Spectrum Analysis

Singular Spectrum Analysis (SSA) is a non-parametric method used for the analysis and forecasting of time series. A key feature of SSA is that the methodology allows the decomposition of a series into trend, seasonality, and noise[2]. This enables the reconstruction

[2] The theoretical and practical foundations of the method are found in Golyandina, Nekrutkin and Zhigljavsky (2001), along with a description in Elsner and Tsonis (1996)

of a less noisy series by considering only the trend and seasonal components. In this study, the univariate version of the model is employed to forecast the series under consideration.

In general, the method is described in two main steps: the decomposition of the series and its reconstruction. The decomposition is responsible for separating the series into trend, seasonality, and noise, while the reconstruction focuses on rebuilding the series using only the trend and seasonal components obtained from the decomposition.

1.3.1.1 Short Description about the SSA

SSA consists of two main stages: decomposition and reconstruction, which can be further subdivided into sub-steps. The decomposition stage includes embedding and singular value decomposition, while the reconstruction stage involves grouping and the application of the diagonal averaging. Thus, the first step is to decompose the series, followed by its reconstruction (Golyandina; Nekrutkin; Zhigljavsky, 2001).

Considerer a time series $y_t = (y_0, \ldots, y_{t-1})$ with a length t. The embedding process is responsible for creating the Hankel matrix, in which K lagged vectors X_i are formed, given by $X_i = X_i = (y_{i-1}, \ldots, y_{i+L-2})^t, \quad 1 \leq i \leq K; \quad K = t - L + 1$, where L is an integer satisfying $L \leq \frac{t}{2}$ and L represents the window length. After this step, the time series is transformed into a Hankel matrix, or trajectory matrix, which can be defined algebraically as $X = [X_1 : \cdots : X_K] = (X_{i,j})_{i,j=1}^{L,K}$, where the columns correspond to the lagged vectors present in an $L - \textit{dimensional}$ space.

$$X = \begin{bmatrix} y_0 & y_1 & \cdots & y_{K-1} \\ y_1 & y_2 & \cdots & y_K \\ \vdots & \vdots & \ddots & \vdots \\ y_{L-1} & y_L & \cdots & y_{t-1} \end{bmatrix} \tag{1.4}$$

It is noted that X is the Hankel matrix, and by definition, Hankel matrices have entries $X_{i,j}$ that depend only on the sum $i + j$. Consequently, the dimensions of X are L and K, with $K = N - L + 1$ observations.

After the embedding step, the Singular Value Decomposition (SVD) of the trajectory matrix is performed. Let us define a matrix S as $S = XX^T$, which provides the eigenvalues. The eigenvalues are ordered in descending order: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L$, and their corresponding orthonormal eigenvectors $U_1 \geq U_2 \geq \cdots \geq U_L$; Thus, the SVD of the trajectory matrix can be expressed as the following equation:

$$X = X_1 + \cdots + X_d \tag{1.5}$$

where $X_i = \sqrt{\lambda_i} U_i V_i^T$, with $i = 1, \ldots, d$; $V_i = \frac{X^T U_i}{\sqrt{\lambda_i}}$. The subscript d denotes the number of non-zero eigenvalues of XX^T and V_i, are the principal components. The triplet

$(\sqrt{\lambda_i}, U_i, V_i)$ is referred to as the "eigentriples" resulting from the SVD process (Golyandina; Nekrutkin; Zhigljavsky, 2001).

After these two decomposition steps, the reconstruction of the series begins. The first step in the reconstruction is the grouping of the obtained eigentriples[3], where the selected eigentriples from the previous step form m disjoint subsets of eigentriples, denoted by $I_1, \ldots, I_m$ (with $I = \{i_1, \ldots, i_p\}$). Based on the subsets of eigentriples obtained from the SVD, Equation (5) can be rewritten as follows:

$$X = X_{I_1} + \cdots + X_{I_2} \tag{1.6}$$

Finally, the diagonal averaging step allows each of the groups obtained during decomposition to be transformed into a new time series. To understand this process, let F be an $L \times K$ matrix composed of elements f_{ij} com $1 \leq i \leq L$ e $1 \leq j \leq K$. The diagonal averaging of the matrix F then produces the series $g_0, \ldots, g_{t-1}$, based on Equation (4) (Golyandina; Nekrutkin; Zhigljavsky, 2001).

$$g_k = \begin{cases} \dfrac{1}{k+1} \sum_{m=1}^{k+1} f^*_{m,\,k-m+2}, & 0 \leq k \leq L^* - 1, \\ \dfrac{1}{L^*} \sum_{m=1}^{L^*} f^*_{m,\,k-m+2}, & L^* - 1 \leq k \leq K^*, \\ \dfrac{1}{N-k} \sum_{m=k-K^*+2}^{N-K^*+1} f^*_{m,\,k-m+2}, & K^* \leq k < N. \end{cases} \tag{1.7}$$

Therefore, after obtaining the results g_k, the reconstruction of the series is performed, yielding $\tilde{X}^{(k)} = (\tilde{f}_0^{(k)}, \ldots, \tilde{f}_{N-1}^{(k)})$. According to the authors, the sum of the m decomposed series from the process equals the original series.

1.3.2 Selection of the Window Length (L) and Eigentriples

Within the SSA method, it is necessary to choose two parameters: the optimal window length (L) and the number of eigentriples (r), or elementary matrices, that will be used to reconstruct the series. According to Golyandina, Nekrutkin and Zhigljavsky (2001), during the decomposition step, it is necessary to define the optimal value of L, which should be at most half the length of the series, or smaller $L \leq \frac{T}{2}$. This choice aims to achieve good separability of the series components, due to the greater number of lagged vectors. Some authors, such as Elsner and Tsonis (1996), argue that the optimal value of L should be equal to one quarter of the series length, that is, $L \leq \frac{T}{4}$.

The fact that the window length (L) is set to half or a quarter of the length of the series is not a strict rule, as L can take smaller values or values reflecting the series'

[3] Not all eigentriples are selected during the reconstruction process, as some may exhibit noisy behavior. Therefore, only the eigentriples identified as representing the trend and seasonal components are grouped. A detailed discussion of this step can be found in Hassani (2007).

periodicity. Therefore, for this study, the approach proposed by authors such as Hassani (2007), Hassani, Heravi and Zhigljavsky (2009) were adopted, which considers the seasonal behavior of the series.

In this study, a periodogram was used to identify the seasonality present in the series. Thus, regarding the periodicity ($\frac{K}{12}$, where k is the number of repetitions of the seasonality, ranging from 1 to 12 — for example, if $K = 1$, the seasonality repeats every 12 months), it can assume values smaller than 1 relative to k. As noted in Golyandina and Korobeynikov (2014), a periodogram is used to identify the frequency of repetitions in a series. It is always necessary to determine the largest number that is less than half the series length ($\frac{L}{2}$) and is also divisible by 12.

Afterward, it is necessary to compute the so-called *w-correlation*, which measures the dependency between two time series $Y_T^{(1)}$ and $Y_T^{(2)}$, for example, as follows:

$$\rho_{12}^{(w)} = \frac{(Y_T^{(1)}, Y_T^{(2)})_w}{\|Y_T^{(1)}\|_w \, \|Y_T^{(2)}\|_w} \tag{1.8}$$

where $(Y_T^{(i)}, Y_T^{(j)})_w = \sum_{k=1}^{T} w_k \, y_k^{(i)} y_k^{(j)}$, com $w_k = \min\{k, L, T{-}k\}$, e $\|Y_T^{(i)}\|_w = \sqrt{(Y_T^{(i)}, Y_T^{(i)})_w}$ with $(i, j = 1, 2)$.

Thus, if two reconstructed components have a *w-correlation* value equal to zero, it indicates that they are well separated. Conversely, if there are high values between two reconstructed components, it is possible to infer that they belong to the same group (in this case, they correspond to a specific seasonality present in the original series).

1.3.3 SSA Forecasting

The SSA forecasting method is applied to time series that approximately satisfy the Linear Recurrent Formula (LRF). Briefly, the linear recurrent formula is described as a relationship between past values in the time series that allows the prediction of future values, Golyandina and Zhigljavsky (2013).Time series belonging to the class of linear recurrent formulas are quite broad, including harmonics, polynomials, and exponential time series.

There are two forecasting algorithms within SSA: the recurrent and the vector approaches. The difference between these algorithms is that the recurrent approach calculates a single future value based on a relationship with past values, whereas the vector approach forecasts a vector of future values based on past observations (Golyandina; Nekrutkin; Zhigljavsky, 2001). The vector approach is an extension of the recurrent approach and is recommended for time series exhibiting unit roots, according to Golyandina, Nekrutkin and Zhigljavsky (2001). The presence of unit roots can prevent the linear recurrent formula (LRF) from being minimized during the forecasting process, thereby necessitating the use of a vector component. Therefore, due to the presence of unit roots in the data used in this study, the vector approach was adopted.

By definition, consider the original series $Y_t = (y_1, \ldots, y_t)$ and the reconstructed series $(\tilde{Y})_t = (\tilde{Y}_1, \ldots, \tilde{Y}_t)$. Given an eigenvector $U \in \mathbb{R}^L$, the vector formed by the first $L-1$ components of U is denoted as $U_\nabla \in \mathbb{R}^{L-1}$. Defining $\mathbf{v}^2 = \pi_1^2 + \cdots + \pi_r^2 < 12$, where π_i is the last component of the eigenvector U_i $(i = 1, \ldots, r)$, it is shown that the last component y_L of any vector $Y = (y_1, \ldots, y_L)^T$ is a linear combination of the first $L-1$ component, that is: $y_L = a_1 y_{L-1} + \cdots + a_{L-1} y_1$, where the coefficient vector $A = (a_1, \ldots, a_{L-1})$ is expressed as:

$$A = \sum_{i=1}^{r} \frac{\pi_i \, U_i^\nabla}{1 - v^2} \tag{1.9}$$

Therefore, the forecasts are obtained as follows:

$$\hat{y}_i = \begin{cases} \tilde{y}_i, & i = 1, \ldots, T \\ \displaystyle\sum_{j=1}^{L-1} a_j \, \hat{y}_{i-j}, & i = T+1, \ldots, T+h \end{cases} \tag{1.10}$$

Equation (10) demonstrates the forecasts using the recurrent form of SSA; therefore, it is necessary to have the matrix corresponding to the Vector SSA. With the Vector SSA matrix, it is possible to make forecasts using this algorithm, which is the method of interest. The matrix is as follows:

$$\Pi = V^\nabla (V^\nabla)^T + (1 - v^2) A A^T \tag{1.11}$$

where: $V^\nabla = [U_1^\nabla, \ldots, U_r^\nabla]$. It is also necessary to consider the linear operator: Where,

$$\theta^{(\nu)} : L_r \longmapsto \mathbb{R}^L \tag{1.12}$$

$$\theta^{(\nu)} U = \left(\Pi U^\nabla \,\middle|\, A^T U^\nabla \right) \tag{1.13}$$

Thus, the vector X_i is defined as follows:

$$X_i = \begin{cases} \tilde{X}_i, & \text{where } i = 1, \ldots, K \\ \theta^{(\nu)} X_{i-1}, & \text{where } i = K+1, \ldots, K+h+L-1 \end{cases} \tag{1.14}$$

where $\tilde{X}_i$'s are the reconstructed columns of the trajectory matrix, obtained after grouping and noise removal. Finally, to construct the matrix $X = [X_1, \ldots, X_{K+h+L-1}]$, in addition to applying diagonal averaging, it was possible to obtain the new series

$y_1, \ldots, y_{K+h+L-1}$, where $y_{t+1}, \ldots, y_{t+h}$ represent the h-ahead forecasting using the Vector SSA.

1.3.4 Accuracy Test

In order to test the forecasting accuracy of the proposed models (SSA, ARIMA, and Exponential Smoothing), the last 100 months of each time series (from May 2015 to August 2023) were set apart to serve as the out-of-sample period. The choice of 100 months was based on Baumeister, Kilian and Lee (2017).Thus, using the 100-month out-of-sample period, it was possible to generate forecasts for different horizons, namely h=1, h=3, h=6, and h=12.

Regarding the choice of a 100-month out-of-sample period, this decision is particularly relevant in the Brazilian context, as it encompasses key moments in the country's recent economic history. During these 100 months, Brazil experienced a transition from an inflation rate of 11.07% in 2015 to 2.26% in 2017 (IPEAData, 2024).In addition to the inflationary cycle, the country faced a nationwide truck drivers' strike in 2018, triggered by the sharp increase in diesel prices, as well as the global COVID-19 pandemic in 2020. Moreover, over this 100-month period, Brazil underwent four different presidential administrations, each with distinct macroeconomic policy approaches. Therefore, this time span captures a variety of economic cycles and structural shocks that Brazil faced over the past decade, providing a comprehensive framework for evaluating the forecasting performance of each model employed.

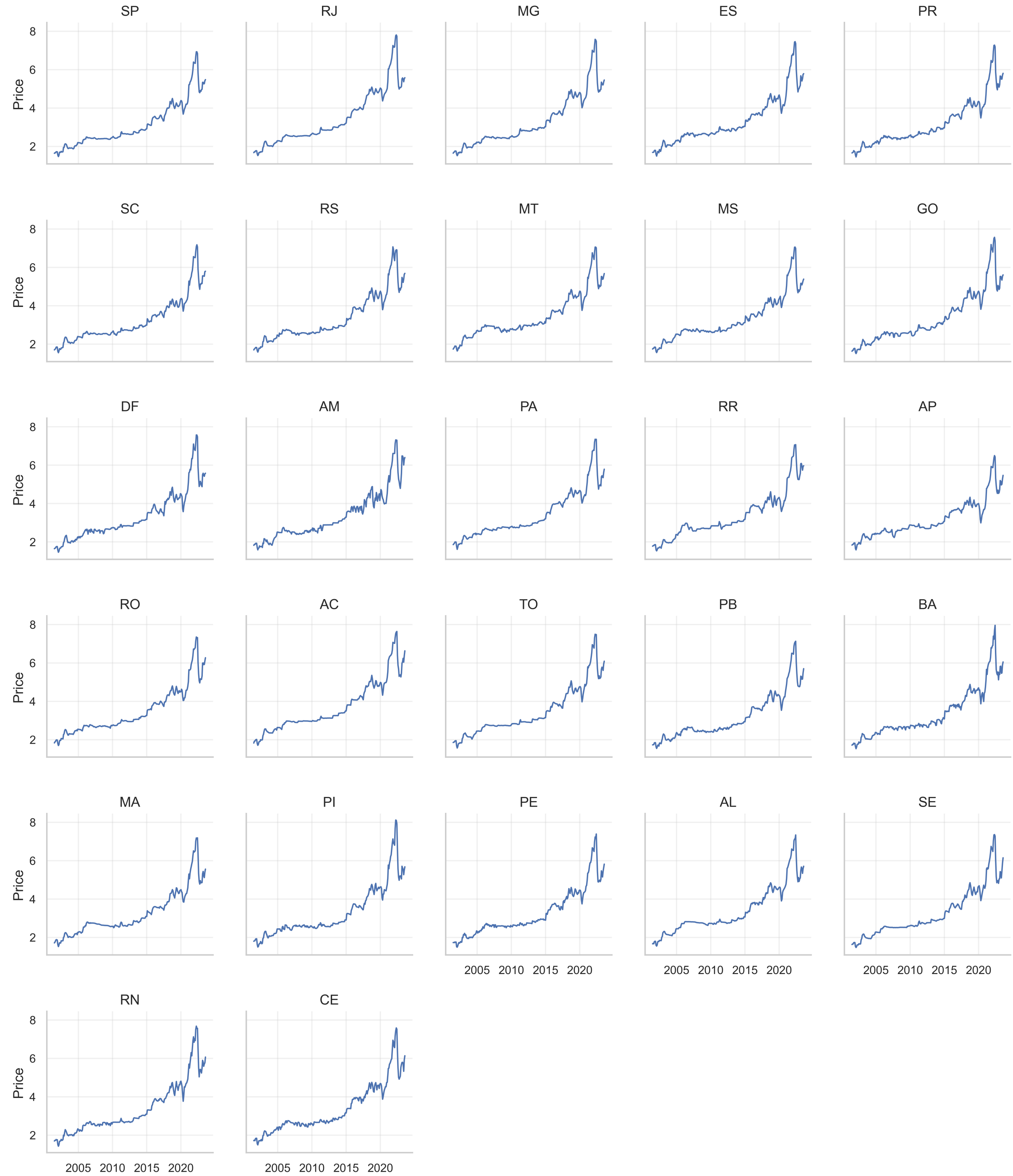
Similar to the approach adopted by Baumeister, Kilian and Lee (2017), a recursive forecasting procedure was employed, whereby each month the models were re-estimated as new information became available. This process can be illustrated as follows: for the estimation in May 2015, the entire in-sample dataset was used; for the forecast of June 2015, the model was re-estimated including the actual observation for May 2015. This procedure was repeated recursively until the end of the out-of-sample period. In addition, a rolling window approach was also applied, allowing forecasts to be generated for each forecast horizon (h = 1, 3, 6, and 12). Consequently, the number of out-of-sample forecast errors was N = 100, 98, 95, and 89 for horizons h = 1, 3, 6, and 12, respectively. The predictive accuracy of each model was evaluated using the Root Mean Squared Error (RMSE). According to Zhang, Patuwo and Hu (1998), it is one of the most commonly used error measures in the forecasting literature.

1.4 Data Base

The data series used in this study were obtained from the National Agency of Petroleum, Natural Gas and Biofuels (ANP)[4]. The series are monthly and cover the period from July 2001 to August 2023. Monthly series were used for all Brazilian states, considering the average retail prices of gasoline and diesel oil. Additionally, there were no data collected in September 2020; therefore, an interpolation between August and October was performed for each series to generate an average value for September 2020. Regarding the state of Pará, the diesel price series had three additional missing months besides September 2020, making it unsuitable for forecasting analysis. For this reason, the state of Pará was excluded from the diesel forecasting analysis. Figures 1.1 and 1.2 present the price series for gasoline and diesel in the Brazilian states, respectively.

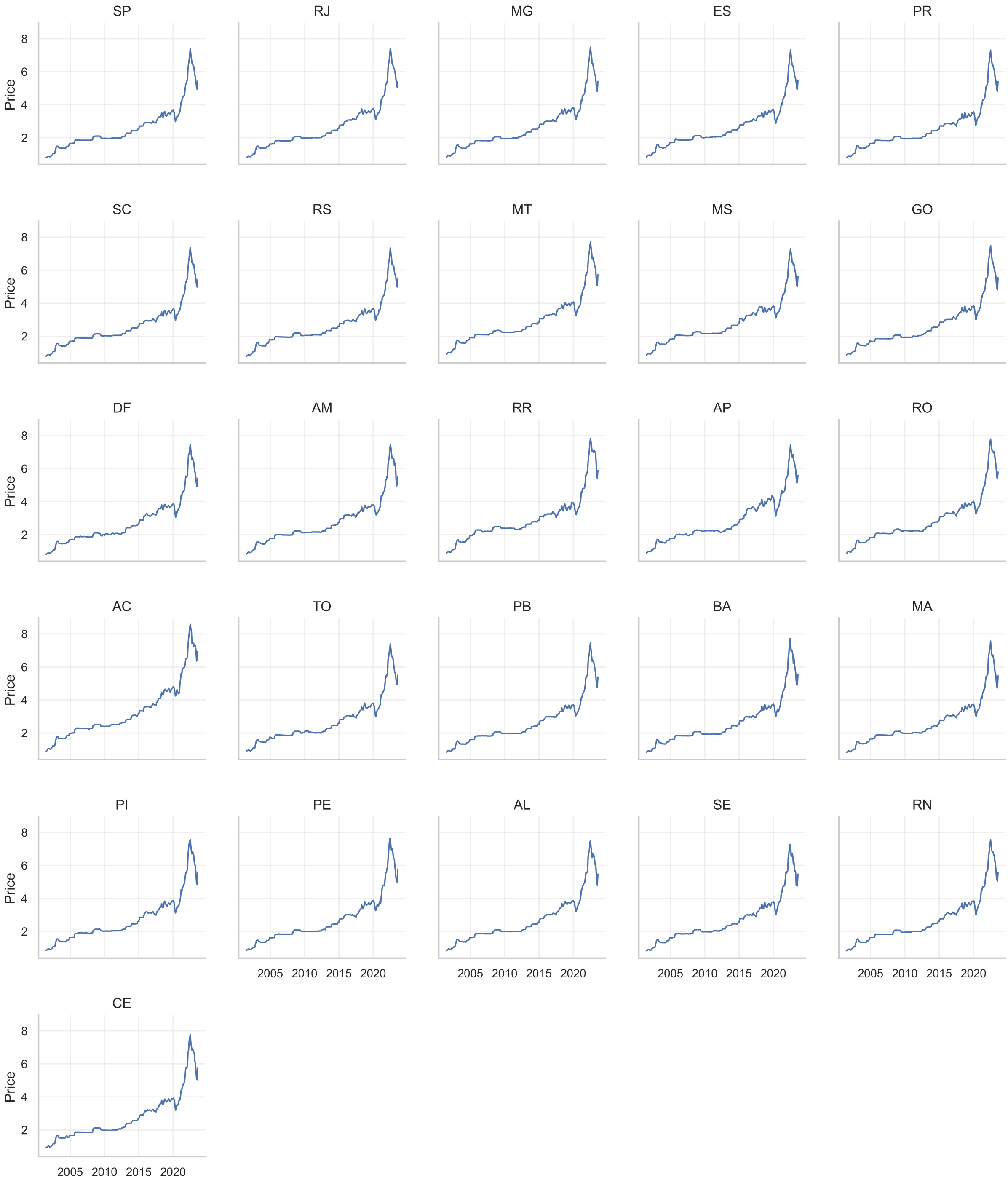
[4] The database is available on ANP website: https://www.gov.br/anp/pt-br/assuntos/precos-e-defesa-da-concorrencia/precos/precos-revenda-e-de-distribuicao-combustiveis/serie-historica-do-levantamento-de-precos.

Figure 1.1 – Gasoline time series prices in Brazil



Source: State abbreviations: Rondônia (RO), Acre (AC), Amazonas (AM), Roraima (RR), Pará (PA), Amapá (AP), Tocantins (TO), Maranhão (MA), Piauí (PI), Ceará (CE), Rio Grande do Norte (RN), Paraíba (PB), Pernambuco (PE), Alagoas (AL), Sergipe (SE), Bahia (BA), Minas Gerais (MG), Espírito Santo (ES), Rio de Janeiro (RJ), São Paulo (SP), Paraná (PR), Santa Catarina (SC), Rio Grande do Sul (RS), Mato Grosso do Sul (MS), Mato Grosso (MT), Goiás (GO), and Federal District (DF).

Figure 1.2 – Diesel time series prices in Brazil



Source: State abbreviations: Rondônia (RO), Acre (AC), Amazonas (AM), Roraima (RR), Pará (PA), Amapá (AP), Tocantins (TO), Maranhão (MA), Piauí (PI), Ceará (CE), Rio Grande do Norte (RN), Paraíba (PB), Pernambuco (PE), Alagoas (AL), Sergipe (SE), Bahia (BA), Minas Gerais (MG), Espírito Santo (ES), Rio de Janeiro (RJ), São Paulo (SP), Paraná (PR), Santa Catarina (SC), Rio Grande do Sul (RS), Mato Grosso do Sul (MS), Mato Grosso (MT), Goiás (GO), and Federal District (DF).

Figures 1.1 and 1.2 illustrate fuel price behavior during the analyzed period, showing that both gasoline and diesel prices trended upward. This upward trend becomes more pronounced in the second half of 2016, following Petrobras's adoption of the Import Parity

Price (IPP) policy, under which domestic fuel prices were aligned with international oil prices and other related market variables. Furthermore, in early 2020, prices declined sharply due to the COVID-19 health crisis, driven by lockdown measures.

Moreover, the increase in diesel prices observed since 2021 was further intensified by the Russian invasion of Ukraine, as Brazil imports a considerable share of its diesel from Russia. Russian products were subject to trade embargoes due to the war, which constrained Brazil's access to diesel and, consequently, led to higher domestic diesel prices (Miller; Kumleben; Nowak, 2024). Subsequently, Tables 1.2 and 1.3 present the descriptive statistics for gasoline and diesel, respectively.

Table 1.2 – Descriptive statistics for gasoline across Brazilian states

States	Observations	Minimum	Median	Mean	Maximum
SP	266	1.472	2.661	3.127	6.934
RJ	266	1.525	2.862	3.438	7.810
MG	266	1.519	2.826	3.299	7.583
ES	266	1.496	2.862	3.321	7.460
PR	266	1.453	2.705	3.199	7.280
SC	266	1.554	2.744	3.228	7.180
RS	266	1.589	2.786	3.347	7.073
MT	266	1.645	2.983	3.435	7.060
MS	266	1.562	2.838	3.296	7.060
GO	266	1.511	2.847	3.338	7.570
DF	266	1.462	2.845	3.325	7.575
AM	266	1.586	2.889	3.333	7.317
PA	266	1.613	2.852	3.419	7.353
RR	266	1.534	2.926	3.336	7.070
AP	266	1.586	2.856	3.181	6.500
RO	266	1.696	2.986	3.465	7.353
AC	266	1.708	3.128	3.698	7.640
TO	266	1.566	2.932	3.459	7.495
PB	266	1.544	2.648	3.182	7.130
BA	266	1.531	2.779	3.367	7.960
MA	266	1.524	2.762	3.248	7.190
PI	266	1.501	2.663	3.298	8.118
PE	266	1.493	2.746	3.265	7.390
AL	266	1.550	2.821	3.384	7.340
SE	266	1.476	2.767	3.269	7.360
RN	266	1.426	2.715	3.349	7.683
CE	266	1.495	2.757	3.352	7.590

Source: Elaborated by the author. SP = São Paulo; RJ = Rio de Janeiro; MG = Minas Gerais; ES = Espírito Santo; PR = Paraná; SC = Santa Catarina; RS = Rio Grande do Sul; MT = Mato Grosso; MS = Mato Grosso do Sul; GO = Goiás; DF = Federal District; AM = Amazonas; PA = Pará; RR = Roraima; AP = Amapá; RO = Rondônia; AC = Acre; TO = Tocantins; PB = Paraíba; BA = Bahia; MA = Maranhão; PI = Piauí; PE = Pernambuco; AL = Alagoas; SE = Sergipe; RN = Rio Grande do Norte; CE = Ceará.

Table 1.2 presents the descriptive statistics for gasoline prices by state in Brazil. It is noteworthy that the state of Piauí recorded the highest monthly gasoline price (8.118), while Rio Grande do Norte registered the lowest monthly price (1.426). Regarding the averages, São Paulo had the lowest mean gasoline price (3.127), whereas Acre exhibited the highest mean (3.698).

Table 1.3 – Descriptive statistics for Diesel across Brazilian states

States	Observations	Minimum	Median	Mean	Maximum
SP	266	0.808	2.103	2.620	7.400
RJ	266	0.797	2.091	2.658	7.420
MG	266	0.840	2.115	2.664	7.490
ES	266	0.841	2.142	2.650	7.330
PR	266	0.803	2.076	2.575	7.310
SC	266	0.788	2.146	2.649	7.370
RS	266	0.779	2.197	2.670	7.340
MT	266	0.905	2.377	2.919	7.710
MS	266	0.868	2.277	2.802	7.300
GO	266	0.871	2.089	2.685	7.500
DF	266	0.813	2.112	2.726	7.460
AM	266	0.822	2.225	2.763	7.460
RR	266	0.895	2.488	2.942	7.830
AP	266	0.869	2.269	2.869	7.450
RO	266	0.864	2.326	2.890	7.790
AC	266	0.874	2.614	3.276	8.580
TO	266	0.919	2.117	2.680	7.390
PB	266	0.851	2.077	2.624	7.450
BA	266	0.841	2.086	2.648	7.720
MA	266	0.824	2.105	2.652	7.570
PI	266	0.861	2.139	2.722	7.560
PE	266	0.856	2.095	2.618	7.640
AL	266	0.841	2.102	2.699	7.490
SE	266	0.838	2.131	2.663	7.270
RN	266	0.838	2.089	2.688	7.560
CE	266	0.923	2.135	2.764	7.760

Source: Elaborated by author. SP = São Paulo; RJ = Rio de Janeiro; MG = Minas Gerais; ES = Espírito Santo; PR = Paraná; SC = Santa Catarina; RS = Rio Grande do Sul; MT = Mato Grosso; MS = Mato Grosso do Sul; GO = Goiás; DF = Distrito Federal; AM = Amazonas; PA = Pará; RR = Roraima; AP = Amapá; RO = Rondônia; AC = Acre; TO = Tocantins; PB = Paraíba; BA = Bahia; MA = Maranhão; PI = Piauí; PE = Pernambuco; AL = Alagoas; SE = Sergipe; RN = Rio Grande do Norte; CE = Ceará.

Considering the descriptive statistics for diesel prices in Brazilian states, shown in Table 1.3, Acre recorded the highest monthly price (8.580), while Santa Catarina had the lowest monthly price (0.788). Regarding the average diesel prices, Acre had the highest monthly average (3.276), whereas Paraná recorded the lowest average (2.575).

1.5 Forecasting Results

1.5.1 Comparison of the accuracy of the forecasts

Considering Table 1.4, the RMSE results for the estimated models using the in-sample data for gasoline and diesel are presented. The first column shows the series used, corresponding to each Brazilian state. Columns 2–4 indicate the RMSE values for gasoline, while columns 5–7 show the RMSE values for diesel. The model parameters are provided in Appendix A (Table A.1).

Table 1.4 – RMSE for the estimated models in the in-sample period for gasoline and diesel

Series	RMSE Gasoline			RMSE Diesel		
	ARIMA	SSA	ES	ARIMA	SSA	ES
SP	0.0477	0.0490	**0.0469**	**0.0345**	0.0361	0.0360
RJ	**0.0447**	0.0574	0.0452	**0.0321**	0.0352	0.0340
MG	**0.0453**	0.0458	0.0487	**0.0339**	0.0369	0.0360
ES	0.0587	**0.0579**	0.0607	0.0360	**0.0313**	0.0380
PR	0.0587	**0.0499**	0.0592	0.0357	**0.0307**	0.0370
SC	0.0550	**0.0473**	0.0574	0.0341	**0.0307**	0.0360
RS	**0.0588**	0.0693	0.0607	0.0392	**0.0332**	0.0410
MT	0.0550	**0.0475**	0.0567	**0.0375**	0.0380	0.0391
MS	**0.0565**	0.0614	0.0580	**0.0382**	0.0384	0.0399
GO	0.0616	**0.0418**	0.0653	**0.0370**	0.0373	0.0380
DF	0.0646	**0.0555**	0.0658	0.0396	**0.0338**	0.0410
AM	0.0681	**0.0565**	0.0703	0.0359	**0.0336**	0.0370
PA	0.0441	**0.0449**	0.0473	–	–	–
RR	0.0518	**0.0506**	0.0557	0.0371	**0.0348**	0.0400
AP	0.0586	**0.0431**	0.0631	**0.0376**	0.0377	0.0400
RO	0.0681	0.0526	**0.0508**	**0.0373**	0.0377	0.0390
AC	**0.0461**	0.0591	0.0502	**0.0402**	0.0411	0.0420
TO	**0.0488**	0.0616	0.0512	**0.0361**	0.0394	0.0390
PB	**0.0539**	0.0541	0.0553	0.0340	**0.0284**	0.0360
BA	0.0723	**0.0564**	0.0739	0.0423	**0.0315**	0.0440
MA	0.0529	**0.0494**	0.0552	0.0347	**0.0289**	0.0370
PI	0.0651	**0.0538**	0.0670	0.0359	**0.0336**	0.0380
PE	0.0605	**0.0563**	0.0607	0.0335	**0.0277**	0.0350
AL	**0.0462**	0.0552	0.0475	**0.0322**	0.0342	0.0340
SE	**0.0452**	0.0545	0.0482	0.0341	**0.0301**	0.0370
RN	0.0602	**0.0592**	0.0618	0.0361	**0.0325**	0.0380
CE	0.0674	**0.0587**	0.0687	0.0371	**0.0336**	0.0390

Source: Elaborated by the author based on research results. Bold values indicate the model that best predicted the series, having the lowest RMSE.

Considering Table 1.4, in 16 states SSA achieved the lowest RMSE values, making it the best in-sample predictor for gasoline; in 16 states, SSA achieved the lowest RMSE value, making it the best predictor; while the ARIMA model performed better in 9 states.

Finally, the ES model provided the best forecasts in 2 states. Therefore, it can be stated that in approximately 60% of the states in Brazil, the SSA model is the most suitable for forecasting gasoline prices; while the ARIMA model is the best predictor in 33% of the Brazilian states, and finally, the ES model provided the best forecasts in 7% of the states.

For diesel, it can be observed that the SSA model provided the best forecasts in 15 Brazilian states (57.69%), while the ARIMA model achieved the best prediction in 11 states (42.30%), and the ES model did not outperform in any state.

Regarding the out-of-sample data, Tables 1.5 and 1.6 present the results for gasoline and diesel, respectively. In Table 1.5, the first column shows the series for the analyzed states; columns 2–4 display the RMSE values of the models for h=1; columns 5–7 show the RMSE values for h=3; columns 8–10 provide the RMSE values for h=6; and finally, columns 11–13 present the RMSE values of each model for h=12. The same logic applies to the columns in Table 1.6.

Table 1.5 – RMSE values for the models estimated based on the out-of-sample data for gasoline

Series	**h=1**			**h=3**			**h=6**			**h=12**		
	SSA	ARIMA	ES	SSA	ARIMA	ES	SSA	ARIMA	ES	SSA	ARIMA	ES
SP	**0.615**	***0.163***	**0.183**	**0.541**	***0.207***	**0.216**	0.652	*0.320*	0.343	0.679	*0.504*	0.518
RJ	**0.670**	***0.204***	**0.226**	0.514	*0.209*	0.226	0.596	*0.330*	0.373	0.683	*0.560*	0.590
MG	**0.682**	***0.200***	**0.225**	**0.543**	***0.223***	**0.244**	**0.690**	***0.351***	**0.402**	0.995	*0.593*	0.671
ES	**0.748**	***0.198***	**0.215**	0.478	*0.232*	0.241	0.588	*0.345*	0.370	0.745	*0.560*	0.568
PR	**0.893**	***0.196***	**0.211**	0.571	*0.233*	0.240	0.702	*0.345*	0.362	0.876	0.533	*0.524*
SC	**0.645**	***0.181***	**0.196**	**0.488**	*0.218*	**0.224**	0.572	*0.322*	0.340	0.727	*0.480*	0.504
RS	**0.552**	***0.195***	**0.213**	**0.457**	***0.254***	0.267	0.531	0.374	0.411	0.646	0.578	0.628
MT	**0.540**	***0.184***	**0.200**	**0.461**	***0.223***	**0.232**	0.574	*0.333*	0.367	0.786	*0.533*	0.585
MS	**0.559**	***0.185***	**0.199**	**0.417**	0.213	0.220	0.520	*0.316*	0.335	0.748	*0.487*	0.513
GO	**0.628**	***0.235***	**0.251**	0.498	*0.267*	0.276	0.644	*0.396*	0.425	0.800	*0.613*	0.655
DF	**1.097**	***0.241***	**0.255**	**0.747**	***0.283***	**0.289**	0.886	*0.411*	0.427	1.172	*0.635*	0.647
AM	**0.538**	***0.262***	**0.270**	**0.501**	***0.333***	**0.340**	0.644	*0.451*	0.472	0.820	*0.596*	0.640
PA	**0.522**	***0.183***	**0.199**	0.388	*0.214*	0.220	0.475	*0.329*	0.356	0.605	*0.532*	0.575
RR	**0.360**	***0.162***	**0.184**	0.340	*0.228*	0.240	0.455	*0.339*	0.362	***0.484***	**0.516**	**0.543**
AP	**0.445**	***0.179***	**0.195**	**0.423**	***0.221***	**0.223**	0.558	*0.320*	0.335	0.630	*0.500*	0.520
RO	**0.679**	***0.194***	**0.209**	0.493	*0.230*	0.231	0.556	*0.354*	0.369	0.653	*0.528*	0.561
AC	**0.554**	***0.194***	**0.208**	0.429	*0.222*	0.253	**0.501**	*0.331*	**0.411**	0.595	*0.501*	0.618
TO	**0.578**	***0.188***	**0.207**	0.455	*0.233*	0.240	0.536	*0.352*	0.370	0.659	*0.549*	0.569
PB	**0.435**	***0.172***	**0.190**	0.394	*0.225*	0.234	0.517	*0.357*	0.366	0.593	0.563	*0.551*
BA	**0.595**	***0.258***	**0.259**	0.450	*0.251*	0.251	0.545	*0.352*	0.366	0.645	*0.560*	0.583
MA	**0.540**	***0.181***	**0.201**	0.411	*0.219*	0.227	0.531	*0.329*	0.357	0.781	*0.525*	0.573
PI	**0.682**	***0.246***	**0.262**	0.459	*0.279*	0.285	0.572	*0.401*	0.419	0.712	*0.637*	0.647
PE	**0.577**	***0.210***	**0.230**	0.435	*0.240*	0.251	0.503	*0.347*	0.369	0.603	0.544	*0.542*
AL	**0.485**	***0.196***	**0.205**	0.392	0.227	*0.222*	0.462	*0.335*	0.345	0.563	*0.521*	0.544
SE	**0.589**	***0.200***	**0.231**	0.455	*0.262*	0.271	0.531	*0.373*	0.391	0.633	*0.553*	0.586

Continues on the next page

Table 1.5 – Continues

Series	**h=1**			**h=3**			**h=6**			**h=12**		
	SSA	ARIMA	ES	SSA	ARIMA	ES	SSA	ARIMA	ES	SSA	ARIMA	ES
RN	**0.603**	***0.226***	**0.239**	**0.464**	***0.261***	**0.267**	0.582	*0.373*	0.392	**0.734**	**0.575**	0.603
CE	**0.653**	***0.229***	**0.241**	0.484	*0.268*	0.276	0.601	*0.375*	0.390	0.786	*0.562*	0.578

Source: Prepared based on the research results. Numbers in bold indicate predictions that were statistically different according to the Modified Diebold-Mariano test (10% significance), considering SSA/ARIMA and SSA/Exponential Smoothing. Numbers in italics represent the forecast with the lowest RMSE value.

Table 1.6 – RMSE values for the estimated models based on the out-of-sample period for diesel

Series	**h=1**			**h=3**			**h=6**			**h=12**		
	SSA	ARIMA	ES	SSA	ARIMA	ES	SSA	ARIMA	ES	SSA	ARIMA	ES
SP	**0.559**	***0.146***	**0.163**	0.374	*0.186*	0.203	0.483	*0.292*	0.390	0.605	0.453	*0.451*
RJ	**0.503**	***0.138***	**0.226**	0.372	*0.177*	0.226	0.493	*0.281*	0.373	0.614	*0.442*	0.590
MG	**0.849**	***0.158***	**0.176**	0.448	*0.194*	0.214	0.515	*0.300*	0.333	0.573	*0.465*	0.493
ES	**0.686**	***0.143***	**0.161**	0.380	*0.182*	0.198	0.463	*0.290*	0.320	0.556	*0.450*	0.499
PR	**0.599**	***0.160***	**0.176**	0.377	*0.197*	0.218	0.478	*0.310*	0.341	0.603	*0.478*	0.504
SC	**0.548**	***0.145***	**0.162**	**0.377**	***0.187***	**0.205**	0.503	*0.294*	0.323	0.702	*0.448*	0.474
RS	**0.969**	***0.156***	**0.169**	0.489	*0.190*	0.205	0.564	*0.292*	0.319	0.584	*0.452*	0.475
MT	**0.520**	***0.158***	**0.177**	0.361	*0.190*	0.210	0.480	*0.299*	0.331	0.626	*0.467*	0.497
MS	**0.551**	***0.151***	**0.167**	**0.450**	***0.193***	*0.213*	0.569	*0.302*	0.334	0.673	*0.458*	0.491
GO	**0.600**	***0.171***	**0.186**	0.391	*0.203*	0.219	0.494	*0.309*	0.338	0.583	*0.471*	0.504
DF	**0.413**	***0.164***	**0.180**	0.318	*0.203*	0.221	0.492	*0.318*	0.347	0.727	*0.483*	0.507
AM	**0.471**	***0.168***	**0.180**	0.366	*0.199*	0.209	0.476	*0.292*	0.320	0.604	*0.456*	0.496
RR	**0.750**	***0.164***	**0.181**	**0.524**	***0.217***	0.231	0.645	*0.315*	0.351	0.673	*0.475*	0.563
AP	**0.632**	***0.161***	**0.173**	**0.517**	***0.218***	**0.234**	0.604	*0.336*	0.378	0.664	*0.497*	0.614
RO	**0.474**	***0.139***	**0.162**	**0.375**	***0.191***	0.214	0.512	*0.304*	0.336	0.674	*0.470*	0.505
AC	**0.474**	***0.171***	**0.186**	0.355	*0.208*	0.217	0.471	*0.297*	0.322	0.593	*0.454*	0.486
TO	**0.462**	***0.155***	**0.167**	0.372	*0.192*	0.198	0.496	*0.302*	0.332	0.660	*0.479*	0.560
PB	**1.308**	***0.152***	**0.174**	0.702	*0.188*	0.210	0.841	*0.297*	0.330	0.926	*0.464*	0.496
BA	**0.609**	***0.190***	**0.201**	0.393	*0.218*	0.232	0.502	*0.329*	0.359	**0.629**	*0.499*	0.537
MA	**0.483**	***0.181***	**0.201**	0.372	*0.200*	0.219	0.531	*0.315*	0.348	0.702	*0.484*	0.537
PI	**1.007**	***0.176***	**0.194**	0.616	*0.208*	0.230	0.752	*0.330*	0.365	0.824	*0.503*	0.543
PE	**0.639**	***0.166***	**0.187**	0.383	*0.207*	0.222	0.486	*0.320*	0.346	0.669	*0.480*	0.516
AL	**0.436**	***0.186***	**0.194**	0.334	*0.195*	0.204	0.452	*0.286*	0.312	0.590	*0.439*	0.472
SE	**0.446**	***0.207***	**0.210**	0.349	*0.226*	0.233	0.492	*0.324*	0.345	0.668	*0.469*	0.494
RN	**0.431**	***0.150***	**0.173**	0.327	*0.206*	0.228	0.445	*0.318*	0.350	0.623	*0.476*	0.506

Continue on the next page

Table 1.6 – Continues

Series	**h=1**			**h=3**			**h=6**			**h=12**		
	SSA	ARIMA	ES	SSA	ARIMA	ES	SSA	ARIMA	ES	SSA	ARIMA	ES
CE	**0.489**	***0.181***	**0.193**	0.355	*0.209*	0.225	0.467	*0.320*	0.349	0.639	*0.483*	0.522

Source: Prepared based on the research results. Bold numbers indicate forecasts that are statistically different according to the Modified Diebold-Mariano test (10% of significance) , considering SSA/ARIMA and SSA/Exponential Smoothing comparisons. Italic numbers indicate the forecast with the lowest RMSE value.

For the out-of-sample results shown in Table 1.5, considering the predictive models, it was observed that the ARIMA model generally provided the best forecasts for all 27 Brazilian states across the different forecast horizons. Specifically, for h=1 and h=6, ARIMA achieved the lowest RMSE values among all models. For h=3, ARIMA had the best prediction in 25 states, while in Alagoas the Exponential Smoothing (ES) model performed best, achieving the lowest RMSE; in Bahia, ARIMA and ES tied with the same RMSE value. Finally, for h=12, ARIMA had the best predictive performance in 23 states, SSA performed best in Roraima, and ES achieved the lowest RMSE in three states (Paraná, Paraíba, and Pernambuco).

The out-of-sample results for diesel shown in Table 1.6 indicate that the ARIMA model provided the best forecasts for all 26 states at h=1, h=3, and h=6. Finally, for h=12, ARIMA achieved the best predictive performance in 25 states, while Exponential Smoothing (ES) had the lowest RMSE for the state of São Paulo.

Therefore, considering all 108 forecasts estimated for gasoline, it can be seen that in 104 forecasts, the ARIMA model outperformed the other models. For diesel, ARIMA was superior in 103 out of 104 forecasts. Consequently, the SSA model did not prove to be an effective predictive model for gasoline and diesel prices across Brazilian states.

Although ARIMA showed the lowest RMSE values, when the Modified Diebold-Mariano test, proposed by Harvey, Leybourne and Newbold (1997) (Diebold-Mariano test original was proposed in 1995 Diebold and Mariano (1995)),was performed, it was observed that the majority of the estimates were statistically indistinguishable. For this test, comparisons were made between the forecasts of SSA/ARIMA and SSA/Exponential Smoothing (ES) for both gasoline and diesel.

Thus, for gasoline, the number of states where the forecasts were statistically different were as follows. In the SSA/ARIMA comparison: for h=1, all 27 states; for h=3, 9 states; for h=6, 1 state; and for h=12, 2 states. In the SSA/ES comparison: for h=1, 27 states; for h=3, 9 states; for h=6, 2 states; and for h=12, 1 state. For diesel, the results were: SSA/ARIMA: for h=1, 26 states; for h=3, 5 states; and for h=6 and h=12, no forecasts were statistically different. In SSA/ES: for h=1, 26 states; for h=3, 3 states; and for h=6 and h=12, there were no statistically different forecasts.

1.5.2 Encompassing Test

In addition to the Modified Diebold-Mariano test, the encompassing test proposed by Harvey, Leybourne and Newbold (1998) was also implemented. The encompassing test works with the forecast errors of the models proposed in this study. The specific form of the encompassing test was chosen due to its simplicity and ease of understanding, although it can be represented in other ways. The equations below are used to test forecast encompassing:

$$e_{t,h}^{SSA} = \lambda \left(e_{t,h}^{SSA} - e_{t,h}^{ARIMA} \right) + \eta_{t,h} \quad (1.15)$$

$$e_{t,h}^{SSA} = \lambda \left(e_{t,h}^{SSA} - e_{t,h}^{AE} \right) + \eta_{t,h} \quad (1.16)$$

The objective of the forecast encompassing test, implemented through the forecast errors $e_{t,h}^{SSA}$ and $e_{t,h}^{ARIMA}$ in Equation (1.15), is to determine whether the forecasts produced by the SSA model contain useful predictive information that is not already incorporated into the forecasts produced by the ARIMA model.

The null hypothesis is $H_0 : \lambda = 0$, whereas the alternative hypothesis is $H_1 : \lambda > 0$. Failure to reject the null hypothesis indicates that the forecasts produced by the SSA model encompass those produced by the ARIMA model, implying that the ARIMA forecasts do not provide additional predictive information beyond that already contained in the SSA forecasts.

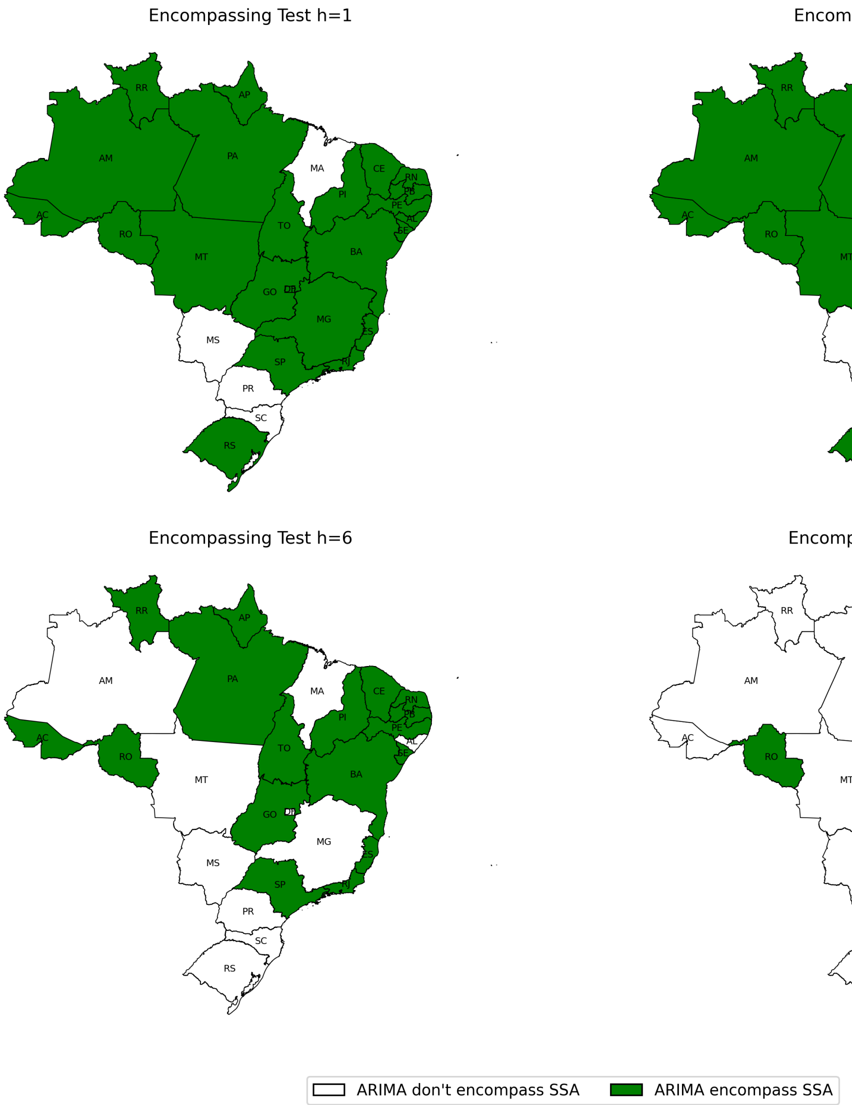
Additionally, the reverse hypothesis must also be tested to assess whether the forecasts produced by the ARIMA model encompass those produced by the SSA model. This reverse test is implemented through Equation (1.16), using the forecast errors $e_{t,h}^{ARIMA}$ and $e_{t,h}^{SSA}$ in Equation (1.16). The null hypothesis is $H_0 : \lambda = 0$, whereas the alternative hypothesis is $H_1 : \lambda > 0$. The same procedure was subsequently applied to the SSA and ES models.

The Encompassing Test results are shown in Figures 1.3 to 1.7 for gasoline and diesel, respectively.

For gasoline, as shown in Figure 1.3, the encompassing test results for SSA and ARIMA were as follows: The ARIMA model encompassed SSA in 23 states for h=1 and h=3; for h=6, in 17 states; and for h=12, in 11 states. In the opposite direction, SSA encompassed ARIMA only for h=12 and in 6 states, as shown in Figure 1.5. Regarding SSA and Exponential Smoothing (ES), present in Figure 1.4, ES encompassed SSA in several states across all forecast horizons: for h=1, in 17 states; for h=3, in 18 states; for h=6, in 12 states; and for h=12, in 3 states. SSA encompassed ES only for h=12 and was present in 5 states, as shown in Figure 1.5.

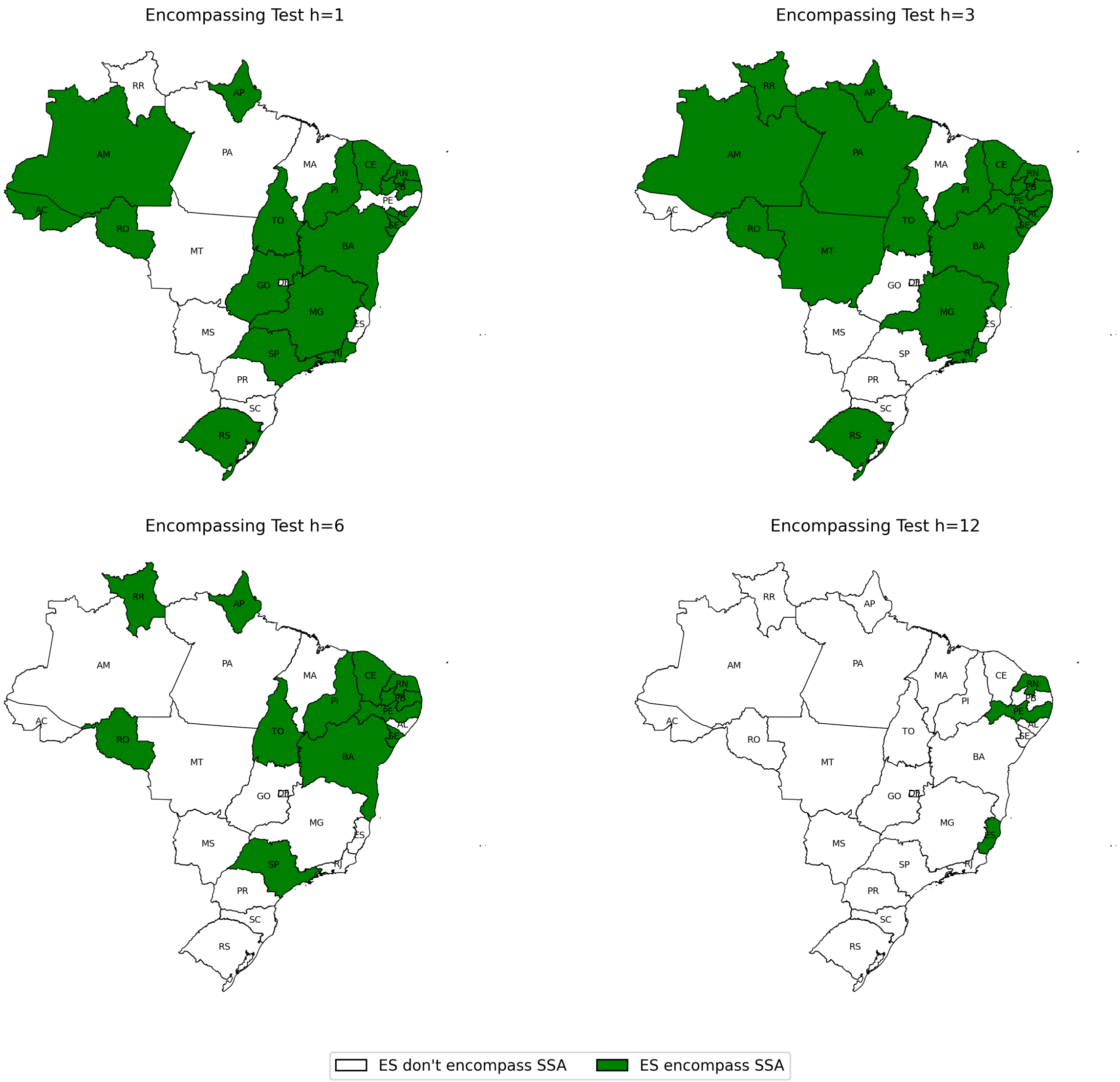
For diesel, as shown in Figure 1.6 (ARIMA/SSA) and 1.7 (ES/SSA), the encompassing test found that the ARIMA and ES models encompassed the forecasts of SSA. Specifically, for SSA and ARIMA, the results were as follows: ARIMA encompassed SSA in 15 states for h=1, 21 for h=3, 9 for h=6, and 4 for h=12. For SSA and ES, ES encompassed SSA in 15 states for h=1, 17 for h=3, 7 for h=6, and 4 for h=12.

Figure 1.3 – ARIMA encompassing test for gasoline



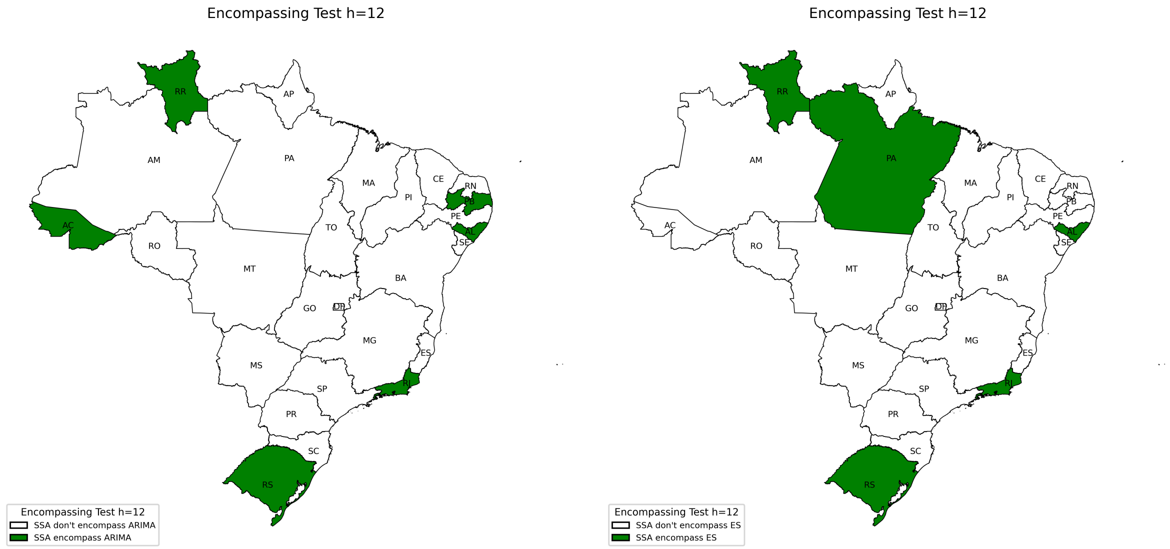
Source: The encompassing test uses a 10% significance level. State abbreviations: Rondônia (RO), Acre (AC), Amazonas (AM), Roraima (RR), Pará (PA), Amapá (AP), Tocantins (TO), Maranhão (MA), Piauí (PI), Ceará (CE), Rio Grande do Norte (RN), Paraíba (PB), Pernambuco (PE), Alagoas (AL), Sergipe (SE), Bahia (BA), Minas Gerais (MG), Espírito Santo (ES), Rio de Janeiro (RJ), São Paulo (SP), Paraná (PR), Santa Catarina (SC), Rio Grande do Sul (RS), Mato Grosso do Sul (MS), Mato Grosso (MT), Goiás (GO), and Federal District (DF).

Figure 1.4 – ES encompassing test for gasoline



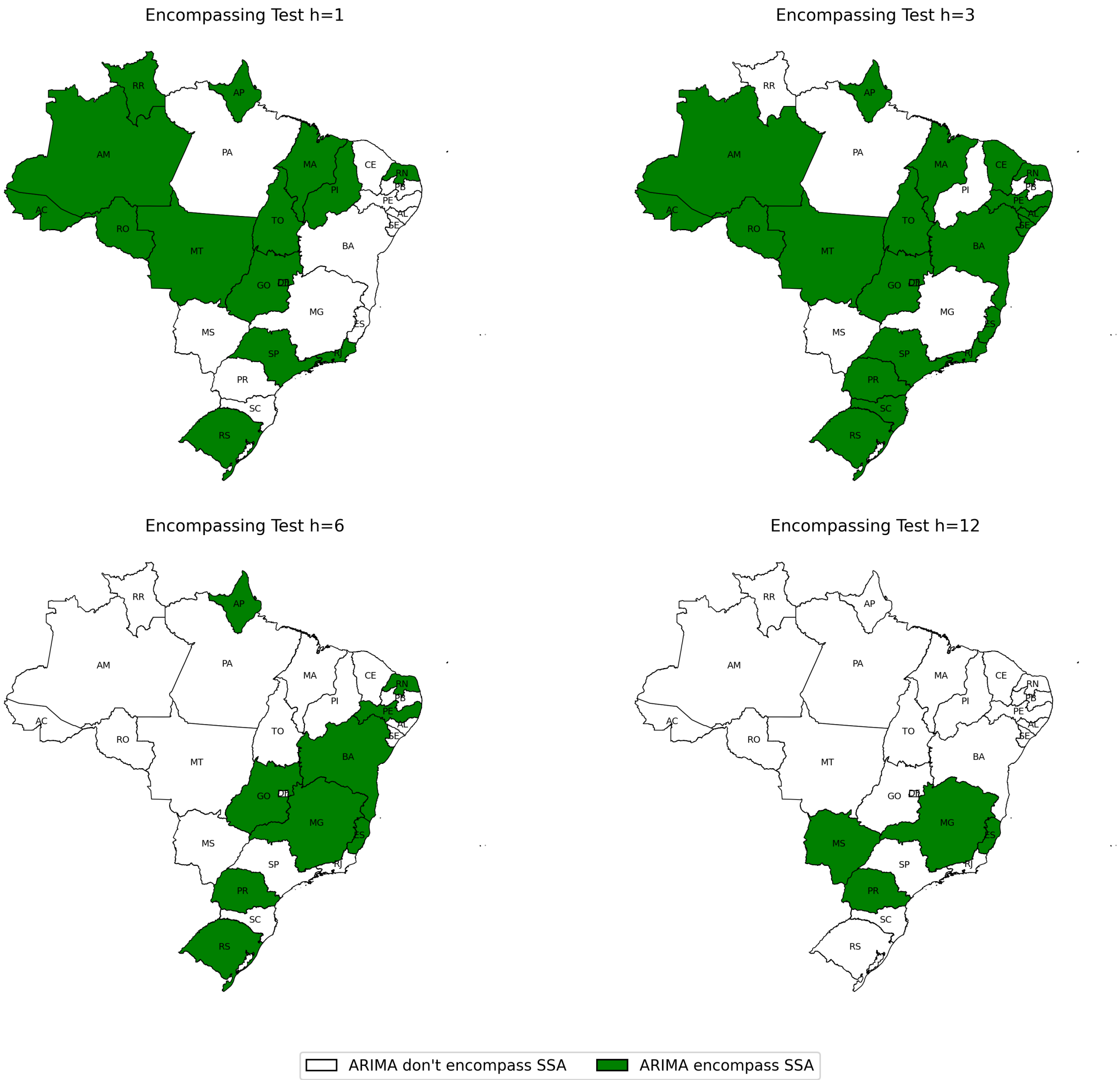
Source: The encompassing test uses a 10% significance level. State abbreviations: Rondônia (RO), Acre (AC), Amazonas (AM), Roraima (RR), Pará (PA), Amapá (AP), Tocantins (TO), Maranhão (MA), Piauí (PI), Ceará (CE), Rio Grande do Norte (RN), Paraíba (PB), Pernambuco (PE), Alagoas (AL), Sergipe (SE), Bahia (BA), Minas Gerais (MG), Espírito Santo (ES), Rio de Janeiro (RJ), São Paulo (SP), Paraná (PR), Santa Catarina (SC), Rio Grande do Sul (RS), Mato Grosso do Sul (MS), Mato Grosso (MT), Goiás (GO), and Federal District (DF).

Figure 1.5 – SSA encompassing test for gasoline



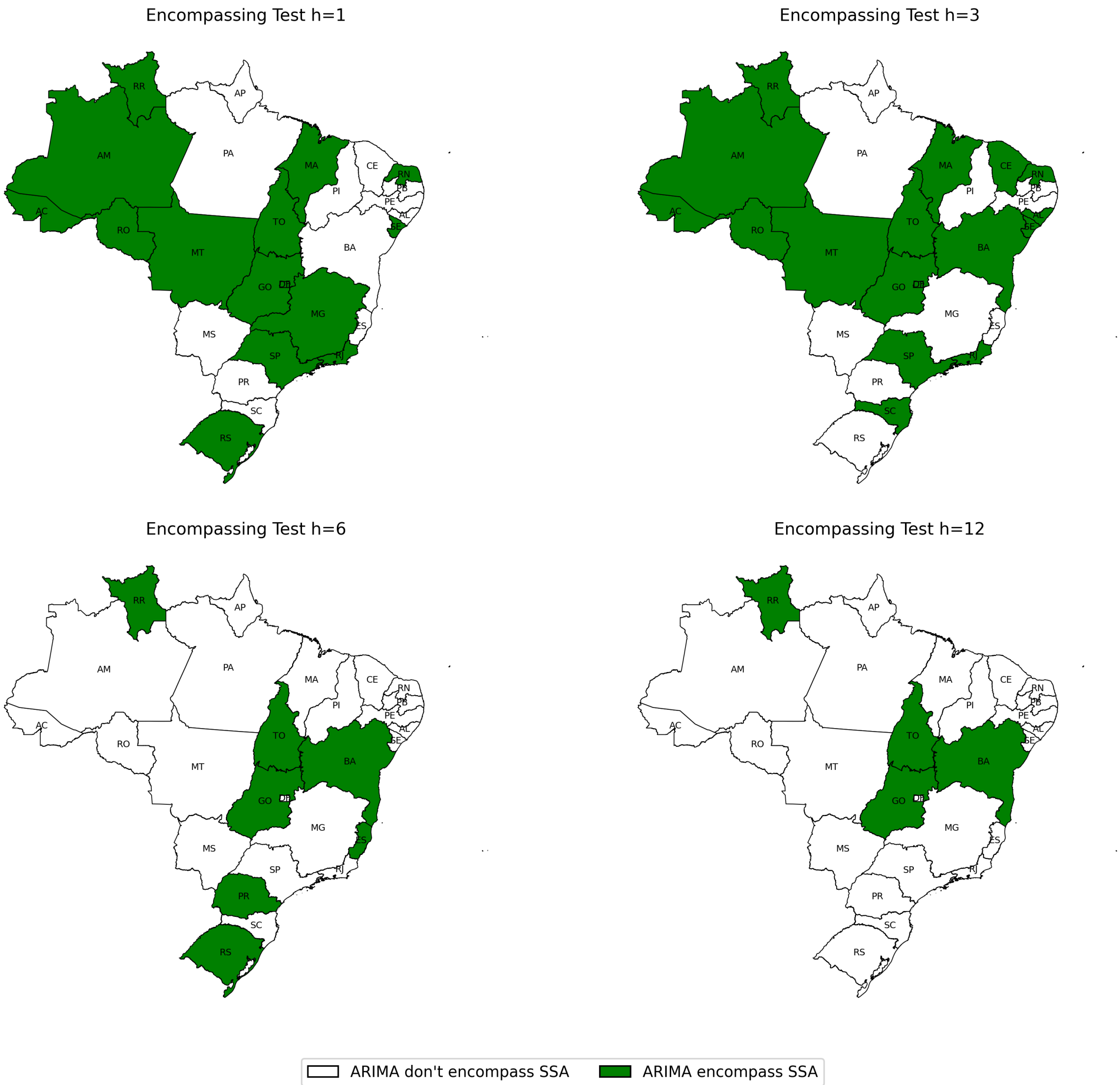
Source: The encompassing test uses a 10% significance level. State abbreviations: Rondônia (RO), Acre (AC), Amazonas (AM), Roraima (RR), Pará (PA), Amapá (AP), Tocantins (TO), Maranhão (MA), Piauí (PI), Ceará (CE), Rio Grande do Norte (RN), Paraíba (PB), Pernambuco (PE), Alagoas (AL), Sergipe (SE), Bahia (BA), Minas Gerais (MG), Espírito Santo (ES), Rio de Janeiro (RJ), São Paulo (SP), Paraná (PR), Santa Catarina (SC), Rio Grande do Sul (RS), Mato Grosso do Sul (MS), Mato Grosso (MT), Goiás (GO), and Federal District (DF).

Figure 1.6 – ARIMA encompassing test for diesel



Source: The encompassing test uses a 10% significance level. State abbreviations: Rondônia (RO), Acre (AC), Amazonas (AM), Roraima (RR), Pará (PA), Amapá (AP), Tocantins (TO), Maranhão (MA), Piauí (PI), Ceará (CE), Rio Grande do Norte (RN), Paraíba (PB), Pernambuco (PE), Alagoas (AL), Sergipe (SE), Bahia (BA), Minas Gerais (MG), Espírito Santo (ES), Rio de Janeiro (RJ), São Paulo (SP), Paraná (PR), Santa Catarina (SC), Rio Grande do Sul (RS), Mato Grosso do Sul (MS), Mato Grosso (MT), Goiás (GO), and Federal District (DF).

Figure 1.7 – ES encompassing test for diesel



Source: The encompassing test uses a 10% significance level. State abbreviations: Rondônia (RO), Acre (AC), Amazonas (AM), Roraima (RR), Pará (PA), Amapá (AP), Tocantins (TO), Maranhão (MA), Piauí (PI), Ceará (CE), Rio Grande do Norte (RN), Paraíba (PB), Pernambuco (PE), Alagoas (AL), Sergipe (SE), Bahia (BA), Minas Gerais (MG), Espírito Santo (ES), Rio de Janeiro (RJ), São Paulo (SP), Paraná (PR), Santa Catarina (SC), Rio Grande do Sul (RS), Mato Grosso do Sul (MS), Mato Grosso (MT), Goiás (GO), and Federal District (DF).

Therefore, in general terms, the encompassing test corroborated the results presented in Tables 1.5 and 1.6, indicating that the SSA model was generally not the best model for forecasting monthly state-level gasoline and diesel prices in Brazil. Nevertheless, the test also indicated that, at the 12-month forecast horizon, the SSA model may retain some predictive relevance, particularly for gasoline prices, as its forecasts provided additional predictive information beyond that contained in the ARIMA and ES forecasts for some states. A possible explanation for this finding is that SSA is better able to capture long-term structures in the price series, such as slowly evolving trends and low-frequency dynamics, which become increasingly important at longer forecast horizons. As forecast uncertainty accumulates, these components may provide additional predictive information that is not fully captured by the ARIMA and ES models.

1.6 Policy Implications

Given the results presented in Tables 1.5 and 1.6, including the outcomes of the Modified Diebold-Mariano test and the Encompassing test for the three predictive models employed, several conclusions can be drawn. First, no single model proved superior across all forecasting horizons. Second, for h=1, the ARIMA model is the most suitable, as it achieved the lowest RMSE and its forecasts were statistically different from those of SSA; for the other forecasting horizons, only a few forecasts were statistically different. Third and finally, the SSA model may contain useful information for forecasting, particularly at the h=12 horizon.

Thus, it can be stated that the predictive results differ from the forecasting literature regarding SSA, as noted in Hassani, Heravi and Zhigljavsky (2009), Hassani et al. (2015), where SSA is highlighted as a forecasting method far superior to other methodologies, including ARIMA and ES. Moreover, the results obtained here also diverge from the fuel forecasting literature[5], which generally indicates ARIMA as the best choice regardless of the forecasting horizon (see Silva (2019), Anjos et al. (2023)). Therefore, it can be concluded that ARIMA is only the most suitable method for h=1, while for the other horizons, any of the three methodologies can be considered for forecasting.

Furthermore, this study demonstrated that it is possible to create an overall forecasting outlook and present it to consumers with the aim of informing them about the trends in fuel prices in Brazil. Such forecasts provide consumers with a higher level of confidence regarding future prices, reducing volatility and increasing certainty for potential future transactions. This, in turn, allows consumers to gain a general understanding of price behavior at the state level, thereby enhancing connectivity via road networks for consumer decision-making.

Nevertheless, it is important to note that even when forecasting fuel prices, unexpected crises may occur and undermine the forecast. One example was the COVID-19 crisis, which stopped global fuel production and raised fuel prices. This type of crisis typically generates a structural break, and it is hard for predictive models to capture such events.

Finally, the existence of fuel forecasting by organizations in the transport sector, such as the CNT, and in the energy sector, such as the ANP, could promote similar studies

[5] Although ARIMA has been superior in the predictions overall, the Encompassing Test and Modified Diebold-Mariano Test noted that most of the forecasting in h=3, 6, and 12, indicated that the predictions were not different from a statistical point of view (Modified Diebold-Mariano Test) and that the SSA model did not encompass the forecasting in terms of other models (Encompassing Test).

to disseminate these results, thereby reaching a larger number of consumers across the different Brazilian states.

1.7 Conclusion

The forecasting of gasoline and diesel prices in a country like Brazil, which relies heavily on road transportation, is a highly relevant topic. Sharp fluctuations in these fuel prices over the past decade have made consumers constantly attentive to changes. Therefore, predicting future prices of these fuels is a challenge, given their volatile behavior and the significant regional price disparities in Brazil due to its continental size.

To address this challenge, two traditional forecasting methods (ARIMA and Exponential Smoothing) were employed for fuel prices, alongside a non-parametric method, the SSA (Singular Spectrum Analysis). According to the reviewed literature, SSA was expected to outperform the traditional models. However, this expectation was not met, despite previous studies suggesting that SSA could provide superior predictive performance.

Nevertheless, it was also observed that although the ARIMA model exhibited the lowest RMSE values for the monthly gasoline and diesel price forecasts, this result was valid only for h=1. For other forecasting horizons, only a few states showed statistically significant differences between the predictions of the models. Therefore, no single predictive method proved to be superior across the overall forecasting landscape for fuel prices.

Moreover, regarding SSA, it was observed that for longer time horizons (h=12), this model exhibited moderately relevant predictive performance compared to the other models analyzed. When examining gasoline, SSA could provide more useful information for monthly price forecasting than the ARIMA and Exponential Smoothing models, according to the Encompassing test, for some Brazilian states. This indicates that SSA may be a useful predictive tool for longer-term horizons, at least for gasoline.

This evidence about the SSA yields better results in the long term because its methodology captures trends and seasonality. Therefore, the method can better handle long-term forecasting, unlike ARIMA and ES, which treat the series as a single unit. In addition, the database generation process contains non-linear features (common in price series), and given that, SSA is a method that separates trend, season, and noise; applying it to time series would be very hard for this method to handle and predict the series.

Finally, this study contributed to the national literature on fuel price forecasting by providing predictions for each Brazilian state, offering a state-level overview of the behavior of fuel prices in Brazil. Such an overview is important given the vast distances across the country, as it can provide consumers with a level of predictability regardless of their location. Moreover, it presents a comprehensive picture of the dynamics of fuel prices in each region of the country.

The main limitation of this study is the absence of the state of Pará in the diesel price forecasts due to the lack of three months in the monthly price time series. This makes it impossible to have a mirrored overview for both fuels across all Brazilian states. As a suggestion for future research, hybrid forecasting models could be employed to combine different predictive methods, aiming to achieve better forecasting performance for gasoline and diesel prices.

2 Essay 2 - Volatility spillovers between oil and gasoline/diesel in Brazil: A Bayesian approach

Abstract

The main objective of this study is to examine the volatility transmission from international crude oil prices to domestic gasoline and diesel prices in Brazil. Since both fuels are petroleum derivatives, understanding the volatility spillover dynamics is particularly relevant during periods of heightened oil price uncertainty. To achieve this, we employ a multivariate GARCH model using the BEKK specification, which is capable of capturing volatility interactions between time series. The analysis covers the period from July 2001 to August 2023. The results indicate that gasoline prices are more responsive to oil price shocks than diesel prices. Moreover, the overall spillover effect between international oil prices and domestic fuel prices is moderate, largely due to government interventions. However, during the implementation of the Import Parity Pricing (IPP) policy, when fuel prices were more market-driven, we observe a persistent and more pronounced volatility spillover from crude oil to both gasoline and diesel prices.

Resumo

O principal objetivo deste estudo é analisar a transmissão de volatilidade dos preços internacionais do petróleo bruto para os preços domésticos da gasolina e do diesel no Brasil. Para isso, emprega-se um modelo GARCH multivariado na especificação BEKK, capaz de capturar as interações de volatilidade entre as séries temporais. A análise abrange o período de julho de 2001 a agosto de 2023.

Os resultados mostram que os preços da gasolina são mais sensíveis aos choques de volatilidade do mercado internacional de petróleo do que os preços do diesel. Além disso, o transbordamento de volatilidade entre os preços internacionais do petróleo e os preços domésticos dos combustíveis é, em geral, moderado, em razão das intervenções governamentais. Entretanto, durante a vigência da política de Preço de Paridade de Importação (PPI), observa-se um transbordamento de volatilidade mais persistente e intenso do petróleo bruto para os preços da gasolina e do diesel.

2.1 Introduction

Oil is the most traded commodity in the world, and plays a central role in the global economy, serving purposes that range from fuels and plastics to a broad set of final goods. The relationship between oil and the international literature has been a recurring topic, particularly regarding the impacts of oil price fluctuations on macroeconomic performance, as evidenced by the works of Hamilton (1983), Akçelik and Ögünç (2016), Barbosa et al. (2023). Furthermore, Barbosa et al. (2023) argues that changes in international oil prices have asymmetric effects on oil-exporting and oil-importing countries. These effects, distinct between import and export oil countries, have already been discussed in prior studies, such as Aastveit, Bjørnland and Thorsrud (2015).

Accordingly, the main objective of this study is to examine how volatility in international oil prices is transmitted to domestic fuel prices (gasoline and diesel) in Brazil. In this article, the West Texas Intermediate (WTI) crude oil price is used as a proxy for international oil prices, while national average retail prices for gasoline and diesel serve as proxies for domestic fuel prices. Therefore, the primary contribution of this study lies in providing a better understanding of how volatility in international oil prices is transmitted to domestic fuel prices in Brazil.

In order to identify how international oil price volatility is transmitted to domestic fuel prices, this study employs a multivariate GARCH model specification. Specifically, a BEKK specification with a Bayesian approach is adopted, as proposed by Rast et al. (2020). This model allows for capturing the co-volatilities present in domestic fuels (gasoline and diesel) arising from fluctuations in international oil prices, represented here by WTI crude oil. Therefore, two models were estimated to capture the effects of oil volatility: one models the relationship between oil and gasoline, and the other between oil and diesel.

In the Brazilian context, the use of petroleum-based fuels such as diesel and gasoline is particularly significant due to the country's continental dimensions. Moreover, Brazil exhibits a high dependence on road transportation, which accounts for approximately 70% of the national transport network (Barbosa et al., 2023; National Association of Railway Transporters, 2024).

Beyond interregional transportation issues, petroleum derivatives also have a direct impact on Brazilian household budgets. According to Brazilian Institute of Geography and Statistics (2024), (IBGE) from the acronym in Portuguese, transportation holds the second largest weight in the Broad National Consumer Price Index (IPCA) — Brazil's main measure of inflation — accounting for 20.68% of the total in 2024. Furthermore, within the transportation category, gasoline represents the highest individual weight at 5.25%.

Inflationary pressures resulting from increases in gasoline and diesel prices have sparked debates regarding the appropriate fuel pricing policy in Brazil. Historically, since the country's re-democratization, all Brazilian governments have remained attentive to fluctuations in fuel prices, aiming primarily to prevent inflationary spirals caused by rising gasoline and diesel costs (Gauto; Delgado; Couto, 2021).

In methodological terms, multivariate GARCH models with a BEKK specification have been used to analyze fuel price volatility in studies by Elder (2004), Elder and Serletis (2010), Wu and Li (2013), Chang and Serletis (2018), Yosthongngam, Tansuchat and Yamaka (2022). Each of these studies sought to examine the oil–fuel relationship and the transmission of volatility. In the present study, applying a Bayesian approach to the BEKK model offers a novel perspective on how international oil price volatility is transmitted to domestic fuel prices (gasoline and diesel). The use of the Bayesian approach is due to the limited number of applications of BEKK models, for example, see Yosthongngam, Tansuchat and Yamaka (2022). Consequently, this work fills a gap in the literature regarding the volatility of international oil prices and their relationship with domestic fuel prices in the Brazilian context. Thereby expanding understanding of how international oil volatility impacts fuel price dynamics in Brazil and affects end consumers.

Beyond this introductory section, the paper is organized into five additional sections. The second section presents a literature review on volatility in the oil market. The third section outlines the empirical strategy employed to measure volatility. The fourth section describes the dataset used in this study. The fifth section presents the results, along with their discussion and policy implications. Finally, the sixth and last section presents the study's concluding remarks.

2.2 Fuel Volatility and the Oil Market

Volatility in the oil and fuel sectors has been widely studied, with studies addressing this issue across different countries. Table 2.1 summarizes the main studies on the subject and their key findings.

The literature review indicates that interest in fuel price volatility dates back to the late 1980s. The pioneering study by Bacon (1991) introduced the identification of the behavior between oil and gasoline prices, known as the "Rockets and Feathers" phenomenon, which posits that fuel prices rise like a rocket (Rockets) following increases in oil prices, but fall like feathers (Feathers) when oil prices decline. This study demonstrated how volatility can affect the relationship between oil and gasoline.

In subsequent years, other authors continued to study the impact of oil price volatility and its effects on fuels (gasoline, diesel, ethanol, and others) in order to assess the effects of volatility on these commodities (see Chevallier (2012), Lamotte et al. (2013), Wu and Li (2013), Boroumand et al. (2015), Rahman (2016)). The general conclusions indicate that oil volatility positively and significantly affects the volatility of gasoline and other fuels, demonstrating that a relationship exists between them.

In addition, it is worth noting that international volatility in oil prices can have different effects on economies. In an oil-importing economy, a positive oil-price shock increases remittance inflows, whereas in an oil-exporting nation, the effect depends on which oil price (Brent or WTI) is used (Dada; Akinlo, 2023). These effects in different types of oil economies have also been investigated in other studies, such as Aastveit, Bjørnland and Thorsrud (2015), Barbosa et al. (2023).

Some authors have analyzed how oil price volatility can affect the prices of energy commodities and, in particular, agricultural commodities. The studies by Wu and Li (2013), Carpio (2019), Barbaglia, Croux and Wilms (2020) showed that oil and its volatility can influence the prices of these commodities in general, and that some agricultural commodities can also affect the prices of energy commodities (gasoline, diesel, and ethanol). This establishes a bi-directional volatility relationship, in which oil and its volatility influence commodities, while commodities and their volatility affect oil.

Furthermore, it is worth noting that volatility analysis is not restricted to a single analytical method, as authors employ a wide range of approaches, as shown in Table 1. The most commonly used methods for volatility analysis are the VAR family methods and GARCH (Univariate and multivariate models), for example, including a combination of these methods (bivariate GARCH-in-mean VAR), as presented in Elder and Serletis (2010), Rahman (2016). In Elder and Serletis (2010), the authors explored how increases in oil uncertainty affect GDP in the USA. Otherwise, in Rahman (2016), the goal was to measure the impact of oil volatility on gasoline prices in the USA. Finally, Mensi, Rehman and Vo (2021) used a Generalized Vector Auto-regression Model (GVAR) to examine the correlation degree between future oil prices and other energy markets, especially gasoline prices. The authors found that oil and gasoline prices exhibit a high degree of correlation, particularly during periods of crisis.

Regarding the MGARCH models, Chevallier (2012) examined how oil shocks and their volatility can affect the natural gas and CO2 futures markets in France. The author used the specification types within the MGARCH models (BEKK, CCC, and DCC) to identify the best-fitting model for oil volatility spillovers. And Wu and Li (2013) examined the effects of oil volatility on the corn and ethanol markets in China. The conclusion is that oil volatility positively affects the corn and ethanol prices in China. Chang and Serletis (2018) concentrated their analysis on the USA gasoline market, seeking to understand how oil volatility affects retail gasoline prices. Finally, Yosthongngam, Tansuchat and Yamaka (2022) examined spillovers between the corn and ethanol markets in the top three ethanol markets in the world (Brazil, China, and the USA). The authors used a BEKK specification and Bayesian inference to calculate the parameters. The conclusion was that ethanol and corn have bidirectional effects in all countries analyzed.

Whereas, in this study, the main goal is to verify how oil volatility spillover affects retail gasoline and diesel prices in Brazil. To reach it, the BEKK specification was used, along with Bayesian inference to estimate the model parameters. This analysis is new in the Brazilian context, as is the use of Bayesian estimation to measure oil volatility and its impacts on domestic fuel prices.

Table 2.1 – Summary of the Literature Review

Study	Time	Variables	Countries	Methodologies	Main results
Fuel and Oil Sector Volatility					
Bacon (1991)	1982-1989 (Monthly)	Oil price and retail gasoline price	United Kingdom	Partial adjustment asymmetric model	The author concluded that there is an asymmetry between oil and gasoline in the relationship of price increases and decreases.
Elder & Serletis (2010)	2/1974 – 1/2008 (Quarterly)	Oil WTI price and Real GDP	United States	Bivariate GARCH-in-mean VAR	The results showed that an increase in oil price uncertainty negatively and significantly affects real GDP.
Chevallier (2012)	22/04/2005–15/12/2008 (Daily).	Oil, natural gas and future CO_2 price	Europe (European Central Bank)	Multivariate GARCH, specification: BEKK, CCC and DCC.	Among the specifications, the DCC model proved to be the most effective in capturing volatility and the relationships between the variables.
Lamotte, Porcher, Schalck & Silvestre (2013)	04/05/1990 – 29/04/2011/ weekly	Gasoline, diesel, Brent oil in Euro and Dollar	France	ARDL	The authors found that oil shocks have a greater impact on gasoline than on diesel. It was not possible to identify periods of low and high volatility.
Wu & Li (2013)	05/07/2003 – 31/08/2012 / daily	Oil, ethanol and corn price.	China	EGARCH and BEKK - MVGARCH	The authors identified that there is a volatility spillover among oil, corn, and ethanol

Continues on next page

Study	Time	Variables	Countries	Methodologies	Main results
Rahman (2016)	January/1978 – november/2014 (Monthly)	Oil and gasoline price.	United Sates	Bivariate GARCH (1,1) in-mean SVAR	It was found that an increase in oil volatility raises gasoline volatility, in addition to generating asymmetric effects on prices
Boroumand, Goutte, Porcher & Porcher (2015)	04/05/1990 – 29/04/2011 (Weekly)	Brent crude oil price, retail diesel price, Euro/USD exchange rate	France	Markov-Switching	The price transmission between oil and diesel is stronger during periods of low volatility than during periods of high volatility
Chang & Serletis (2018)	January/1976 – September/2014 (Monthly)	WTI crude oil price and retail gasoline price	United States	Multivariate GARCH: specification BEKK.	The results confirm the existence of an asymmetric relationship between oil and gasoline in the USA; moreover, oil price uncertainty has a positive effect on gasoline prices.
Carpio (2019)	July/2001 – December/2018 (Monthly)	Prices of oil, gasoline, ethanol, and sugar	Brazil	VARX, VECX, and Volatility indices calculated by the EFA	Oil volatility has a limited impact on gasoline, ethanol, and sugar due to government interventions during price adjustments
Barbaglia, Croux & Wilms (2020)	03/10/2012 – 28/10/2016 (Daily)	Commodity prices: oil, natural gas, corn, soybeans, coffee, cotton, wheat, and sugarcane.	United States	VAR, t-lasso estimator, volatility indices	The results showed that volatility is bidirectional between energy and agricultural commodities. Increases in oil lead to rises in gasoline and agricultural commodity prices

Continues on next page

Study	Time	Variables	Countries	Methodologies	Main results
Mensi, Rehman & Vo (2021)	02/01/1997 - 08/02/2021 (Daily)	Energy Future Markets: WTI crude oil, Brent crude oil, Heating oil, gas oil, gasoline, and natural gas.	Global	Wavelet Decomposition and Generalized Vector Auto-Regression (GVAR)	The model's results indicate that oil and gasoline exhibit a strong correlation, particularly during crises. So, the high volatility present in oil prices spills over to gasoline prices.
Yosthongngam, Tansuchat & Yamaka (2022)	01/01/2015 – 31/12/2020/ daily	Corn and ethanol prices	Brazil, China and United States	Multivariate GARCH: BEKK (1,1)	The results showed that volatility between corn and ethanol is positive and bidirectional for all countries analyzed.

Source: Elaborated by the author.

2.3 Empirical Strategy

2.3.1 Multivariate GARCH Model

GARCH models were developed by Bollerslev (1986) with the purpose of measuring volatility in a time series. Although the model captures the volatility present in the series, it is limited to analyzing only a single series. To extend volatility measurement to a multivariate context, Bollerslev, Engle and Wooldridge (1988) developed the Multivariate GARCH model. For Multivariate GARCH models, there are three main specifications: BEKK (Baba, Engle, Kraft, and Kroner), proposed by Engle and Kroner (1995); CCC (Constant Conditional Correlation), proposed by Bollerslev (1990); and DCC (Dynamic Conditional Correlation), proposed by Engle (2002). The use of the BEKK model is due to its ability to model the volatility of the entire system; therefore, it can model the volatility of fuel prices simultaneously. Another model that can be used is DCC; however, this model can only model volatility using separate fuel series. This way, the BEKK is the only model that can capture the system volatility and its effects, considering all series used. Therefore, the use of this specification to model the volatility were used in the following studies: (Elder, 2004; Wu; Li, 2013; Chang; Serletis, 2018; Yosthongngam; Tansuchat; Yamaka, 2022).

Thus, a BEKK (1,1) specification was employed to capture the volatilities present in the analyzed series. The BEKK model is represented by Equation (2.1).

$$H_t = C'C + A'(\epsilon_{t-1}\epsilon'_{t-1})A + B'H_{t-1}B \tag{2.1}$$

Where: H_t is $d \times d$ matrix representing the conditional variance-covariance of the model at $t = 1, ..., T$; C is a symmetric and positive definite $d \times d$ matrix that captures the constant (baseline) component of the variance–covariance matrix; ϵ_t is a $d \times 1$ vector of residuals associated with the covariance matrix H_t, augmented by an independent white noise process; and finally, A and B are $d \times d$ parameter matrices that govern, respectively, the influence of past shocks and past conditional variances on the dynamics of H_t.

Equation (2.1) is rewritten in matrix form, as shown in Equation (2.2):

$$\begin{aligned} H_t &= \begin{bmatrix} \sigma_{11,t} & \sigma_{12,t} \\ \sigma_{21,t} & \sigma_{22,t} \end{bmatrix} \\ &= \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{\top} \begin{bmatrix} \epsilon^2_{1,t-1} & \epsilon_{1,t-1}\epsilon_{2,t-1} \\ \epsilon_{2,t-1}\epsilon_{1,t-1} & \epsilon^2_{2,t-1} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \\ &+ \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}^{\top} \begin{bmatrix} h_{11,t-1} & h_{12,t-1} \\ h_{21,t-1} & h_{22,t-1} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \end{aligned} \tag{2.2}$$

Where: each sigma $(\sigma_{11}, ..., \sigma 22)$ represents the conditional variance present for gasoline/diesel and oil, where each conditional variance is in t.

The coefficients in matrix C represent the intercepts of matrix H_t, capturing the constant variances within the model. Furthermore, it is correct to state that C is a $d \times d$ positive definite and symmetric matrix.

Matrix A captures the ARCH effect present in the system. Accordingly, the element a_{12} measures the impact of past shocks in the price of oil on the conditional variance of gasoline, while a_{21} captures the impact of past shocks in the price of gasoline on the conditional variance of oil. In this context, the elements a_{11} and a_{22} measure the impact of past shocks in gasoline and oil prices, respectively, on their own current conditional variances.

Considering matrix B, it captures the GARCH effect, which measures the persistence of volatility over time. Thus, b_{12} indicates the extent to which past volatility in oil affects future volatility in gasoline, while b_{21} measures how past volatility in gasoline influences future volatility in oil. Similarly, b_{11} represents the impact of past gasoline volatility on its current volatility, whereas b_{22} reflects the effect of past oil volatility on its own current volatility. This analysis applies analogously to the relationship between oil and diesel prices in Brazil.

In summary, the ARCH effect measures how a past shock in period $(t-1)$ affects the conditional variance in period t, while the GARCH effect captures the persistent impact of past conditional variances in periods $(t-1, t-2, ..., t-n)$ on the conditional variance in period t.

The BEKK specification also incorporates a structure for modeling the mean dynamics of the series. In the present study, since the BMGARCH package developed by Rast et al. (2020) is employed, three possible specifications are available for the mean equation: a constant-only model, a VAR (1) model, and an ARMA (1,1) model. The package estimates the parameters using Bayesian inference. The use of the Bayesian approach is motivated by high dimensionality and strong dependence among the parameters; this may complicate inference with maximum-likelihood methods (Bayarri; Berger, 2004). Moreover, the lack of Bayesian applications for modeling volatility in fuel markets was one reason to apply this type of inference in this study.

The priors used are described in Rast et al. (2020), which consider a VAR(1) model. The model is defined by equation (2.3):

$$Y_t = \phi_0 + \Phi Y_{t-1} \tag{2.3}$$

Where: y_t represents the vector of variables included in the VAR model, ϕ_0 denotes the $K \times 1$ vector of constant intercepts, and Φ is the $K \times K$ matrix of autoregressive coefficients.

Furthermore, the priors for ϕ_0, Φ, also the priors to matrix A, B, and C are presented in table 2.2.

Table 2.2 – Prior distributions

Prior	**Distribution**	**Informative**
Matrix A	$Uniform(a_l, a_u)$	
Matrix B	$Uniform(b_l, b_u)$	
Matrix C	$B \sim Normal(0, 3I)$	Matrix C is divided into standard deviation (s) and correlation matrix (R_c). $$\text{Diag}(s) = \exp(XB)$$ and $$R_c \sim \text{LKJ}(1)$$ is the prior of the correlation matrix.
Mean equation		
ϕ_0	$N(\text{mean}(Y_{k1}), \text{var}(Y_{k1}))$	$\text{mean}(Y_{k1})$ and $\text{var}(Y_{k1})$ are the empirical mean and variance of the k-th time series at the initial point ($t = 1$). It is assumed that the elements of ϕ_0 are a priori independent.
Φ	$Normal(0, 1)$	For all $i, j = 1, \ldots, k$, this prior distribution implies that all entries of Φ are independently and identically distributed as $$Normal(0, 1)$$

Source: Elaborated by the author.

Furthermore, this study employed a Markov Chain Monte Carlo (MCMC) sampling algorithm for gasoline and diesel in the volatility modeling process, using Hamiltonian maximization. The MCMC algorithm relies on the Hamiltonian Monte Carlo (HMC) method proposed by Neal (2011), which aims to generate random samples to approximate the posterior distribution of the variables of interest. The functioning of this algorithm is inspired by physics, particularly Hamiltonian mechanics, where each parameter is treated as a particle moving through the probability distribution using a gradient. The use of gradients allows the algorithm to efficiently explore high-probability regions of the posterior distribution without relying solely on random steps, as in the Metropolis–Hastings algorithm. Ultimately, HMC enables the exploration of complex probability distributions more efficiently, yielding an accurate approximation of the posterior.

2.4 Database

In this study, to examine oil spillover in Brazilian national fuel prices, the database used comprised monthly national gasoline and diesel pump prices and the international oil price (WTI), as used in (Barbosa et al., 2023). The period analyzed was from July 2001 to August 2023. Moreover, an interpolation average was used to calculate gasoline and diesel prices for September 2020, as COVID-19 prevented price collection that month. A complete description of the data is provided in Table 2.3.

Finally, the observations were transformed to measure the growth rate of fuel prices: the original time series was taken in log form, first-differenced, and multiplied by 100.

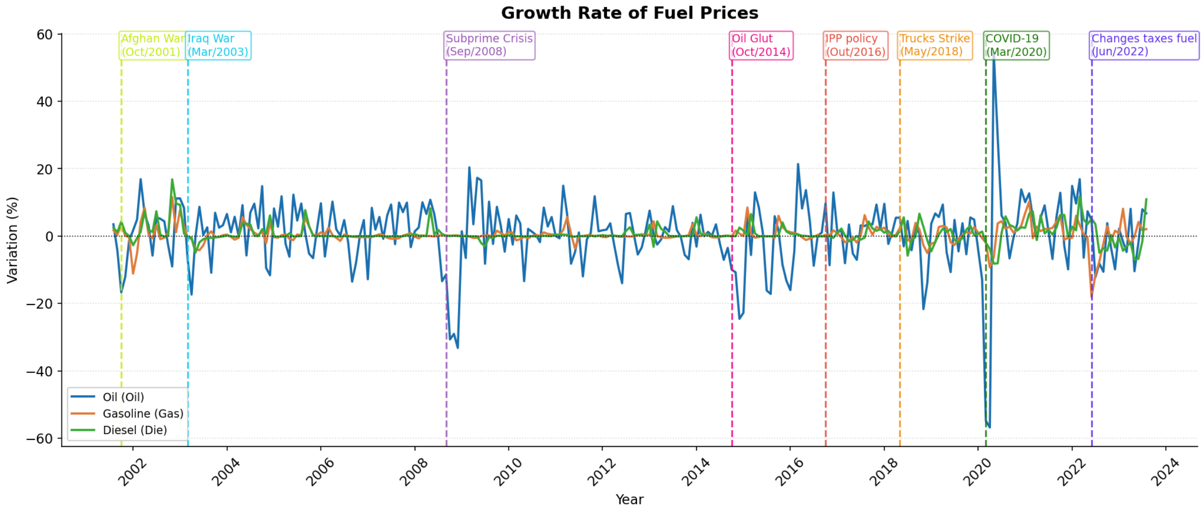
Table 2.3 – Description of the variables

Variable	**Series**	**Source**
Gasoline	Monthly historical series (R$/L)	ANP (<https://www.gov.br/anp/pt-br>)
Diesel	Monthly historical series (R$/L)	ANP (<https://www.gov.br/anp/pt-br>)
Oil	Crude Oil Prices: West Texas Intermediate (WTI) - Cushing, Oklahoma (US$/bbl)	FED Saint Louis (<https://fred.stlouisfed.org/series/MCOILWTICO>)

Source: Elaborated by authors.

Figure 2.1 presents the growth rate of fuel prices. The purpose is to illustrate the growth rate of fuel prices over time, allowing for the identification of large variation points.

Figure 2.1 – Growth Rate of Fuel Prices



Source: Elaborated by the author.

According to Figure 2.1, there were eight periods over time when oil prices or Brazilian national prices had large variations in growth rates. It is important to note that oil price variations have been larger than fuel price variations, and there have been more instances when oil price variations were higher than fuel price variations.

This way, the Afghanistan War (2001), Iraq War (2003), Subprime Crisis (2008), Oil Glut (2015), and Covid-19 (2020) were the periods in which oil price variations exceeded fuel price variations. Moreover, these moments affected the whole world, so oil prices were the first to feel the impact.

Regarding national fuel variations, the magnitude of variation increased after the implementation of the IPP policy in October 2016, with the growth rate becoming more volatile. These variations in fuel price growth rates reached their peak in May 2018, during a Truckers' strike. Beyond this, in 2022, the government removed federal taxes from fuel prices and set a cap of 17% to 18% on state taxes. This measure immediately reduced fuel prices (Brazilian Chamber of Deputies, 2022).

2.5 Results

2.5.1 Multivariate Model Estimation

In the present study, the BEKK specification was estimated using 15,000 draws, with the first 7,500 discarded. Tables 2.4, 2.5, and 2.6 present the main results for the matrix coefficients C, A, and B, respectively. In each table, the reported values correspond to the posterior mean in the first column, the posterior standard deviation in the second, the posterior median in the third, the 2.5% and 97.5% credible intervals for the median in the fourth and fifth, and the $\hat{R}$ values in the sixth.

In addition, it is important to emphasize that the standard deviation reported by the model refers to the posterior mean of the standard deviations of the parameter distributions estimated by the BEKK model. This way, the standard deviation indicates the degree of uncertainty associated with each model parameter. Therefore, the closer the standard deviation is to zero, the more concentrated the posterior distribution is around its mean (Gelman et al., 2013).

The $\hat{R}$, also known as the Gelman–Rubin convergence diagnostic or potential scale reduction factor $\hat{R}$, is a metric used to assess the convergence of Bayesian models, particularly those estimated through Markov Chain Monte Carlo (MCMC) methods. Essentially, $\hat{R}$ evaluates whether the different chains have converged to the same posterior distribution. Convergence is achieved when the within-chain variance (intra-chain) becomes close to the between-chain variance (inter-chain). In practice, an $\hat{R}$ value close to 1 indicates that the chains have converged properly (or, in other words, are well mixed) (Gelman; Rubin, 1992).

Regarding the Markov Chain, only one was used; normally, four are used, as suggested in Rast et al. (2020). Gelman and Rubin (1992) emphasizes the necessity to use at least two Markov chain to generate the posterior estimation and estimate the $\hat{R}$ value. The expected value of $\hat{R}$ is one, which means that both chains converged in the posterior. However, in this study, it was not possible to use two or more Markov chain because they were computationally intensive, and even with a large number of iterations, they did not converge. Despite this, the BEKK-MGARCH was implemented to allow for a single Markov Chain, and the algorithm splits into two segments, enabling the calculation of $\hat{R}$ and the estimation of the posterior. Finally, the interpretation of columns 1–7 in Table 2.4 also applies to columns 1–7 in Tables 2.5 and 2.6.

Table 2.4 – Constant variance components associated with matrix C for gasoline and diesel

	Mean	S. deviation	Median	2.5%	97.5%	$\hat{R}$
Gasoline						
var_{gas}	0.02	0.02	0.02	0.00	0.07	1
var_{oil}	22.39	8.49	21.51	8.76	40.85	1
$corr_{gas/oil}$	0.02	0.53	0.03	-0.93	0.93	1
Diesel						
var_{die}	0.01	0.00	0.01	0.00	0.02	1
var_{oil}	20.51	7.62	19.80	7.73	37.25	1
$corr_{die/oil}$	0.32	0.42	0.35	-0.65	0.97	1

Source: Elaborated with results of the research. These coefficients (var_{gas} and var_{die}) represent constant variance for gasoline and diesel, and *corr* is the constant correlation between oil and gasoline/diesel. Both coefficients are made from the C matrix.

Table 2.4 presents the results for the components present in the C matrix of gasoline and diesel, which captures the constant variances in the model. In Table 2.3, the estimated mean values for the constant conditional variance var_{oil} related to gasoline (22.39) and diesel (20.51) are quite similar, as are their median values. This suggests that crude oil affects both fuels—gasoline and diesel—in a similar manner. Furthermore, the estimated mean values for gasoline and diesel var_{gas} or var_{die}, which represent the constant conditional variance, are close to zero for both fuels (gasoline = 0.02 and diesel = 0.01), indicating that the observed volatilities are mainly explained by the ARCH and GARCH effects. Finally, the constant conditional correlations between oil/gas and oil/die were 0.02 and 0.32, respectively. For both fuels, the correlation was positive, but for gasoline, it was near zero, indicating weak contemporaneous comovement between the series.

Table 2.5 – Coefficients of the A matrix for gasoline and diesel

	Mean	S. Deviation	Median	2.5%	97.5%	$\hat{R}$
Gasoline						
a_{11}	0.25	0.06	0.25	0.16	0.38	1
a_{12}	-0.03	0.30	-0.04	-0.59	0.60	1
a_{21}	0.01	0.01	0.01	0.00	0.03	1
a_{22}	0.22	0.23	0.29	-0.42	0.48	1
Diesel						
a_{11}	0.59	0.07	0.59	0.47	0.74	1
a_{12}	-0.24	0.25	-0.24	-0.72	0.25	1
a_{21}	0.00	0.00	0.00	-0.01	0.00	1
a_{22}	0.25	0.07	0.25	0.11	0.38	1

Source: Elaborated with results of the research.

Table 2.5 captures the ARCH effect for gasoline and diesel. The interpretation is that for a_{11}, a random shock to the gasoline price in the previous period results in an increase of 0.25 in the conditional variance in the current period. Similarly, for diesel, a shock to its price at $t-1$ leads to an increase of 0.59 in the conditional variance at period t.

Considering the ARCH effect for crude oil and its past shocks a_{22}, the estimated values for the gasoline/oil relationship are 0.22 and for the diesel/oil relationship 0.25. Therefore, past own shocks in gasoline and diesel affect their respective conditional variances with similar intensity.

The coefficient a_{12} captures the cross-market ARCH effect, measuring how past shocks in the oil market are transmitted to the current conditional variance of gasoline (diesel). The estimated coefficients are negative, indicating that oil price shocks reduce current conditional volatility. Specifically, the coefficient is 0.04 for gasoline, suggesting a relatively small negative spillover effect, whereas for diesel the coefficient is 0.24, indicating a stronger negative transmission of oil price shocks to diesel's conditional variance.

Finally, the coefficient a_{21} captures how past shocks in gasoline/diesel affect the conditional variance of oil. It was observed that a_{21} for gasoline was 0.01 and 0.00 for diesel, indicating that past shocks in these fuels do not impact the conditional variance of oil. These results were expected, given that Brazil is a small open economy where domestic fuel prices do not influence the international oil price.

Moreover, the results indicate that diesel is more volatile to changes in oil prices compared to gasoline. One possible explanation for this greater sensitivity of diesel is that gasoline contains a percentage of ethanol in its blend. Therefore, gasoline prices do not depend solely on oil but also on sugarcane prices. This relationship has been documented in the literature: Carpio (2019) highlights the interdependence between gasoline and ethanol in Brazil, while Barbaglia, Croux and Wilms (2020) also identify this relationship between gasoline, oil, and ethanol/sugarcane in the United States.

Table 2.6 – Coefficients of the B matrix for gasoline and diesel

	Mean	S. Deviation	Median	2.5%	97.5%	$\hat{R}$
Gasoline						
b_{11}	0.86	0.05	0.87	0.76	0.94	1
b_{12}	0.70	0.53	0.69	-0.33	1.75	1
b_{21}	0.00	0.02	0.00	-0.03	0.04	1
b_{22}	-0.49	0.29	-0.57	-0.82	0.47	1
Diesel						
b_{11}	0.54	0.05	0.54	0.44	0.63	1
b_{12}	0.63	0.42	0.62	-0.18	1.48	1
b_{21}	0.00	0.01	0.00	-0.01	0.02	1
b_{22}	-0.57	0.20	-0.60	-0.84	-0.10	1

Source: Elaborated with results of the research.

Table 2.6 presents the GARCH effect. The GARCH effect measures the extent to which past volatilities $(t-1, t-2, t-3, ..., t-n)$ influence the current level of volatility at period t. Consequently, high values of the coefficients b indicate that volatility is persistent over time, whereas low values suggest that volatility does not persist.

Given that b_{11} represents the coefficient capturing the extent to which past gasoline volatilities affect current gasoline volatility, its value is 0.86. In a case, for diesel, the b_{11} coefficient is 0.54. This indicates that gasoline volatility exhibits a high level of persistence, as it carries over volatilities from previous periods, whereas diesel shows a lower level of persistence. Regarding oil, b_{22}, the coefficient measuring how past volatilities impact current volatility, shows values of -0.49 (gasoline/oil relationship) and -0.57 (diesel/oil relationship). These statistically significant coefficients indicate the presence of volatility spillover between the respective markets. Due to the quadratic structure of the BEKK model, however, the negative signs are not interpreted individually.

The coefficient b_{12} measures the extent to which past oil volatilities affect current gasoline/diesel volatility. The results were 0.70 (gasoline) and 0.63 (diesel), indicating that both fuels inherit a substantial portion of volatility from past oil volatilities.

So, based on the results of a_{12} and b_{12}, it is possible to see a difference between them. In the case of oil price shocks, volatility has a negative impact on gasoline and diesel prices; however, the accumulation of volatility in oil leads to higher volatility in gasoline and diesel. Thus, a single oil shock can reduce volatility for gasoline and diesel, but the sum of volatilities from oil shocks has mostly passed to gasoline and diesel volatility.

Finally, the coefficient b_{21}, which measures the extent to which past gasoline/diesel volatilities affect current oil volatility, was found to be 0 for both gasoline and diesel. This indicates that neither fuel impacts the current volatility of oil, a result that is expected given that Brazil is a small open economy; hence, domestic fuel price changes do not influence international oil prices.

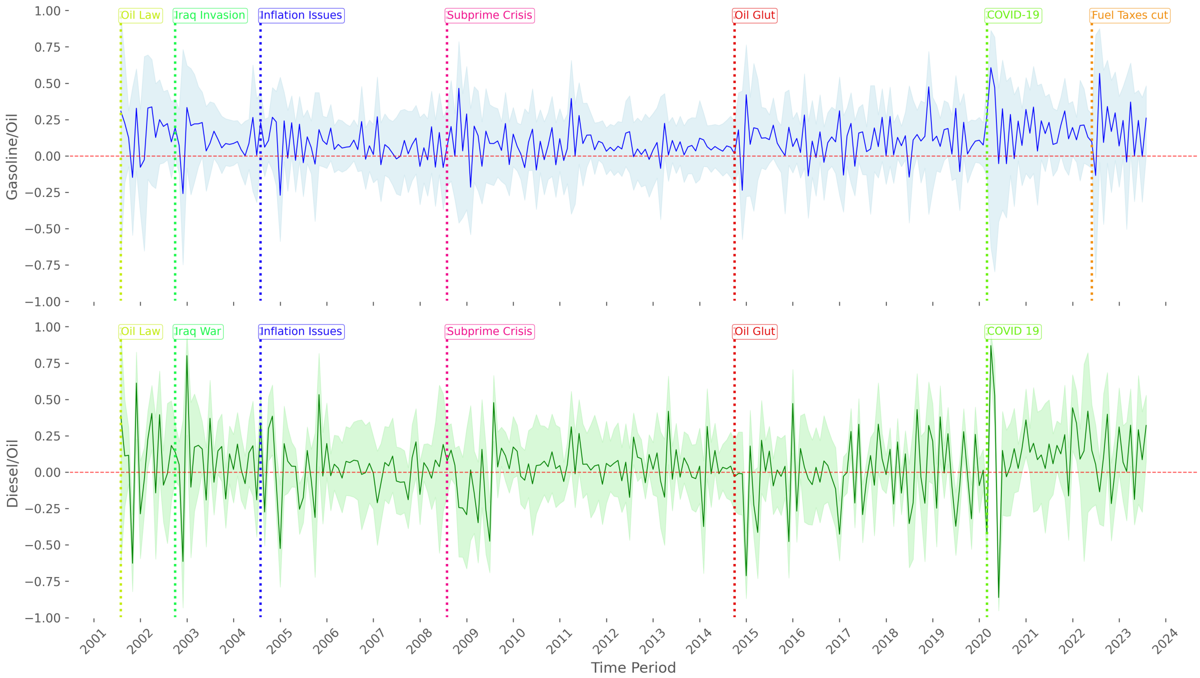
Moreover, it is important to note that in both effects (ARCH and GARCH), the uncertainty degree was very high. Once that, the posterior distribution was very wide, from negative to positive for many coefficients. This indicates that even if the mean or median has a positive (or negative) value, there is some chance that one of the posterior possibilities has a different signal.

In order to assess the robustness of the results, particularly for gasoline, an estimation was performed between the international gasoline price and the Brazilian gasoline price. The use of international gasoline for robustness checks is justified by the fact that a portion of Brazilian gasoline is of foreign origin, as indicated by Technical Note SDR/ANP No. 068 developed by National Agency of Petroleum, Natural Gas and Biofuels (2018), which establishes guidelines for fuel pricing in Brazil. Accordingly, the RBOB gasoline price was considered, with data available from the New York Mercantile Exchange (NYMEX).

Thus, based on the estimated parameters (the complete table is attached in the Appendix B), only the parameter a_{12} differed. In the oil/gasoline relation, it was -0.03, whereas in the international gasoline/Brazilian gasoline relation it was 0.19. This indicates that a shock at $t-1$ in international gasoline increases the conditional variance of domestic Brazilian gasoline by 0.19. Therefore, a past shock in the international gasoline price has a positive effect on the conditional variance of Brazilian gasoline, in contrast to the oil/gasoline relationship.

To illustrate the ARCH and GARCH effects produced by the BEKK-MGARCH model for both fuels (gasoline and diesel) and their relationship with the international oil price, the conditional correlations between oil and gasoline and between oil and diesel are presented in Figure 2.2. According to Rast et al. (2020), the conditional correlation, R_t, is a transformation of the conditional covariance, H_t, over time (as shown in Equation 2).

Figure 2.2 – Conditional Correlations between Gasoline, Diesel and Oil Prices



Source: Author's elaboration based on research results. The light gray area represents the 90% credibility interval, and the black line depicts the correlation measure, specifically the posterior mean estimated by the model.

From Figure 2.2, it can be observed that for 90% of the analyzed period, the conditional correlation between oil and gasoline is positive. This indicates that gasoline prices tend to follow the movements of oil prices. Furthermore, the estimated model successfully captured the price correlation between oil and gasoline, as periods of higher correlation amplitude correspond to times of international oil market stress, which consequently impacted Brazilian gasoline prices.

Figure 2.2 shows seven periods in which the conditional correlation experienced substantial changes, including episodes of negative correlation. These periods coincide with major national and international events that may have influenced the relationship between international oil prices and domestic fuel prices. For ease of interpretation, each event is identified in the figure with a flag: the Oil Law, the Iraq Invasion, Inflationary Issues, the Subprime Crisis, the Oil Glut, the COVID-19 pandemic, and the Russian invasion of Ukraine.

So, the first moment was primarily influenced by the 1997 Oil Law. This law established a liberalization of fuel prices, allowing distributors and retail fuel stations to set prices freely. Although enacted in 1997, the law had a transition period of 36 months, concluding in August 2000, and came fully into effect on January 1, 2002 (National Agency of Petroleum, Natural Gas and Biofuels, 2001). Moreover, given that the beginning of the sample is July 2001, the implementation and application of the law impacted the gasoline price series analyzed in this study, which justifies considering a law enacted years prior to the sample period.

The second period was characterized by two factors that contributed to increased gasoline price volatility. First, the 2002 presidential elections in Brazil featured two candidates with different economic proposals, which caused the exchange rate to rise and impacted the fuel and commodities markets in Brazil (Kanczuk, 2015). Additionally, on the international front, the 2003 invasion of Iraq created stress in the global oil market, affecting OPEC member countries, including Iraq, and leading to increases in the price per barrel of oil (Kanczuk, 2015).

The third period refers to the inflation issues in Brazil between 2004 and 2005. In this period, the country faced electric issues that increased the food and good prices. Trying to solve these questions, during this time, the government implemented a series of measures to combat inflation, including raising the SELIC interest rate and intervening in fuel prices (Kanczuk, 2015).

The fourth period occurred in 2008 due to the Subprime crisis in the United States. This crisis led to a recession in the country, causing a slowdown in the global economy, reducing demand for oil and affecting its prices. The fifth period is related to disagreements over oil production levels among OPEC members, particularly the Arab members. These disputes caused fluctuations in oil barrel prices between 2014 and 2015 (Energy Research Company, 2017; Energy Research Company, 2021).

The sixth period, beginning in February 2020, is related to the COVID-19 health crisis, which slowed down the global economy, reducing oil consumption and consequently causing a drop in prices. Finally, the Russian invasion of Ukraine in 2022 led Western markets to freeze oil trade with Russia, causing a surge in prices, given that Russia is a major oil producer (Miller; Kumleben; Nowak, 2024).

The conditional correlation between oil and diesel is also shown in Figure 2.2. There is a difference from the conditional correlation for oil/gasoline, given that, in oil/diesel, the correlation oscillates between positive and negative more often. In Brazil, there is a history of government intervention in diesel prices during periods of rising oil prices. Almeida, Oliveira and Losekann (2015) highlight that between 2011 and 2015, the Brazilian government maintained stable prices for Brazilian fuels, especially diesel, to meet the growing domestic demand for these fuels. Consequently, international oil price increases did not fully pass through to domestic diesel prices.

Furthermore, according to a study by FGV Europe Projects (2018), following the 2018 truck drivers' strike, the government froze the retail diesel price for 60 days post-strike in response to the truckers' demands, aiming to mitigate the impact of rising oil prices.

In addition to domestic price control measures, diesel experienced increased volatility (reflected in a higher amplitude of the conditional correlation) during certain periods that affected its price; each period is marked as a flag in Figure 2.2.

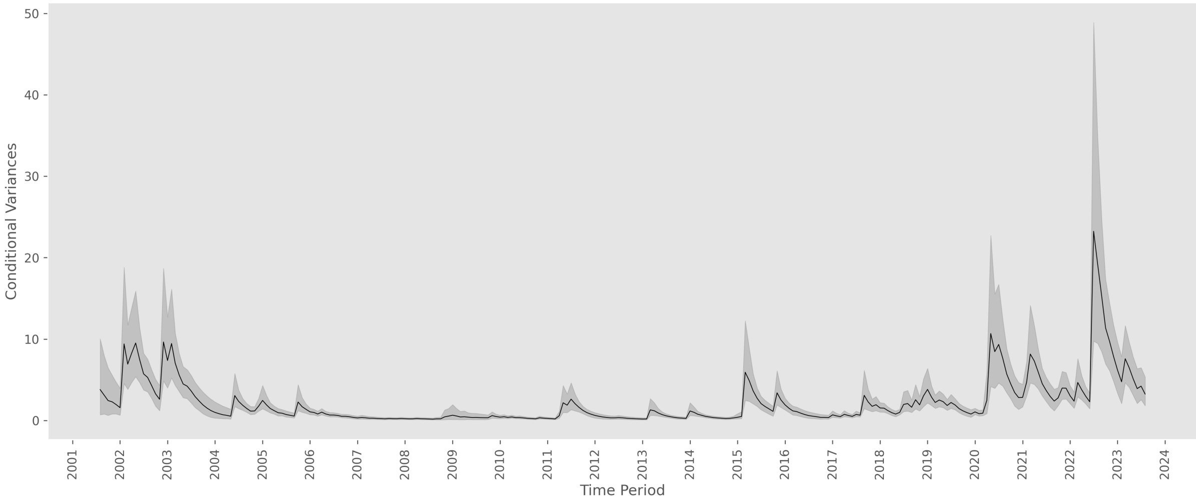
So, the first period is associated with internal factors that occurred in Brazil in 2001. Two events stand out: first, the Petroleum Law, enacted in 1997, came fully into force at the end of 2001, resulting in the termination of subsidies for gasoline and diesel. In addition, the country faced a severe hydroelectric crisis that caused blackouts and energy shortages. To address the lack of electricity, the government promoted the use of thermoelectric plants powered by diesel, which significantly increased the demand for this fuel and raised its price (National Agency of Petroleum, Natural Gas and Biofuels, 2001; Kanczuk, 2015). The second period corresponds to the U.S. invasion of Iraq between August 2002 and August 2003. Since Iraq is a major oil producer, the conflict generated disruptions and uncertainty in the global oil market, consequently leading to higher prices.

While, the third period refers to the interval between 2004 and 2005, characterized by inflationary pressures in Brazil. In this context, the government implemented price controls on fuels, including diesel and gasoline (Kanczuk, 2015). The fourth period is linked to the 2008 subprime crisis, which led to the collapse of key sectors of the global economy and also affected the oil and fuel markets. The fifth period, between 2014 and 2015, was marked by uncertainty regarding oil production levels in the Middle East, generating volatility and upward pressure on international oil prices (Energy Research Company, 2017; Energy Research Company, 2021).

Finally, the sixth period corresponds to the sanitary crisis caused by the COVID-19 pandemic (2020–2021), which resulted in a significant global economic slowdown and, subsequently, a sharp increase in oil prices following the recovery in demand.

In addition to the conditional correlations, there are also the conditional variances for the gasoline/oil and diesel/oil relationships. These measure how each variable within the system responds to the ARCH/GARCH effects captured by the BEKK model. In this context, the conditional variances reflect how the variance of each variable behaves, taking into account the ARCH and GARCH effects represented by matrices A and B, respectively, for each relationship (Rast et al., 2020). Figures 2.4 and 2.5 present the conditional variances for gasoline and oil, respectively.

Figure 2.3 – Conditional Variance of Gasoline



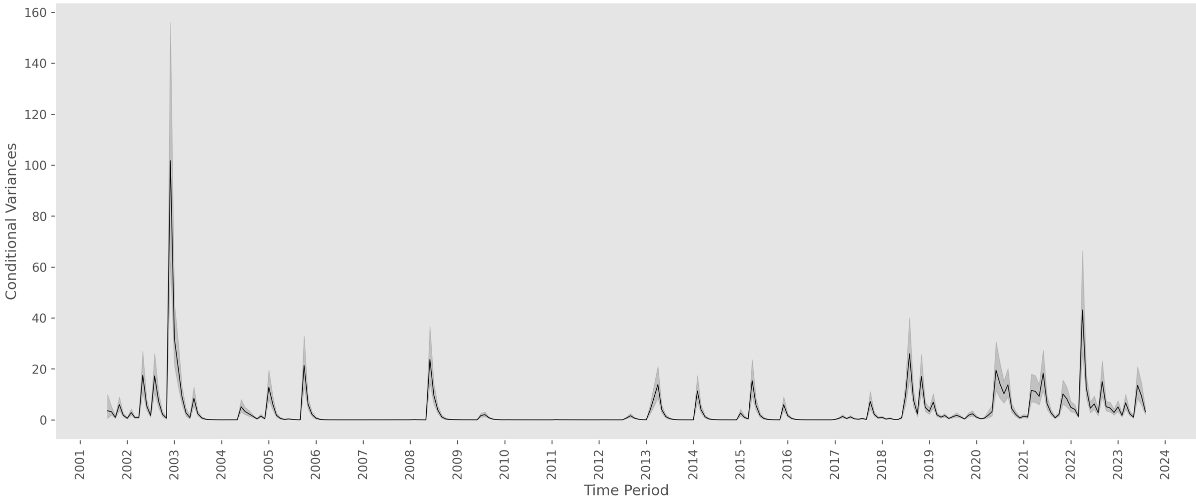
Source: Author's elaboration based on research results. The light gray area represents the 90% credibility interval, and the black line depicts the correlation measure, specifically the posterior mean estimated by the model.

Figure 2.4 presents the conditional variance for gasoline. According to (Rast et al., 2020), the conditional variance measures the level of volatility present in gasoline prices. It can be observed that there are specific periods (five in total) in which gasoline volatility experiences a significant increase.

Considering the flags marked in Figure 2.2, the first two volatility peaks are associated with two events: the 1997 Petroleum Law and the Iraq invasion, respectively. The second high-volatility period corresponds to production issues in the Arab world between 2014 and 2015. The third period is related to the COVID-19 pandemic, and finally, the last volatility peak is linked to the Russian invasion of Ukraine.

It is noteworthy that two events marked as flags in Figure 2.2 did not strongly impact gasoline price volatility to the extent observed in the four previously analyzed periods. These events were the inflationary pressures in Brazil between 2004 and 2005 and the Subprime Crisis in the U.S. in 2008. Furthermore, it is important to highlight that Figure 5 also shows periods of low volatility, which correspond to moments when gasoline prices remained stable, with no constant or abrupt increases relative to oil prices specifically, from the end of 2005 to March 2011.

Figure 2.4 – Conditional Variance of Diesel



Source: Author's elaboration based on research results. The light gray area represents the 90% credibility interval, and the black line depicts the correlation measure, specifically the posterior mean estimated by the model.

In Figure 2.5, seven periods can be identified in which the volatility of diesel prices in Brazil increased. By considering the events marked as flags in Figure 2.2, it is possible to link these events with the periods of elevated volatility observed in Figure 2.5.

The first period of diesel price volatility is linked to the 2003 Iraq invasion, corresponding to the second event in Table 2.8; the second period relates to the inflationary pressures in Brazil between 2004 and 2005; the third period corresponds to the 2008 Subprime crisis in the United States; the fourth period is associated with production problems in the Arab oil-producing countries between 2014 and 2015; the sixth period relates to the COVID-19 health crisis; and the seventh period corresponds to the Russian invasion of Ukraine in 2022.

Considering the flags in Figure 2.2 and the periods of high volatility depicted in Figure 2.5, the absence of the fifth period of elevated volatility is noticeable. It correspond to the truck drivers' strike in May and June 2018. The strike caused a diesel supply shortage across Brazil, driving prices up, with prices normalizing in the subsequent months (FGV Europe Projects, 2018).

Furthermore, there were periods of low volatility in diesel prices. For example, between the second and third periods of high volatility, there was an interval of relative stability, indicating that during this gap, diesel prices remained stable, end of 2005 until the start of 2008.

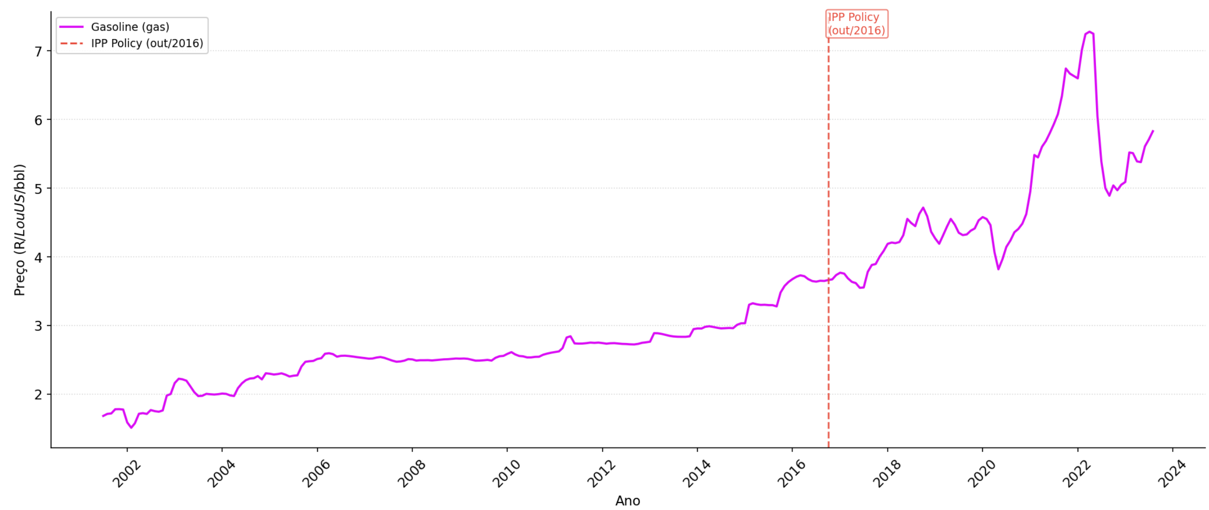
2.5.2 Estimations for Gasoline and Diesel during the IPP Period by Petrobras

Starting in October 2016, Petrobras implemented a new pricing policy called the IPP (Import Parity Price). The IPP is a methodology to calculate the price of a given product by considering the international price, the country's exchange rate, transportation costs, taxes, and port fees. Therefore, the objective of the IPP is to set a domestic market price that reflects the international cost of the product.

Based on this framework, the calculation of the IPP for domestic fuels, implemented by Petrobras in 2016, relied on four fundamental pillars: 1) the international price of crude oil; 2) the country's exchange rate (in Brazil's case, the Brazilian real/US dollar); 3) transportation costs (including port fees); and 4) a margin set by Petrobras itself, functioning as a type of safeguard to prevent financial losses for the company.

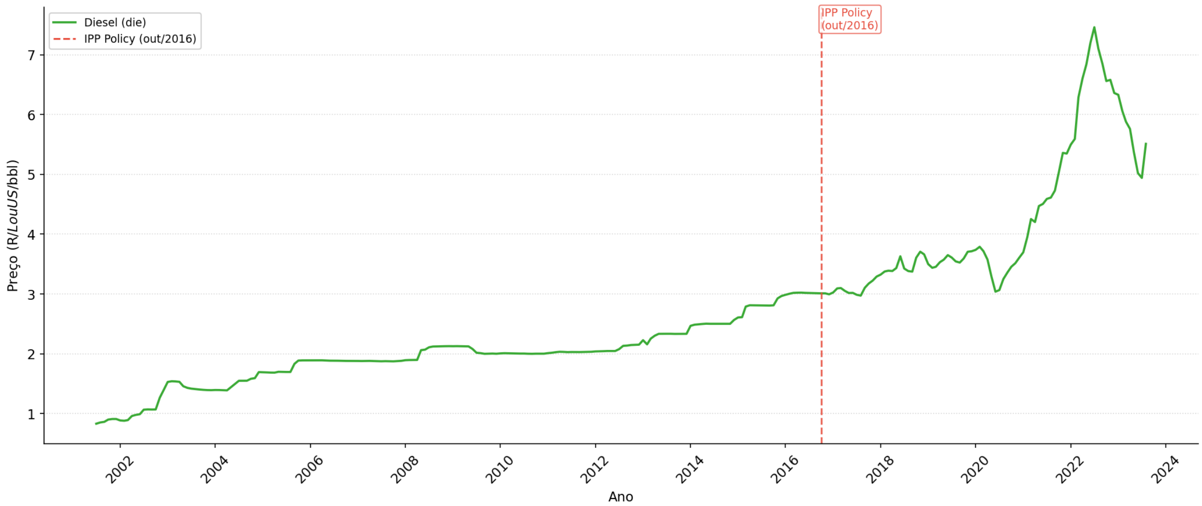
The implementation period of the IPP led to successive increases in gasoline and diesel prices in Brazil, as shown in Figures 2.6 and 2.7, which display the respective fuel prices. In these graphs, the red dashed line indicates the month when the IPP was implemented in the pricing of domestic fuels. These successive price hikes raised questions from civil society regarding the necessity of applying the IPP to domestic fuels and culminated in one of the most significant moments of tension under the IPP: the 2018 truck drivers' strike across several regions of Brazil.

Figure 2.5 – Gasoline price historical



Source: Elaborated by the author. The red dashed line indicates the start of Petrobras' IPP (October, 2016).

Figure 2.6 – Diesel price historical



Source: Elaborated by the author. The red dashed line indicates the start of Petrobras' IPP (October, 2016).

With the alignment of fuel prices in Brazil to the evolution of international oil prices, greater volatility was observed, suggesting that the parameters of the estimated BEKK model may have undergone substantial changes. Therefore, the model was re-estimated considering the period of the IPP implementation (October 2016 to May 2023).

However, during the estimation of the BEKK model, it was observed that the VAR (1) specification did not achieve convergence for the Markov chain. As an alternative, a Bayesian VAR (1) was estimated, using the R package BVAR. After estimation, the residuals were saved and used to estimate the volatility. In this volatility, the BEKK specification was used, with a mean structure as constant.

The Bayesian VAR used 15,000 iterations, excluding the first 5,000 (the first third of the iterations). This number of iterations was chosen because it represents the minimum required for the Bayesian VAR to produce consistent results (Kuschnig; Vashold, 2021). Additionally, the Minnesota prior and the Metropolis-Hastings selection algorithm were employed during estimation. The Minnesota prior was chosen due to its strong predictive power, as its distribution allows the reduction of the number of parameters to be estimated to only two hyperparameters (Kilian; Lütkepohl, 2017). Furthermore, the Metropolis-Hastings selection algorithm was used because of its ability to generate random draws to create a sequence of values approximating a desired distribution, as well as to eliminate potential autocorrelation among the selected observations (Metropolis et al., 1953).

Regarding the BEKK model, a single Markov chain was used for both fuels (gasoline and diesel); however, gasoline employed 80,000 iterations (excluding the first 40,000), while diesel used 240,000 iterations (excluding the first 120,000). This number of iterations was chosen as it represents the minimum required to estimate the model without issues related to Markov chain mixing or other estimation problems (Gelman; Rubin, 1992). This high number of interactions is due to the IPP sample being smaller than the entire sample. In this way, the volatility results for gasoline and diesel were obtained, as presented in Tables 2.7, 2.8, and 2.9.

First, the components of the C matrix are presented in Table 2.7. The C matrix corresponds to the intercept values of the BEKK model, where the intercept represents the constant variances. Thus, there are two coefficients in the C matrix: var_{gas} or var_{die}, which is the constant variance for gasoline and/or diesel, and var_{oil}, which is the constant variance for oil.

Table 2.7 – Constant variance components associated with matrix C for gasoline and diesel during the IPP

	Mean	S. Deviation	Median	2.5%	97.5%	$\hat{R}$
Gasoline						
var_{gas}	2.19	1.51	2.16	0.05	5.31	1
var_{oil}	20.09	14.18	19.95	0.11	49.06	1
$corr_{gas/oil}$	0.32	0.48	0.4	-0.84	0.97	1
Diesel						
var_{die}	0.84	1.15	0.38	0.01	4.33	1
var_{oil}	30.01	12.80	29.73	1.81	56.33	1
$corr_{die/oil}$	0.04	0.55	0.06	-0.93	0.94	1

Source: Elaborated from results of the research. These coefficients (var_{gas} and var_{die}) represent constant variance for gasoline and diesel, and *corr* is the constant correlation between oil and gasoline/diesel. Both coefficients are made from the C matrix.

Thus, the posterior mean values of the var_{gas} and var_{die}, respectively, coefficients for gasoline and diesel are 2.19 and 0.84, respectively, and are significantly higher than those observed in Table 2 for the full sample period. This increase in the components of the C matrix indicates that, considering the IPP period, the level of volatility in IPP corresponding to the constant is higher than that in the entire sample. So, it has no oil shock; the constant volatility present in fuels is high in IPP.

Table 2.8 – Coefficients of the A matrix for gasoline and diesel during the IPP

	Mean	S. Deviation	Median	2.5%	97.5%	$\hat{R}$
Gasoline						
a_{11}	0.26	0.16	0.24	0.02	0.62	1
a_{12}	0.47	0.50	0.51	-0.66	1.38	1
a_{21}	-0.01	0.05	-0.02	-0.10	0.08	1
a_{22}	-0.06	0.52	-0.28	-0.80	0.73	1
Diesel						
a_{11}	0.19	0.15	0.16	0.01	0.56	1
a_{12}	-0.13	0.81	-0.19	-1.49	1.44	1
a_{21}	0.00	0.03	0.00	-0.07	0.06	1
a_{22}	0.01	0.60	0.30	-0.85	0.86	1

Source: Based on results of the research.

Table 2.8 presents the ARCH effect for gasoline during the IPP period. The ARCH effect captures how past shocks—in this case, at time $(t-1)$ affect the current variance at time t. Thus, it can be observed that during the PPI period, there were changes in the values of some coefficients. For the A_{11} coefficient, the value for diesel was 0.19, indicating that, during the IPP period, a past shock in diesel leads to a 0.19 increase in conditional variance. For gasoline, the A value was 0.26.

For A_{12}, it is observed that during the IPP period, the value for gasoline was 0.47, while for diesel it was 0.13. Finally, it can be seen that for A_{21} there were no significant changes during the PPI period compared to the full sample period, with values close to 0.00.

Table 2.9 – Coefficients of the B matrix for gasoline and diesel during the IPP

	Mean	S. Deviation	Median	2.5%	97.5%	$\hat{R}$
Gasoline						
b_{11}	0.48	0.31	0.43	0.02	1.01	1
b_{12}	0.49	1.23	0.47	-1.98	2.72	1
b_{21}	-0.04	0.09	-0.04	-0.25	0.16	1
b_{22}	-0.08	0.40	-0.10	-0.78	0.67	1
Diesel						
b_{11}	0.39	0.25	0.36	0.02	0.92	1
b_{12}	0.39	0.84	0.37	-1.26	2.10	1
b_{21}	0.00	0.22	0.05	-0.31	0.29	1
b_{22}	0.03	0.27	0.04	-0.50	0.53	1

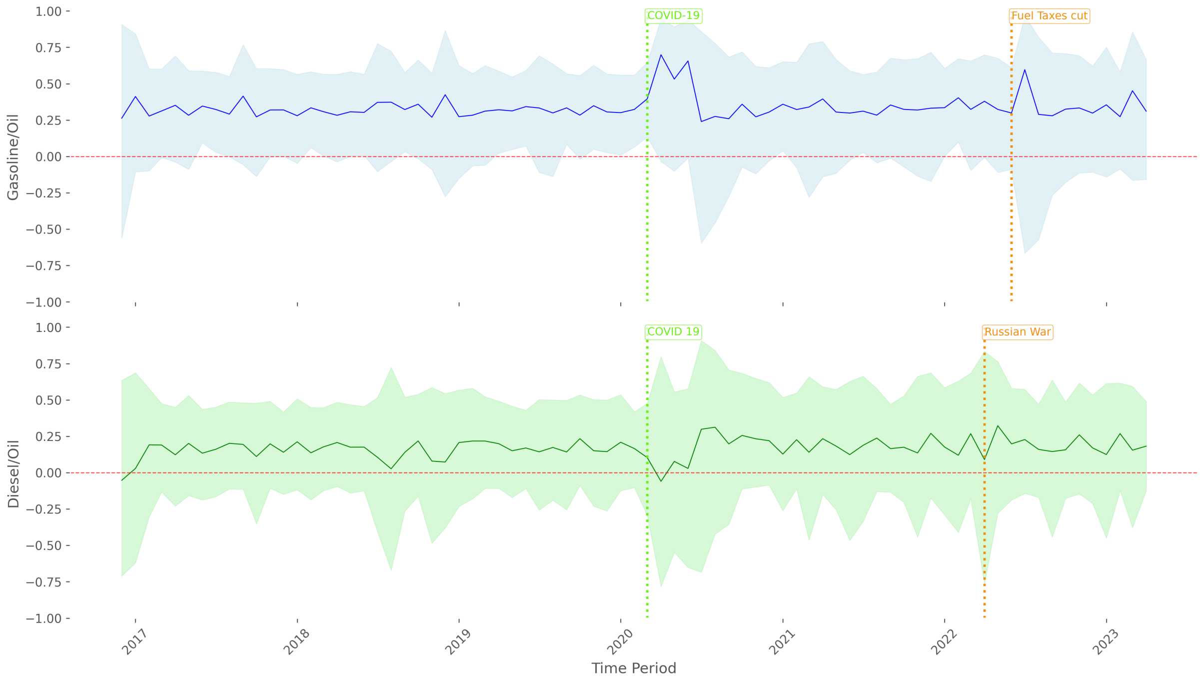
Source: Based on results of the research.

Next, Table 2.9 presents the GARCH effect, which measures how past volatilities affect current volatility. The fact that the GARCH effect considers past volatility to assess its impact on current volatility indicates that this effect captures the persistence of shocks over time.

Thus, from Table 2.9, the B_{11} coefficients can be observed, being 0.48 for gasoline and 0.39 for diesel. The B_{12} values were 0.49 for gasoline and 0.39 for diesel. The cross relationship between the domestic fuel and oil, represented by B_{21}, showed values of 0.04 for gasoline and 0.00 for diesel. Finally, the B_{22} values were 0.08 for gasoline and 0.03 for diesel. Therefore, it is important to note that the GARCH parameter has the same magnitude for oil and own fuels. So, past volatility in oil or in gasoline/diesel has the same impact on current volatility.

Additionally, analyses of conditional correlation between oil/gasoline and oil/diesel are presented in Figure 2.7, as well as the conditional variances for gasoline and diesel in Figures 2.8 and 2.9, respectively.

Figure 2.7 – Conditional Correlations between Gasoline, Diesel and Oil Prices in the IPP Period



Source: Based on the research results. The light gray area represents the 90% credible interval, and the black line represents the correlation.

Figure 2.7 presents the conditional correlation between gasoline and oil during the IPP and shows some differences compared to Figure 2.2. The first difference is that in Figure 2.7, the conditional correlation between gasoline and oil is entirely positive (black line), indicating that increases in oil prices were accompanied by increases in gasoline prices. In contrast, Figure 2.2 shows moments of inverse correlation, where oil prices fell while gasoline prices rose. The fact that the correlation remained positive throughout the IPP period was expected, given that gasoline prices were adjusted according to changes in oil prices, making a positive correlation during this period consistent.

The second difference is that, due to the shorter analysis period, Figure 2.7 shows only two periods of large variation in the correlation. The first period, during 2020, corresponds to the onset of the COVID-19 pandemic, which led to increases in the prices of various products, including oil. The second period refers to the Russian invasion of Ukraine in early 2022, as Western countries imposed sanctions on Russia, preventing the trade of petroleum products and causing oil prices to rise (Miller; Kumleben; Nowak, 2024).

Regarding the conditional correlation between oil/diesel, it was positive for almost the entire period, except at the onset of the COVID-19 pandemic, when there was a brief period of negative correlation. Additionally, there were three periods of large variation in the diesel/oil conditional correlation during the IPP.

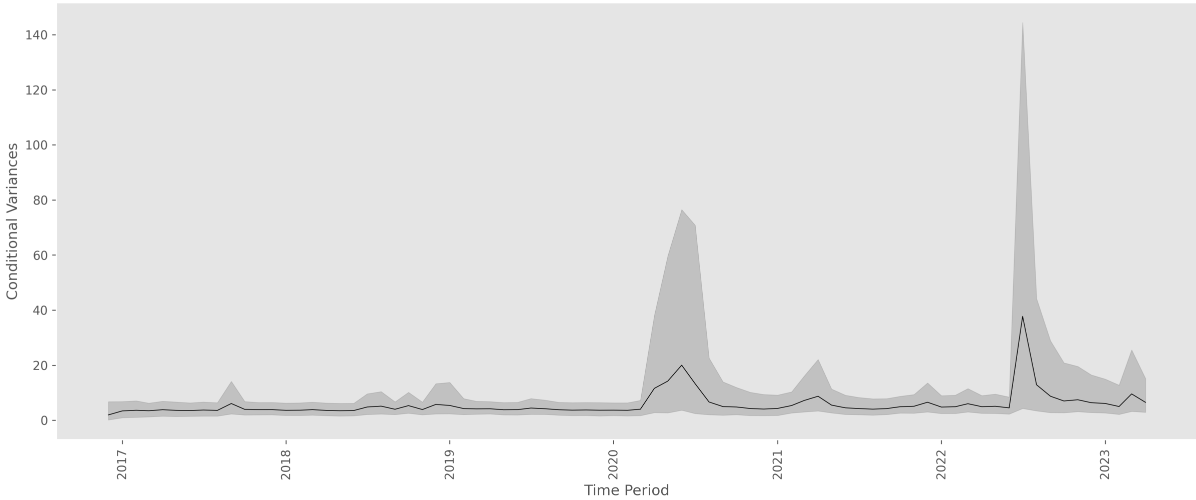
The first period occurred in May 2018, corresponding to the truck drivers' strike in Brazil. Between May 20 and 30, 2018, numerous strikes by independent truck drivers took place across the country, protesting high diesel prices. During this short period, fuel shortages occurred at several gas stations in multiple states, causing a temporary increase in the prices of various fuels (Central Bank of Brazil, 2018).

The second period occurred in 2020, corresponding to the COVID-19 health crisis, which led to shortages in the production of various products, including fuels, due to sanitary restrictions. Finally, the third period, in February 2022, corresponds to the Russian invasion of Ukraine. With the invasion, Russia faced economic sanctions that prevented it from selling large quantities of oil. As a result of the reduced Russian capacity to sell oil and its derivatives, global oil prices tended to rise.

Figure 2.8 shows the conditional variance of gasoline during the IPP. It can be observed that there was little variance, except for the 2020 period at the onset of the COVID-19 pandemic and May 2022. In early 2022, the Russian invasion of Ukraine occurred, and Russia faced economic sanctions that impacted oil prices (Miller; Kumleben; Nowak, 2024).

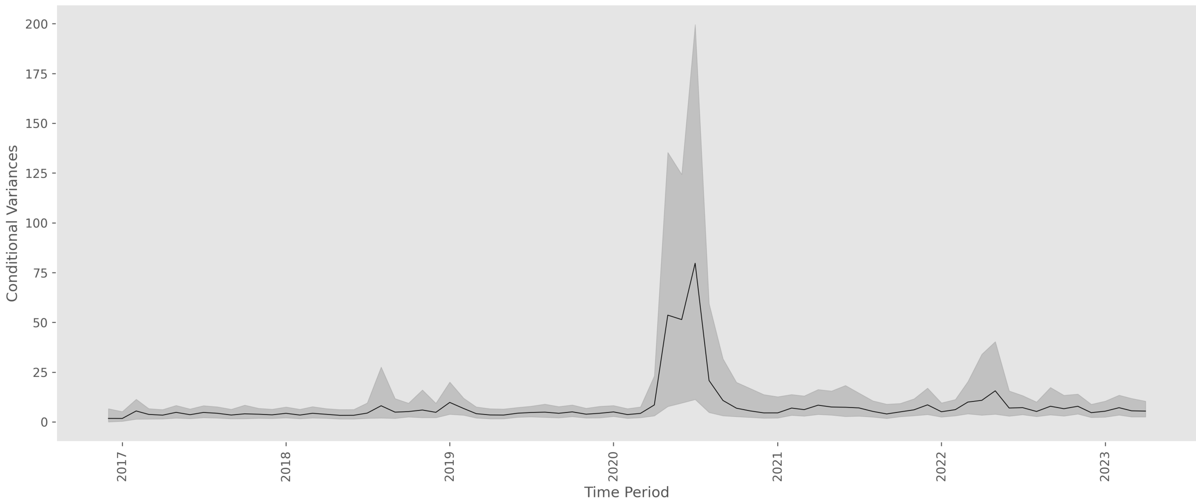
Regarding diesel, the conditional variance of diesel/oil is shown in Figure 2.9, where three variance peaks are observed. The first peak corresponds to the truck drivers' strike in May/June 2018, which caused instability in the supply of goods, services, and products in the country. The second peak occurred in 2020, reflecting the effects of the COVID-19 health crisis, which led to product shortages (including fuels) due to sanitary restrictions imposed by the pandemic. Finally, the third peak is associated with the Russian invasion of Ukraine, when Russia was sanctioned by Western countries, limiting its ability to trade its products (mainly oil), which generated a shortage of oil and derivatives and caused international prices to rise.

Figure 2.8 – Conditional variance of gasoline during the IPP



Source: Based on the research results. The light gray area represents the 90% credible interval, and the black line represents the correlation.

Figure 2.9 – Conditional variance of diesel during the IPP



Source: Based on the research results. The light gray area represents the 90% credible interval, and the black line represents the correlation.

2.6 Conclusions

Using data from the ANP for gasoline and diesel, along with international oil prices from the FED, this study sought to understand how volatility is transmitted from international oil prices to domestic fuel prices (gasoline and diesel) using a multivariate GARCH model with a BEKK specification. The results indicate that gasoline is the fuel most susceptible to oil price volatility, whereas diesel prices are more stable in the face of oil price fluctuations.

Considering the periods analyzed in this study, an interesting result was found for the IPP period. During this period, the Brazilian government adopted a fuel pricing policy using international oil prices as a benchmark, which led to a reduction in fuel volatility, particularly for gasoline.

During the IPP period, a sequence of adjustments in gasoline and diesel prices occurred in response to changes in oil prices. As a result of these constant adjustments, volatility was lower for both fuels; however, baseline volatility was higher during the IPP compared to the full sample period. In contrast, during the full sample period, when price freezes and non-transmissions of international oil prices occurred, particularly for diesel, greater volatility was observed. This is because when prices were eventually adjusted, larger increases were required to align with international oil prices. These delays in price adjustments caused substantial variations in domestic fuel prices and an increase in fuel price volatility.

It was also observed, during this period, that the conditional correlation between oil and fuels was entirely positive, particularly for gasoline. This indicates that gasoline prices followed oil prices during this period, in line with the free-market pricing policy. Such a correlation was expected, given the pricing policy established under the IPP.

Finally, it can be stated that the IPP reduced volatility amplitude for both fuels, although it also generated a constant flow (an increase in baseline volatility) of this volatility. This is due to frequent price adjustments during the period. Thus, it can be observed that in the full sample period, volatility was higher, whereas baseline volatility was lower (due to irregular price adjustments). During the IPP, frequent adjustments led to reduced volatility, despite a higher baseline volatility.

Therefore, considering the differences in the estimations for gasoline and diesel across the two periods analyzed, several possible reasons can be highlighted: (1) the Brazilian government has a long history of intervening in fuel prices (gasoline and diesel) to control prices and prevent inflationary pressures in the country (Moraes; Bacchi, 2014; Almeida; Oliveira; Losekann, 2015; National Agency of Petroleum, Natural Gas and Biofuels, 2018; Official Gazette of the Union, 2021; Official Gazette of the Union, 2022); (2) Brazil has a long-standing program supporting the production of ethanol and biofuels (biodiesel) to avoid or mitigate price fluctuations in diesel and gasoline stemming from oil prices. Since 2000, Brazil has systematically increased the ethanol content in gasoline to stabilize gasoline prices (Civil House, 2015; National Association of Automobile Manufacturers, 2024b; Ministry of Mines and Energy, 2025b); (3) the country needs to import foreign oil to refine domestic crude through blending due to the characteristics of the Brazilian refining system and the qualities of the oil extracted domestically (Delgado; Gauto, 2021).

Moreover, it is recommended that the Brazilian government adopt a more transparent pricing policy for the country's consumers, preferably managed and presented by the ANP. Such a pricing policy needs to be transparent about future price adjustments tied to oil prices, explaining how these increases will occur and their impact on consumers. One possible solution would be to restructure the IPP to enable healthy monitoring of fuel prices relative to oil, while informing consumers of potential price increases and decreases. Additionally, this approach would demonstrate proper management of Petrobras, preventing strain on the company's cash flow due to the absence of price adjustments.

Regarding the limitations, this study has the following: the first limitation is the use of only a single multivariate GARCH specification (BEKK) to analyze the volatility between international oil prices and domestic fuel prices (gasoline and diesel); the second limitation is the use of only one Markov chain in the model, due to the high computational cost of employing two or more chains and the nature of the data, which complicates the convergence of multiple Markov chains.

To further enrich the discussion on gasoline and diesel volatility in the Brazilian context and its relationship with oil prices, future research could consider alternative specifications (such as DCC and CCC) and compare oil volatility spillover across them. Moreover, it is possible to add an analysis of the exchange rate and the price of anhydrous ethanol, and explore how these variables respond to oil price shocks.

3 Essay 3 - Fuel Price Forecasting in Brazil under Cross-State Dependence: An EMAR Framework

Abstract

Fuel price forecasting has been extensively studied in academic literature. Most existing research in Brazil focuses on specific fuels, regional markets, or wholesale prices, with retail price forecasting receiving comparatively less attention. This paper introduces the Envelope Matrix Autoregressive Model (EMAR) as an innovative method for predicting retail fuel prices in Brazil. EMAR enables analysis across multiple dimensions, considering various fuel types and regions. By employing envelope dimension reduction, EMAR efficiently distinguishes relevant information from irrelevant data. Forecasts produced by EMAR are evaluated against those generated by the Vector Autoregressive Model (VAR) and the Matrix Autoregressive Model (MAR) using the MSE criteria. The findings indicate similar results across the MAR and EMAR models.

Resumo

A previsão dos preços dos combustíveis tem sido amplamente estudada na literatura acadêmica. A maior parte das pesquisas realizadas no Brasil concentra-se em combustíveis específicos, mercados regionais ou preços no atacado, enquanto a previsão dos preços ao consumidor tem recebido relativamente menos atenção. Este estudo apresenta o Envelope Matrix Autoregressive Model (EMAR) como um método inovador para prever os preços dos combustíveis ao consumidor no Brasil. O EMAR permite analisar múltiplas dimensões simultaneamente, considerando diferentes tipos de combustíveis e regiões. Por meio da redução de dimensão via envelope, o modelo separa eficientemente as informações relevantes das irrelevantes. As previsões geradas pelo EMAR são comparadas às obtidas pelos modelos Vector Autoregressive (VAR) e Matrix Autoregressive (MAR), utilizando o Erro Quadrático Médio (MSE) como critério de avaliação. Os resultados indicam desempenho preditivo semelhante entre os modelos MAR e EMAR.

3.1 Introduction

Brazil is a country of continental proportions that has historically relied on road transportation to connect its different parts and deliver goods, services, and people. According to a study by the National Association of Railway Transporters (2024) (ANTF) from the acronym in Portuguese, the transportation matrix in Brazil is composed of 70% of highways and roads, as previously reported by Barbosa et al. (2023). Therefore, even small changes in fuel prices can significantly affect Brazilian families' budgets, because small increases in fuel prices over time lead to higher freight prices for goods and services.

Finally, it can lead to social unrest, as occurred during the truck drivers' strikes in May 2018.

Due to the volatility of fuel prices, there is a need for greater price predictability. Thus, the National Agency of Petroleum, Natural Gas and Biofuels (2018) (ANP), from the acronym in Portuguese, developed a technical note and established that contracts between producers and distributors of oil and its derivatives must include information about the future prices of the products involved. This requirement increases transparency in contractual negotiations and provides greater predictability in oil contracts, contributing to the stability of oil derivative prices.

Furthermore, the predictability present in these contracts can also be extended to final consumers through fuel price forecasting. By using forecasting models, it is possible to estimate future fuel prices and provide information about potential price movements. This information can help retail consumers in Brazil anticipate price variations and make better-informed decisions.

Historically, forecasting fuel prices has been a recurrent theme in the international literature. Fuel forecasting is divided into two categories: univariate and multivariate forecasting (The appendix table contains the main references for univariate and multivariate forecasting). The univariate forecast uses its own fuel time series to predict future prices; for example, it uses only the gasoline monthly time series price to forecast its own future. Moreover, there are several different types of methodologies to forecast the future fuel prices (e.g., ARIMA, Exponential Smoothing, ARMA, and Neural Networks). These methods have been applied in the following studies Ye, Zyren and Shore (2005), Asuamah and Ohene-Manu (2015), Baumeister, Kilian and Lee (2017), Safari and Davallou (2018), Silva (2019), Klein et al. (2020), Queiroz et al. (2022), Pei et al. (2023).

On the other hand, the multivariate approach predicts future fuel prices by considering variables such as macroeconomic factors, behavioral consumer consumption, and other types. For example, the following studies predicted fuel prices based on their own fuel prices and other variables Baghestani (2015), Su et al. (2019), Berrisch and Ziel (2022), Pincheira-Brown et al. (2022).

Considering the univariate approach, it is important to highlight the following study:García-Martos, Rodríguez and Sánchez (2013) predicted natural gas and coal prices in the Spanish market using ARMA (Autoregressive Moving-Average) and VARMA (Vector Autoregressive Moving-Average) models on daily data from 2009 to 2011. Additionally, Baumeister, Kilian and Lee (2017) predicted gasoline prices at the pump in the U.S. market using several methods.

Meanwhile, in a multivariate approach, it's worth citing Stefanakos and Schinas (2014), who used a VARMA model to predict future bunker prices at some harbors worldwide. Using this method allows us to predict fuel prices by considering prices at different ports worldwide. However, the VARMA model does not consider dimensionality reduction as in this study. Thus, the main contribution of this paper is to link fuel prices across locations and types to forecast them.

In Brazil, studies that predicted fuel prices have focused on specific fuels and states, or on the national context. For example, Silva (2019) predicted ethanol prices using a univariate method for São Paulo state. On the other hand, Santos et al. (2021) and Queiroz et al. (2022) predicted the ethanol and diesel prices respectively in a national context. Lastly, Klein et al. (2020) used an artificial neural network to forecast the prices of gasoline, diesel, LPG, and LNG (Liquefied Natural Gas), but forecasting each fuel separately.

This paper employs a multivariate methodology to extract the informational content embedded in prices across various Brazilian states and fuel types, with the objective of forecasting retail prices in Brazil. In this regard, the article contributes to the literature by overcoming the limitations of previous studies that made forecasts only for a single type of fuel or for just one region or state in Brazil.

To achieve this objective, the Envelope Matrix Autoregressive (EMAR) model, which was developed by Samadi and Alwis (2025), is employed. The Matrix Autoregressive (MAR) model encompasses a class of models that represent time series data as intersections of two classifications. For instance, consider different fuel prices—such as diesel, gasoline, and ethanol—across various locations like the states of São Paulo, Minas Gerais, and Rio de Janeiro in Brazil. These intersections, where fuel types form the rows and states the columns, naturally create matrix-valued time series. Chen, Xiao and Yang (2021) introduced a Matrix-Valued Autoregressive specification (MAR), and the EMAR model expands on this by employing an envelope approach that converts matrix variables into vectors. Utilizing subspace concepts, the EMAR model extracts relevant information from these vectors and reduces the paramameter's dimensionality efficiently.

So, the methodology can extract the desired information by reducing the subspaces used through the vectors. For example, with four types of fuel, the methodology can reduce all four fuel prices into two common vectors; each vector carries helpful information for predicting future fuel prices. The idea of vector can be applied iteratively across fuel price states as well. Furthermore, this reduction in the number of common vectors from a large number of variables allows the model to estimate with fewer parameters than other models, e.g., VAR and MAR.

To verify these qualities (fewer parameters and the capacity to extract useful information) relative to other models, the EMAR was used to forecast fuel prices in Brazil. The database considered four types of fuel (gasoline, ethanol, diesel, and LPG (Liquefied Petroleum Gas) and twenty-five states in Brazil (São Paulo (SP), Rio de Janeiro (RJ), Minas Gerais (MG), Espírito Santo (ES), Paraná (PR), Santa Catarina (SC), Rio Grande do Sul (RS), Mato Grosso (MT), Mato Grosso do Sul (MS), Goiás (GO), Dsitrito Federal (DF), Amazonas (AM), Roraima (RR), Acre (AC), Tocantins (TO), Rondônia (RO), Maranhão (MA), Piauí (PI), Ceará (CE), Rio Grande do Norte (RN), Bahia (BA), Paraíba (PB), Sergipe (SE), Pernambuco (PE), and Alagoas (AL)) from July 2001 to September 2025.

The results of this study indicate that the EMAR model was slightly superior to the other methods presented, VAR (Vector Autoregressive) and MAR (Matrix Autoregressive), for forecasting. Although the MSE for the EMAR method was the lowest, the difference between MAR and EMAR was too small to be significant.

Therefore, this study contributed to the Brazilian literature in several ways. First, it introduced a new forecasting approach and applied it to fuel prices, a novelty in the international and national scenario. Second, the analysis of forecasting using information from multiple states and several fuel prices is unique to Brazil. Third, the two-way, in which both states and fuels can help forecast their own prices, is relevant in the Brazilian context given the extensive use of highways to transport goods, services, and people. In summary, this study sought to bring a new perspective on fuel forecasting and its dynamics.

The remainder of the paper is structured as follows. Section 2 presents the literature review on fuel forecasting; Section 3 presents the Empirical Strategy; Section 4 describes the data and summary statistics; Section 5 presents the main results; and finally, Section 6 presents the conclusions.

3.2 Literature Review on fuel forecasting

The literature on forecasting fuel in Brazil and worldwide can be divided into two main groups: univariate and multivariate approaches. The univariate approach typically uses its own fuel time series to predict future prices. Meanwhile, the multivariate approach uses additional variables (macroeconomic factors, consumer behavior, and others) to help forecast future fuel prices. In Table 1, the three main branches this study focused on are described: the univariate and multivariate approaches and fuel forecasting for Brazil. Furthermore, Table 1 presents a description of the main analyses for each study.

Table 3.1 – Literature Review

Study	**Time**	**Variables**	**Countries**	**Methodologies**	**Main results**
Forecasting in the oil and fuels sector around the world					
Ye, Zyren & Shore (2005)	January/1992–April/2003 (Monthly)	Oil WTI, Industrial Activity of the OECD.	United States	Naive, RSTK, MALT model.	The RSTK model provided best forecasts of WTI prices at 1-, 2-, 3-, and 6-month horizons.
García-Martos, Rodríguez & Sánchez (2013)	06/March/2009–29/April/2011 (Daily)	Oil, natural gas and coal.	Spain	ARIMA and VARMA.	For forecast horizons up to 15 days, the VARMA model outperformed ARIMA on average in predicting prices.
Asumah & Ohene-Manu (2015)	January/2000–December/2011 (Monthly)	Premium fuel (High Octane rating).	Ghana	ARIMA (1,1,0), ARIMA (0,1,1) and ARIMA (1,1,1).	The ARIMA (1,1,1) model achieved the best forecasts according to the RMSE. However, the model tended to underestimate prices compared to actual values.
Baumeister, Kilian & Lee (2017)	January/1992–March/2014 (Monthly)	Gasoline	United States	ARMA (1,1); IMA (1); ARIMA (1,1,1); AR (12); Bayesian AR (12); AR (AIC); Bayesian AR (AIC); Exponential Smoothing, UC-SV and no-change forecast.	The UC-SV model outperformed the comparison model (no-change forecast) at forecast horizons of 3, 6, 9, 12, 15, 18, 21, and 24 months.

Continues on next page

Study	Time	Variables	Countries	Methodologies	Main results
Safari & Davallou (2018)	January/2003–September/2016 (Monthly)	Oil	Countries members of the OPEC (Organization of the Petroleum Exporting Countries).	ARIMA, Neural Networks, and Exponential Smoothing, with the use of hybrid models based on these three methods.	The best forecasting method was a hybrid model (PHM), according to the RMSE. The model incorporates time-varying weights and was applied for a one-month-ahead forecast only.
Pei, Huang, Shen & Wang (2023)	7/january/1997–13/September/2022	Natural gas price	United States	Temporal Convolutional Network, one-dimensional convolutional neural network, gate recurrent unit, long short-term memory.	The proposed model (Temporal convolutional network) was the best-performing model, with Mean Absolute Percentage Error (4.96%), Mean Absolute Error (0.21), 0.68, and Root Mean Squared Error (0.687).
Gosh & De (2024)	01/January/2020–27/January/2023 (Daily)	Very Low Sulphur fuel Oil	Amsterdam, Antwerp, Gothenburg, Hamburg, and Rotterdam Harbors	Least Square Boosting (LS-Boosting) and Facebook Prophet	The results indicated that LS-Boosting was the best predictive model during calm moments, but during turbulent moments (the Russian-Ukrainian War), adding information from Facebook Prophet can be useful for forecasting maritime fuel prices.
Forecasting using multivariate models					

Continues on next page

Study	Time	Variables	Countries	Methodologies	Main results
Stefanakos & Schinas (2014)	06/July/2012–28/December2012 (Daily)	Marine Bunker Oil	Rotterdam, Fujairah, Singapore and, Houston Harbors	Vector AutoRegressive Moving Average	The model forecasted the original series very well and showed that using octa-variate combinations is the best for forecasting bunker prices across four harbors. Thus, the VARMA methodology was efficient at forecasting maritime prices.
Baghestani (2015)	January/2003–June/2014 (Monthly)	gasoline price and Harbor conventional gasoline regular spot price	United States	Vector AutoRegressive Model and Moving Average	The results indicated that the VAR model was superior to the MA model for forecasting gasoline prices. Furthermore, the VAR model had more directional accuracy than the MA.

Continues on next page

Study	Time	Variables	Countries	Methodologies	Main results
Su, Zhang, Zhu & Zha (2019)	January/2001–December/2017 (Daily, Weekly and Monthly)	Natural gas spot price, Heating Oil Prices (HO), WTI oil Prices (WTI), Baker Hughes US Natural Gas Rotary Rig Count (NGRRC), Total US Natural Gas Marketed Production (NGMP), Total US Natural Gas Consumption (NGC), Total US Natural Gas Underground Storage Capacity (NGUSC), and Total US Natural Gas Imports (NGI)	International	Linear Regression, Linear support vector machine (SVM), Quadratic SVM, Cubic SVM, LSBoosting.	To forecast natural gas, the best model was LSBoosting at daily, weekly, and monthly frequencies, as it had the lowest MAE, MSE, and RMSE.
Ou, Lin, Xu, Hao, Li, Gao, He, Przesmitzki & Bouchard (2020)	January/2011–October/2018	retailed gasoline price and Brent crude oil	China	the asymmetric error correction model (AEM), the threshold error correction model (TEM) and the threshold and asymmetric error correction model (TAEM).	According to the forecast of retail gasoline prices, the TAEM model was the best at predicting future gasoline prices in China, as it included more exogenous variables than the other two models. Given this, the adjusted R-square was the highest for the TAEM model (0.78) compared with the other models.

Continues on next page

Study	Time	Variables	Countries	Methodologies	Main results
Escribano & Wang (2021)	08/January/2010–11/November/2019 (Weekly)	Brent oil, pre-tax retail 95-octane gasoline, spanish retail gasoline.	Spain	STAR, Random Forest, Mixed Random Forest.	The best predictive model, in terms of goodness of fit, was a Mixed Random Forest, both in-sample and out-of-sample. Once the model was the best in all weeks ahead forecasting (from 1 to 12). The novelty of this model was to blend cointegration with Random Forest.
Berrisch & Ziel (2021)	January/2011–March/2020 (Daily and Monthly)	Natural gas, oil ,and coal prices	Europe	Autoregressive structure with Exponential smoothing process with autoregressive moving average (ARMA) errors. Furthermore, the volatility process was modeled using Threshold GARCH. VAR, Random Forest and, ARIMA.	The results indicated that the proposed model was the best for forecasting future natural gas prices in Europe, both day-ahead and month-ahead, with the lowest MAE and RMSE values.
Pincheira-Brown, Bentancor, Hardy & Jarsun	October/1999–December/2019 (Monthly)	Brent and WTI oil, propane, regular gasoline, heating oil, copper prices; London Metal Exchange Index, Chilean Exchange Rate and Chilean effective Exchange Rate.	Chile	AutoRegressive Model, Random Walk, no-change model, spread model based on crude oil futures prices and spread model based on return of non-oil commodities.	The spread model was the best predictive model for forecasting oil derivatives in Chile, as it accounted for the exchange rate.

Continues on next page

Study	Time	Variables	Countries	Methodologies	Main results
Guo, Huang, Li & Li (2023)	26/March/2018–28/February/2023 (Daily)	Crude oil price and over 12 other variables (monetary, future market, stock market and economic policy uncertainty).	China	Machine learning methods: long short-term memory, gate recurrent unit,recurrent neural network, multilayer perceptron, convolutional neural network, backpropagation and support vector regression.	The following results, the best predictive model was gate recurrent unit. This model has the lowest value across the evaluation methods: MAE, MSE, RMSE, and MAPE.
Forecasting for fuels in Brazil					
Silva (2009)	November/2002–December/2017 (Monthly)	Hydrous Ethanol	São Paulo state/Brazil	ARMA (1,1); ARMA (1,2); ARMA (2,1); ARMA (2,2); ARMA (1,3) and ARMA (3,1).	The ARMA(1,1) model achieved the best forecasting performance, with the lowest RMSE.
Klein, Silva, Veiga, & Mariani (2020)	May/2004–February/2015 (Weekly)	gasoline, diesel, ethanol, liquefied petroleum gas and, liquefied natural gas.	Brazil.	Artificial Neural Network: feedforward and Elman architectures and wavelets (Rickert and Morlet).	The results showed that there was no best model for predicting across all fuels. However, the study presented a comprehensive portfolio of forecasting methodologies for fuels in Brazil. Therefore, this was the study's novelty.
Santos, Barboza, Veiga & Silva (2021)	25/January/2010–11/December/2020 (Daily)	Hydrous ethanol	Brazil	Machine learning methods: Long-Short term memory, Random Forest, Support Vector Machine with Linear, Radial Basis Function.	The results indicated that the long-short-term memory was the best predictive model for forecasting ethanol prices in Brazil, with horizons of 63, 126, and 252 business days.

Continues on next page

Study	Time	Variables	Countries	Methodologies	Main results
Queiroz, Alves, Oliveira & Aguiar (2022)	December/2012–May/2020 (Monthly)	Diesel S10 and S500	Brazil	Fuzzy models.	Among the proposed models, VS-ePL-KRLS achieved the lowest RMSE values and was the best predictive method for diesel prices in Brazil.

Source: Elaborated by the author.

Based on Table 1, the existing literature on fuel price forecasting follows two main approaches. The first uses only historical fuel prices (univariate models) to predict future prices. The second incorporates additional explanatory variables, such as crude oil prices and macroeconomic indicators, to improve forecasting performance.

Considering the univariate forecast for fuel prices, the study by García-Martos, Rodríguez and Sánchez (2013) predicted natural gas and coal prices in the Spanish market using ARMA and VARMA models on daily data from 2009 to 2011. Additionally, Baumeister, Kilian and Lee (2017) predicted gasoline prices at the pump in the U.S. market using several methods. In multivariate approaches, studies typically use macroeconomic or consumer behavior variables to predict future fuel prices, as in Su et al. (2019), Pincheira-Brown et al. (2022), Guo et al. (2023).

Meanwhile, the main goal of this study is to forecast future fuel prices, taking into account the potential contributions of other fuels. Moreover, the methodology employed in this study allows us to analyze the influence of fuel prices in other states to predict the price in a specific state. For instance, considering a panel containing four fuels (diesel, ethanol, gasoline, and Liquefied Petroleum Gas) and five states (São Paulo, Rio de Janeiro, Minas Gerais, Espírito Santo, and Paraná) in order to forecast the gasoline price in Rio de Janeiro state, the other fuels will provide useful information, as well as the states.

The methodology used in this study, called EMAR (Samadi; Alwis, 2025), reduces dimensionality across different fuel types and states. This dimensionality reduction allows us to transform the data into common factors for fuels and states and use these factors to forecast each fuel price in each state. Therefore, this paper introduces a novel approach to forecasting fuel prices.

Even so, Stefanakos and Schinas (2014) used a VARMA model to predict future bunker prices at some harbors worldwide. Using this method allows us to predict fuel prices, considering the price of this fuel at different ports worldwide. However, the VARMA model does not consider a dimensionality reduction as in this study. Thus, the main contribution of this paper is to link fuel prices across locations and types to forecast them.

Finally, this study offers a new perspective on fuel forecasting. Being different from other approaches across several fuel types and states used to generate the forecast. Therefore, leads to a deeper understanding of fuel behavior in Brazil.

3.3 Empirical Strategy

This section presents the models used to forecast domestic fuel prices. In this study, three models were employed: the Envelope Matrix Autoregressive Model (EMAR), the Matrix Autoregressive Model (MAR), and the Vector Autoregressive Model (VAR).

3.3.1 Matrix Autoregressive Models

Chen, Xiao and Yang (2021) developed a bilinear version of the Matrix Autoregressive Model (MAR), which allows for combining different types of fuel for different Brazilian states. This intersection between different types of fuel and Brazilian states can be interpreted as a time-series matrix, which can be represented by a Vector Autoregressive model (VAR)[1], as presented in equation 1.

$$VEC(\mathbf{X}_t) = \boldsymbol{\phi} VEC(\mathbf{X}_t) + VEC(\mathbf{E}_t) \tag{3.1}$$

Here, the VAR model is a vectorized version, and according to Chen, Xiao and Yang (2021), it fails to recognize the strong connections present in the columns and rows. As a result, each column represents a state, and each row represents a type of fuel. Furthermore, the parameter phi is very difficult to interpret due to the large dimensions of the coefficient matrix phi ($mn \times mn$), where m is the number of rows and n is the number of columns and both in the same t.

To address this problem, Chen, Xiao and Yang (2021) brought up the MAR model of order p (MAR(p), which can handle it, and being defined as a bilinear process in which a matrix-valued time series $\{\mathbf{X}_t\}_{t=1}^{N}$ of dimension $m \times n$ is described by the equation below.

$$\mathbf{X}_t = \boldsymbol{\mu} + A_1 \mathbf{X}_{t-1} B_1 + \cdots + A_p \mathbf{X}_{t-p} B_p + \mathbf{E}_t \tag{3.2}$$

Where $\boldsymbol{\mu} \in \mathbb{R}^{m \times n}$ is the intercept matrix, $\mathbf{A_i} \in \mathbb{R}^{m \times n}$ and $\mathbf{B_i} \in \mathbb{R}^{m \times n}$ with $i = 1, \ldots, p$, are the row and columns coefficient matrices, respectively. In this study, the row will represent the types of fuel and the columns the Brazilian states, and $\mathbf{E_t} \in \mathbb{R}^{m \times n}$ is a matrix representing a white noise process Cov(vec($\mathbf{E}_t$)) $= \boldsymbol{\Sigma}$, i.e., vec($\mathbf{E}_t$) $\sim WN(\mathbf{0}, \boldsymbol{\Sigma})$. For $t \neq s$, $\mathbf{E_t}$ and $\mathbf{E_s}$ are uncorrelated, however may have concurrent correlations. Moreover, for each, t, $\mathbf{E_t}$ is independent of lagged matrices $\mathbf{X_t - 1}$, $\mathbf{X_t - 2}$, etc. The MAR(p) process can be rewritten as a MAR(1) as

$$\mathbf{X}_t = \boldsymbol{\mu} + \sum_{i=1}^{p} A_i \mathbf{X}_{t-i} B_i + \mathbf{E}_t \tag{3.3}$$

Where $\mathbf{X}_t = blockdiag(\mathbf{X}_t \ \cdots \ \mathbf{X}_{t-p+1}) \in \mathbb{R}^{m \times np}$, $\quad A = (A_1 \ \cdots \ A_p) \in \mathbb{R}^{m \times mp}$, $\quad B = (B_1 \ \cdots \ B_p) \in \mathbb{R}^{n \times np}$.

The structure in the MAR model can be explained as follows: matrix A reflects row-wise interactions (in this case, the types of fuels), and matrix B reflects column-wise interactions (in this case, the Brazilian states). To improve understanding of the model, note the special cases.

[1] The VAR model was proposed by Sims (1980)

First, let us assume A=I, as the bilinear form will be

$$X_t = X_{t-1}B^{'} + E_t$$

As the conditional expectation of the first column of Xt is given by

$$\begin{pmatrix} \text{gas} \\ \text{die} \\ \text{eth} \\ \text{lpg} \end{pmatrix}_t^{SP} = b_{11} \begin{pmatrix} \text{gas} \\ \text{die} \\ \text{eth} \\ \text{lpg} \end{pmatrix}_{t-1}^{SP} + b_{12} \begin{pmatrix} \text{gas} \\ \text{die} \\ \text{eth} \\ \text{lpg} \end{pmatrix}_{t-1}^{RJ} + \cdots + b_{1n} \begin{pmatrix} \text{gas} \\ \text{die} \\ \text{eth} \\ \text{lpg} \end{pmatrix}_{t-1}^{PI}$$

where gas is gasoline, die is diesel, eth is ethanol, and lpg is liquefied petroleum gas, using 25 Brazilian states (excluding Pará and Amapá). Besides that, the matrix, which means at time t, the conditional expectations of a fuel price given one state, is a linear combination of the same fuel from all states at t-1, and this linear combination is the same for different fuels. Therefore, this model (only B) captures the column-wise interactions among the Brazilian states. However, there are no interactions among the fuels.

On the other hand, in case B=I, the interactions among the fuel will be found, and the model becomes

$$X_t = AX_{t-1} + E_t$$

In this case, each column of X_t follows

$$X_{t,j} = AX_{t_1,j} + E_{t,j}, j = 1, \cdots, n.$$

The matrix form will be given by

$$\begin{pmatrix} \text{gas} \\ \text{die} \\ \text{eth} \\ \text{lpg} \end{pmatrix}_t^{SP} = A \begin{pmatrix} \text{gas} \\ \text{die} \\ \text{eth} \\ \text{lpg} \end{pmatrix}_{t-1}^{SP} + \begin{pmatrix} \text{Egas} \\ \text{Edie} \\ \text{Eeth} \\ \text{Elpg} \end{pmatrix}_t^{SP}$$

and

$$\begin{pmatrix} \text{gas} \\ \text{die} \\ \text{eth} \\ \text{lpg} \end{pmatrix}_t^{PI} = A \begin{pmatrix} \text{gas} \\ \text{die} \\ \text{eth} \\ \text{lpg} \end{pmatrix}_{t-1}^{PI} + \begin{pmatrix} \text{Egas} \\ \text{Edie} \\ \text{Eeth} \\ \text{Elpg} \end{pmatrix}_t^{PI}$$

In other words, for each state, its type of fuel follows a VAR (1) model (of its own past) of dimension m; the same is valid for other states. Furthermore, both models are too restrictive, because it is very hard to find a reason for São Paulo's (SP) types of fuel to follow the same model as Piauí's (PI), and affirm that there is no interaction between these states. Therefore, it is assumed that there is an interaction among the fuels and states.

The matrices A and B are not identified, since they can be multiplied by the same nonzero constant and the model remains the same. In order to solve this problem, Chen, Xiao and Yang (2021) proposed to normalize matrix A by ensuring that its Frobenius norm is equal to 1. Furthermore, it was assumed that the structured covariance matrix is as follows

$$\sum_{\mathbf{e}} = Cov(Vec(E_t)) = \sum_{c} \otimes \sum_{r} \tag{3.4}$$

where $\sum_{\mathbf{c}} \in \mathbb{R}^{n \times n}$ and $\sum_{\mathbf{r}} \in \mathbb{R}^{m \times m}$ are positive defined column and row covariance matrices. Hence $\sum_c$ and $\sum_r$ are exclusively defined up to a proportionality constant ($\sum_e = \alpha \sum_c \otimes \alpha^{-1} \sum_r, \alpha \neq 0$), this will make $\sum_e$ non-identifiable, so it was imposed the constraints $\|\sum_r\|_F = 1$ and positivity in its first row and column. After all, the total number of parameters (NOP) in the MAR model is $nm + p(n^2 + m^2) + n(n+1)/2 + m(m+1)/2 - 2$. According to this formula, it can be seen that the size increases significantly in high dimensions and as more lags are added, requiring a high level of efficiency and parsimony to avoid the curse of dimensionality.

The stability of the MAR(p) model (2) is guaranteed by checking the characteristic roots of the corresponding VAR(p), which is obtained by vectorizing both sides of equation (2) as follows

$$\mathbf{vec}(\mathbf{X_t}) = \mathbf{vec}(\mu) + \boldsymbol{\Theta}\mathbf{vec}(\mathbf{V_{t-1}}) + \mathbf{vec}(\mathbf{E_t}), \tag{3.5}$$

where $V_{t-1} = (X_t, \cdots, X_{t-p})$, and $\Theta = (\Theta_1, \cdots, \Theta_p) = B * A \in \mathbb{R}^{mn \times mnp}$ with $\Theta_i = (B_i \otimes A_i)$. Let L be the lag operator, and $\Theta(L) = I_{mn} - \Theta_1 L - \cdots - \Theta_p L^p$ is the characteristic polynomial. The MAR(p) process will be stationary if all roots of det $(\Theta(L)) = 0$ have moduli greater than one. For $p = 1$, stationary holds if $\rho(A).\rho(B) < 1$ (Chen; Xiao; Yang, 2021).

3.3.1.1 One-Sided Matrix Autoregressive Models

Based on the previous explanation about the two sides of the MAR matrix, either A or B. Now, it is possible to estimate both matrices separately, as was done by (Chen; Xiao; Yang, 2021) and (Samadi; Alwis, 2025). The reason for this choice is that, in a one-sided MAR model, it is possible to concentrate the estimation on either the rows or the columns. Furthermore, in this study, two one-sided MAR models are used: one in which the estimation is centered on the rows, and another in which it is centered on the columns. The equation below shows the equation for each side of the MAR models.

$$X_t = A_1 X_{t-1} + \cdots + A_p X_{t-p} + E_t = A Z_{t-1} + E_t, \; if \; B_i = I_n \tag{3.6}$$

$$X_t = X_{t-1} B_1^\top + \cdots + X_{t-p} B_p^\top + E_t = V_{t-1} B^\top + E_t, \; if \; A_j = I_m \tag{3.7}$$

Where $Z_{t-1} = (X_{t-1}^\top, \cdots, X_{t-p}^\top)^\top \in \mathbb{R}^{mp \times n}$, and $V_{t-1} = (X_{t-1}, \cdots, X_{t-p}) \in \mathbb{R}^{m \times np}$. In model (6), the columns present in X_t represent each Brazilian state, and follow the same VAR(p) model dimension m. In the same way, in model (7), each row of X_t represents each type of fuel used in this study, and follows the same VAR(p) model of dimension n. Considering the estimation process, in this study, the parameters of the MAR model were estimated via maximum likelihood, assuming that the error terms follow a matrix-normal distribution, i.e, $E_t \sim \mathcal{N}_{m,n}(0, \sum_r \otimes \sum_c)$ (Samadi; Alwis, 2025).

Finally, the use of one-sided estimation for rows or columns facilitates overall estimation. Besides that, it provides a better way to explain the model.

3.3.2 Vectorized Matrix-Variate Time Series

In this study, a transformation VEC(.) was applied, where the intention behind this was to transform a matrix time series X_t into a vector time series $y_t = VEC(X_t) \in \mathbb{R}^q$, where $q = mn$. Therefore, the VAR(p) model of y_t is then given by

$$y_t = \alpha + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + \varepsilon_t = \alpha + \beta X_t + \varepsilon_t, \tag{3.8}$$

Where $\beta \in \mathbb{R}^{q \times qp}$ is the coefficient matrix, $\alpha \in \mathbb{R}^q$ is the intercept, $X_t \in \mathbb{R}^{qp}$ is the lagged vector, and finally, $\varepsilon_t \in \mathbb{R}^q$ is white noise ($\varepsilon_t \sim WN_q(0, \sum_\varepsilon)$).

Equation (8) presents the classical VAR model, in the same way as equation (1), and both equations are very similar to the stacked VAR model (5). However, their coefficients, β_i and Θ_i are different. In contrast to the vectorized MAR model (5), the VAR model (8) lacks a Kronecker structure. Where the maximum likelihood estimation for its parameters are $\alpha = \bar{y}$, $\beta = \sum_{t=i}^N (y_t - \bar{y})^\top X_t (\sum_{t=1}^N X_t X_t^\top)^{-1}$ and $\sum_\varepsilon = \frac{1}{N} \sum_{t=1}^N (y_t - \beta X_t)(y_t - \beta X_t)^\top$.

3.3.3 Envelope-Based Matrix Autoregressive Analysis

The envelope as a method was raised by Cook, Li and Chiaromonte (2010) in 2010. The purpose of this method is for multivariate linear regression, with the main goal of reducing the dimensionality present in the regression. Dimension reduction improves estimation efficiency by focusing only on the essential information contained within minimal subspaces while discarding irrelevant information. This approach enhances both estimation and efficiency by reducing the number of free parameters and utilizing the minimal reducing subspaces, which relate the mean function to the covariance matrix.

Two definitions summarize the idea of an envelope, according to Samadi and Alwis (2025).

Definition 1

A subspace $\mathcal{R} \subseteq \mathbb{R}^q$ is referred to as a reducing subspace of matrix $\mathbf{M} \in \mathbb{R}^{q\times q}$ if $\mathcal{R}$ decomposes $\boldsymbol{M}$ as $M = P_{\mathcal{R}} M P_{\mathcal{R}} + Q_{\mathcal{R}} M Q_{\mathcal{R}}$. In this case, we say that $\mathcal{R}$ reduces $\boldsymbol{M}$.

Definition 2

A subspace $\boldsymbol{M}$-envelope of a set $\mathcal{S}$, denoted by $\mathcal{E}_M(\mathcal{S})$, is defined as the intersection of all reducing subspaces of $\boldsymbol{M}$ that contain $\mathcal{S}$, where $\boldsymbol{M} \in \mathbb{R}^{q\times q}$ and $\mathcal{S} \subseteq span(\boldsymbol{M})$.

This reduction of the subspaces is the cornerstone of envelope development, widely used in functional analysis (Conway, 1990). In this way, according to Definition 2, there is a guarantee of the existence and uniqueness of envelopes, affirming that the intersection of any two reducing spaces of $\mathbf{M}$ is also a reducing space.

3.3.3.1 Envelope Matrix Autoregressive (EMAR) Models

Even though the MAR model (2) is more parsimonious in contrast to the VAR model (9), when this occurs, the linear combination of the rows or columns of $X_t \in \mathbb{R}^{m\times n}$ may have some distributions that remain unchanged with variations in the lagged variables. To address this issue, Samadi and Alwis (2025) proposed utilizing an envelope structure for both rows and columns in the MAR model. This approach guaranteed improved accuracy and reduced NOP (total number of parameters) compared to the standard bilinear MAR model. This new approach was called the envelope MAR (EMAR). The authors established two envelope structures for both the rows and columns of the MAR process, and they assumed the existence of two subspaces $\mathcal{S}_r \subseteq \mathbb{R}^m$ and $\mathcal{S}_c \subseteq \mathbb{R}^n$ satisfying

$$(a)\, Q_{\mathcal{S}_r} X_t \mid \mathcal{V}_{t-1} \sim \mathcal{Q}_{\mathcal{S}_r} X_t \qquad (b)\, \mathrm{Cov}_r\Big(\mathcal{P}_{\mathcal{S}_r} X_t, \mathcal{Q}_{\mathcal{S}_r} X_t \mid \mathcal{V}_{t-1}\Big) = 0 \tag{3.9}$$

$$(a)\, X_t \mathcal{Q}_{\mathcal{S}_\rfloor} \mid \mathcal{V}_{t-1} \sim X_t \mathcal{Q}_{\mathcal{S}_\rfloor} \qquad (b) Cov_c(X_t \mathcal{P}_{\mathcal{S}_\rfloor}, X_t \mathcal{Q}_{\mathcal{S}_\rfloor} \mid \mathcal{V}_{t-1}) = 0, \tag{3.10}$$

Where $Cov_c(X_t) = \mathbb{E}[(X_t - \mathbb{E}(X_t))^\top (X_t - \mathbb{E}(X_t))]$ and $Cov_r(X_t) = \mathbb{E}[(X_t - \mathbb{E}(X_t))(X_t - \mathbb{E}(X_t))^\top]$ and assuming $\mathcal{B}_r = span(A)$ and $\mathcal{B}_c = span(B)$. Therefore, the conditions (9, a) and (10, a) imply that $\mathcal{B}_r \subseteq \mathcal{S}_r$ and $\mathcal{B}_c \subseteq \mathcal{S}_c$, respectively, showing that the marginal distributions of $Q_{\mathcal{S}_r} X_t$ and $Q_{\mathcal{S}_c} X_t$ do not depend on the lagged variables $\mathcal{V}_{t-1}$. Moreover, the conditions (9,b) and (10,b) is true if and only if $\mathcal{S}_c$ and $\mathcal{S}_c$ are reducing spaces of $\sum_r$ and $\sum_c$, in a respective way, where $\sum_r = P_{\mathcal{S}_r} \sum_r P_{\mathcal{S}_r} + Q_{\mathcal{S}_r} \sum_r Q_{\mathcal{S}_r}$ and $\sum_c = P_{\mathcal{S}_c} \sum_c P_{\mathcal{S}_c} + Q_{\mathcal{S}_c} \sum_c Q_{\mathcal{S}_c}$, implying that $P_{\mathcal{S}_r} \sum_r Q\mathcal{S}_r = 0$, and $P_{\mathcal{S}_c} \sum_c Q\mathcal{S}_c = 0$. In additional, the conditions (9,b) and (10,b) imply that $Q_{\mathcal{S}_r} X_t$ and $X_t Q_{\mathcal{S}_c}$ do not depend on the lagged variables $\mathcal{V}_{t-1}$ through a linear association $P_{\mathcal{S}_r} X_t$ and $X_t P_{\mathcal{S}_c}$. Therefore, the relation of dependence present in X_t on its lagged values $\mathcal{V}_{t-1}$ is only in material parts of the autoregression ($P_{\mathcal{S}_r} X_t$ and $X_t P_{\mathcal{S}_c}$), while the immaterial parts ($Q_{\mathcal{S}_r} X_t$ and $X_t Q_{\mathcal{S}_c}$) is not affected by changes in $\mathcal{V}_{t-1}$ and behave as white noise becoming them irrelevant for estimating the coefficient matrices.

About the normality issue, the conditions for (9, b) and (10, b) imply the conditional independence assumption to be ($P_{\mathfrak{f}_r} X_t \perp Q_{\mathcal{S}_r} X_t$ and $X_t P_{\mathcal{S}_c} \perp X_t Q_{\mathcal{S}_c} \mid \mathcal{V}_{t-1}$). These assumptions of conditional independence are stronger than conditions (9, b) and (10, b). Overall, the intersection of all reduction subspaces $\mathcal{S}_j$ of $\sum_j$ also contain $\mathcal{B}_j (j = r, c)$ is named the $\sum_j$ envelope of $\mathcal{B}_j$, and denoted by $\varepsilon_{\sum_j}(\mathcal{B}_j)(j = r, c)$. Therefore, these subspaces always exist, and the efficiency gains using the envelope methodology are described in Samadi and Alwis (2025).

In summary, the idea behind the EMAR model is to concentrate estimation in the subspace containing the main information used to forecast fuel prices in Brazil, in this case. The authors use the idea of subspaces to reduce the number of parameters to be estimated while extracting the most useful information. This new forecasting approach seeks the best possible estimation with fewer parameters.

3.3.3.2 Envelope Formulation for EMAR models

To create the envelope presents in the EMAR models, Samadi and Alwis (2025) used the following envelopes, $\sum_r$ and $\sum_c$, where the authors stated that the dimensions are $\sqcap_\infty = dim(\varepsilon_{\sum_r}(\mathcal{B}_r))$ and $\sqcap_\in = dim(\varepsilon_{\sum_c}(\mathcal{B}_c))$. Given that, it was assumed that (R, R_0) and (C, C_0) are orthogonal matrices such that $R \in \mathbb{R}^{m \times u_1}$ $(u_1 \leq m)$ and $C \in \mathbb{R}^{n \times u_2}$ $(u_2 \leq n)$ forming semi-orthogonal bases for $\varepsilon_{\sum_r}(\mathcal{B}^r)$ and $\varepsilon_{\sum_c}(\mathcal{B}_c)$, respectively. Considering the following spans, $span(A) \subseteq \varepsilon_{\sum_r}(\mathcal{B}_r) = span(R)$, and $span(B) \subseteq \varepsilon_{\sum_c}(\mathcal{B}_c) = span(C)$, coordinate matrices $\eta_1 \in \mathbb{R}^{u_1 \times mp}$ and $\eta_2 \in \mathbb{R}^{u_2 \times np}$ has it existence such that $A = R\eta_1$ and $B = C\eta_2$. Therefore, the EMAR model can be written as

$$\begin{gathered} X_t = \mu + R\eta_1 \mathcal{V}_{t-1} \eta_2^\top C^\top + E_t, \\ \sum_r = R\Omega_r R^\top + R_0 \Omega_{r0} R_0^\top, \sum_c = C\Omega_c C^\top + C_0 \Omega_{c0} C_0^\top, \end{gathered} \tag{3.11}$$

Where $\Omega_j > 0$ and $\Omega_{j0} > 0$ are unknown $(j = r, c)$. As for the matrices R and C, neither of them is identifiable; however, their column spaces, that is, span(R) and span(C), are identifiable (Ding; Cook, 2018).

Overall, the increase in envelope dimensions (u_1 and u_2) due to the dimensions of X_t (m and n) indicates that the reduction in the number of parameters becomes increasingly significant. Finally, the EMAR models reach the best estimation over the standard MAR models in two ways.

Firstly, the EMAR model has a significant reduction in the estimated parameter space. Second, this new approach is distinguished by the use of minimal reducing subspaces that are capable of exploiting the structural connections between coefficients and covariances, starting from the envelope bases R and C. Since it captures information from both the conditional means and the covariance structures, EMAR models are superior to MAR models due to reduced parameter counts and a significant efficiency gain from the structural covariance matrix.

Lastly, the EMAR model was estimated using the two-step iterative procedure developed by Ding and Cook (2018). The authors used maximum likelihood to estimate the EMAR parameters. Moreover, in Samadi and Alwis (2025), the authors assumed that the error matrices are normally distributed, with the possibility of relaxing the normality assumption. For more details, see Samadi and Alwis (2025).

3.3.4 EMAR Model estimation

The EMAR model estimation by Samadi and Billard (2025) followed the approach proposed by Ding and Cook (2018). Thus, the estimation was performed in a two-step iterative procedure to obtain the Maximum Likelihood Estimation (MLEs) of the envelope MAR model parameters, as presented in equation (11). This way, these parameters were estimated under the assumption of normally distributed random error matrices, while relaxing the normality assumption and using the 1D algorithm introduced by Cook and Zhang (2016). So, assuming the normality of E_t Et and if the lagged matrix $\mathcal{V}_{\sqcup-\infty}$ is centered ($\overline{\mathcal{V}}_{\sqcup-\infty} = \prime$), the MLE for u is X. Therefore, the log-likelihood function of the EMAR model with fixed B and $\sum_c$ is given by.

$$\begin{aligned} L(R, \eta_1, \Omega_r, \Omega_{r0}) = & -\frac{Nnm}{2}\log(2\pi) - \frac{Nn}{2}\log|\Sigma_c| - \frac{Nm}{2}\log|\Omega_r| - \frac{Nm}{2}\log|\Omega_{r0}| \\ & - \frac{1}{2}\sum_{t=1}^{N} \operatorname{tr}\Big[\Sigma_c^{-1}(R^\top X_t - R^\top \bar{X} - \eta_1 \mathcal{V}_{\sqcup-\infty} B^\top)^\top \\ & \qquad \times \Omega_r^{-1}(R^\top X_t - R^\top \bar{X} - \eta_1 \mathcal{V}_{\sqcup-\infty} B^\top)\Big] \\ & - \frac{1}{2}\sum_{t=1}^{N} \operatorname{tr}\Big[\Sigma_c^{-1}(X_t - \bar{X})^\top (R_0 \Omega_r R_0^\top)^{-1}(X_t - \bar{X})\Big]. \end{aligned} \tag{3.12}$$

The complete derivation of the log-likelihood function is provided in Samadi and Billard (2025). Thus, the adoption of the two-step iteration procedure drives the MLEs of the unknown matrix parameters R, C, A, B, $\sum_r$, and $\sum_c$ under the EMAR model:

Step 1: Fixing B and $\sum_c$, the remaining parameters are estimated as follows.

$$
\begin{aligned}
\hat{R} &= \arg\min_{U_1^\top U_1 = I_{u_1}} \left(\log\left|U_1^\top \hat{\Gamma}_{res,r|c} U_1\right| + \log\left|U_1^\top \hat{\Gamma}_{0,r|c}^{-1} U_1\right| \right), \\
\hat{\Gamma}_{0,r|c} &= (Nn)^{-1} \sum_{t=1}^{N} (X_t - \bar{X}) \Sigma_c^{-1} (X_t - \bar{X})^\top, \\
B_{r|c} &= \Big(\sum_{t=1}^{N} (X_t - \bar{X}) \Sigma_c^{-1} B \mathcal{V}_{t-1}^\top \Big) \Big(\sum_{t=1}^{N} V_{t-1} B^\top \Sigma_c^{-1} B \mathcal{V}_{t-1}^\top \Big)^{-1}, \\
\hat{\Gamma}_{res,r|c} &= \frac{1}{Nn} \sum_{t=1}^{N} \Big[(X_t - \bar{X} - B_{r|c} \mathcal{V}_{t-1} B^\top) \Sigma_c^{-1} (X_t - \bar{X} - B_{r|c} \mathcal{V}_{t-1} B^\top)^\top \Big].
\end{aligned}
\tag{3.13}
$$

The intermediate estimators for A and Σ_r are derived as

$$
\hat{A}_{r|c} = P_{\hat{R}} B_{r|c}, \quad \hat{\Sigma}_{r|c} = P_{\hat{R}} \hat{\Gamma}_{\text{res},r|c} P_{\hat{R}} + P_{\hat{R}_0} \hat{\Gamma}_{0,r|c} P_{\hat{R}_0},
$$

where $(\hat{R}, \hat{R}_0)$ is an orthogonal matrix.

Step 2: the same way, keeping A and $\sum_r$ fixed, and estimating the remaining parameters as

$$
\begin{aligned}
\hat{C} &= \arg\min_{U_2^\top U_2 = I_{u_2}} \left(\log\left|U_2^\top \hat{\Gamma}_{\text{res},c|r} U_2\right| + \log\left|U_2^\top \hat{\Gamma}_{0,c|r}^{-1} U_2\right| \right), \\
\hat{\Gamma}_{0,c|r} &= \frac{1}{Nm} \sum_{t=1}^{N} (X_t - \bar{X})^\top \, \Sigma_r^{-1} \, (X_t - \bar{X}), \\
B_{c|r} &= \left(\sum_{t=1}^{N} (X_t - \bar{X})^\top \, \Sigma_r^{-1} A \mathcal{V}_{t-1} \right) \left(\sum_{t=1}^{N} \mathcal{V}_{t-1}^\top A^\top \, \Sigma_r^{-1} A \mathcal{V}_{t-1} \right)^{-1}, \\
\hat{\Gamma}_{\text{res},c|r} &= \frac{1}{Nm} \sum_{t=1}^{N} \left(X_t^\top - \bar{X}^\top - B_{c|r} \mathcal{V}_{t-1}^\top A^\top \right) \Sigma_r^{-1} \left(X_t^\top - \bar{X}^\top - B_{c|r} \mathcal{V}_{t-1}^\top A^\top \right)^\top.
\end{aligned}
\tag{3.14}
$$

The intermediate estimators for B and Σ_c are obtained as

$$
\hat{B}_{c|r} = P_{\hat{C}} B_{c|r}, \quad \hat{\Sigma}_{c|r} = P_{\hat{C}} \hat{\Gamma}_{\text{res},c|r} P_{\hat{C}} + P_{\hat{C}_0} \hat{\Gamma}_{0,c|r} P_{\hat{C}_0},
$$

where $(\hat{C}, \hat{C}_0)$ is an orthogonal matrix. For more details, see in Samadi and Alwis (2025).

As seen, the parameters were estimated under the assumption of normally distributed random error matrices, meaning that during estimation, a random component is used to calculate the parameters via MLEs. However, the random component may yield different results on each attempt, so a solution is to replicate the estimation n times and average the results for each model.

Therefore, 5 replications were performed to stabilize the model. This way, after all, the final result was obtained by averaging all these replications, especially for out-of-sample performance. For example, for h=1 (h-steps ahead), the forecast was calculated 5 times, and the average was taken.

3.3.5 Accuracy test

To measure and compare the different models employed in this study, the Mean Squared Error (MSE) was used. The use of MSE to calculate the forecasting error was also adopted by Samadi and Alwis (2025).

MSE was used on the training set from July 2001 to April/2023; this period corresponds to 90% of the sample. A forecasting strategy was adopted using a rolling-window scheme with 90% of the sample (261 observations). The horizon forecasts were generated for h=1, 3, 6, and 12 steps ahead, with the evaluation period (e.g., May 2023 to September 2025, 10% of the sample), and the results were compared across the models. To assess forecasting accuracy using h-step ahead mean squared forecast error ($MSFE_h$), with the metrics for MSE and $MSFE_h$ calculated as follows

MSE was used on the training set from July 2001 to April 2023; this period corresponds to 90% of the sample. A rolling-window forecasting strategy was adopted using 90% of the sample (261 observations). Forecasts were generated for horizons of h=1, 3, 6, and 12 months ahead, and evaluated over the out-of-sample period (May 2023 to September 2025, corresponding to the remaining 10% of the sample). For each forecasting horizon, the forecast error used to compute the h-step-ahead mean squared forecast error ($MSFE_h$) was calculated using only the final observation of the forecast horizon. For example, for h=3, the forecast error corresponds to the third month ahead (e.g., March in a January–March forecast), rather than averaging the errors from the first, second, and third months. The MSE and $MSFE_h$ metrics were calculated as follows.

$$MSE = \frac{1}{N}\sum_{t=1}^{N_0} ||X_t - \hat{X}_t||^2, MSFE_h = \frac{1}{N - N_0 - h + 1}\sum_{t=N_0}^{N-h} ||X_{t+h} - \hat{X}_{t+h}||_f^2 \quad (3.15)$$

where N_0 and N are the start and end of the evaluation sample, respectively. h represents the forecast horizon, $\hat{X}_t$ is the predict value by the model and the X_t the real value.

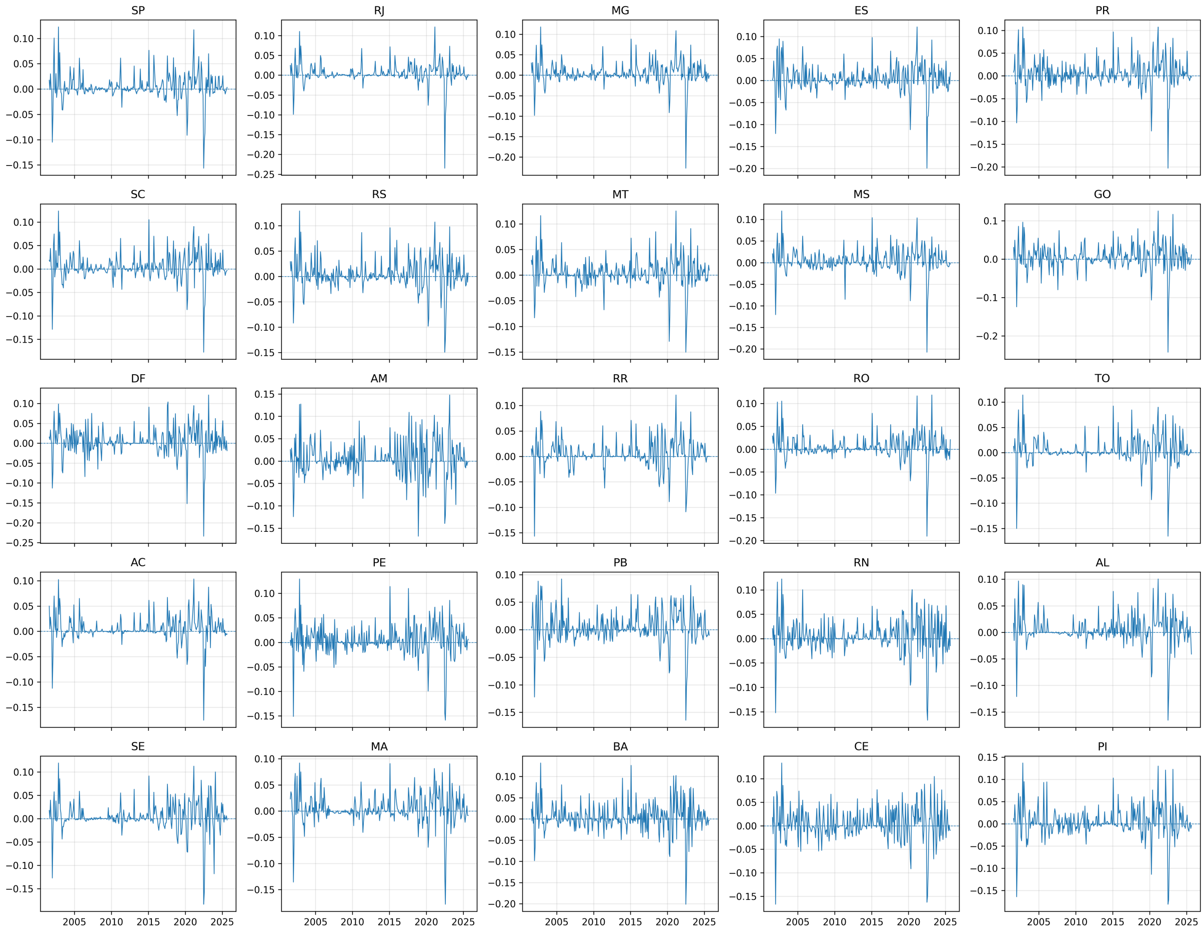
3.4 Data and Summary Statistics

For the purpose of forecasting fuel prices across several Brazilian states, we used the fuel price series available from the National Agency of Petroleum, Natural Gas and Biofuels (ANP)[2]. The following fuel types were extracted: gasoline, diesel, ethanol, and liquefied petroleum gas (LPG). A total of 25 states were considered, excluding Pará and

[2] The series are available from the Competition Defense Coordination (Coordenadoria de Defesa da Concorrência) through the Price Survey System (Sistema de Levantamento de Preços) at the following website: https://www.gov.br/anp/pt-br/assuntos/precos-e-defesa-da-concorrencia/precos/precos-revenda-e-de-distribuicao-combustiveis/serie-historica-do-levantamento-de-precos

Amapá. These two states were removed because they lacked data on diesel in Pará and ethanol in Amapá.

Figure 3.1 – Gasoline price in Brazilian states



Source: Elaborated by the author based on ANP data. SP = São Paulo; RJ = Rio de Janeiro; MG = Minas Gerais; ES = Espírito Santo; PR = Paraná; SC = Santa Catarina; RGS = Rio Grande do Sul; MT = Mato Grosso; MS = Mato Grosso do Sul; GO = Goiás; DF = Distrito Federal; AM = Amazonas; RR = Roraima; AP = Amapá; RO = Rondônia; AC = Acre; TO = Tocantins; PB = Paraíba; BA = Bahia; MA = Maranhão; PI = Piauí; PE = Pernambuco; AL = Alagoas; SE = Sergipe; RGN = Rio Grande do Norte; CE = Ceará.

This study uses the following dataset, which includes these fuels (gasoline, diesel S500, ethanol, and LPG), with monthly data spanning from July 2001 to September 2025. The fuel prices correspond to the retail price[3] (at the pump) in each state in Brazilian currency Real. Due to the COVID-19 pandemic, fuel prices were not collected in September 2020. Therefore, we applied a simple average interpolation between the August and October values to obtain an estimate for September. Finally, all series employed in this study were transformed by taking the first logarithmic difference to achieve stationarity.

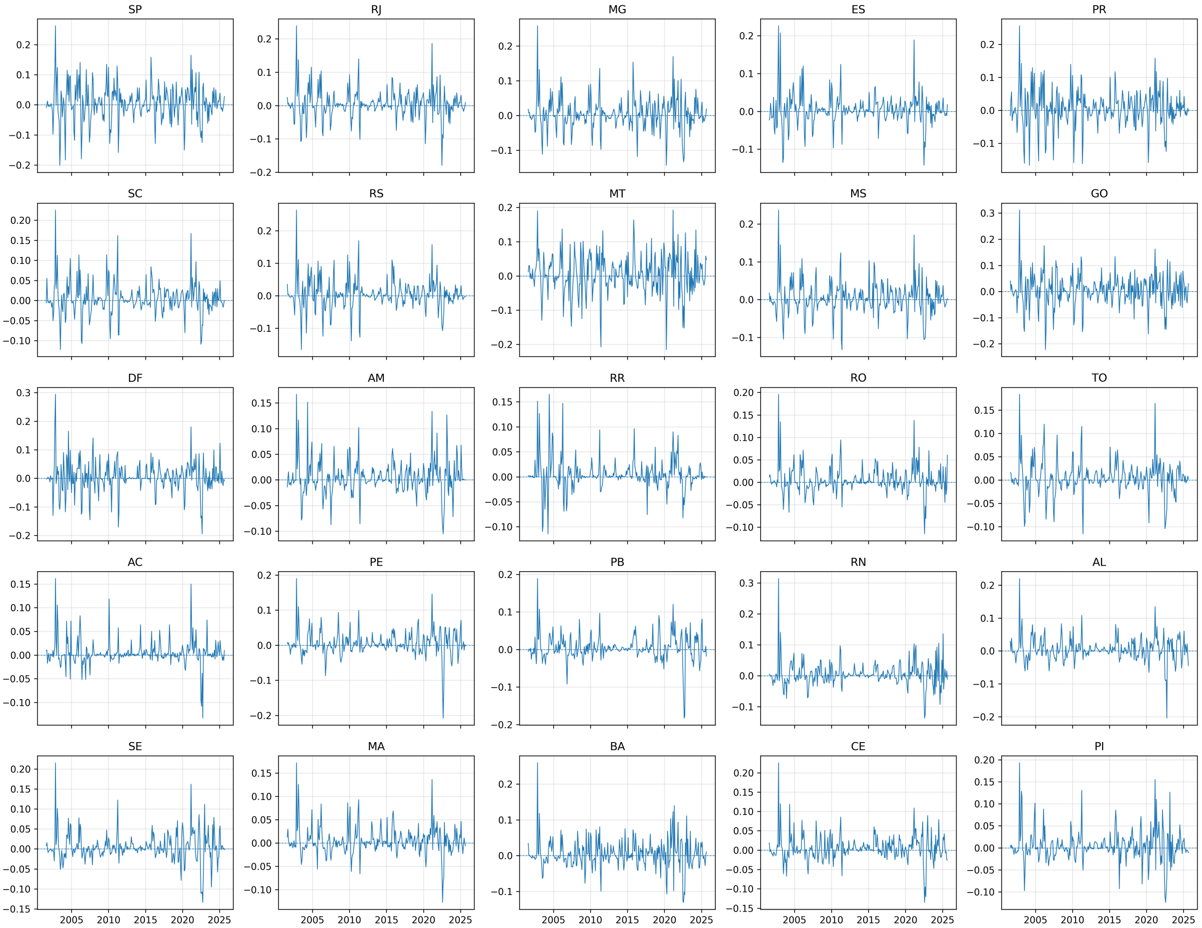
---

3 According to ANP, until October 30, 2004, the average price present in each state was the simple arithmetic mean. After that, the average retail prices were calculated considering the weights based on sales information sent to ANP. This calculation of means is valid at the state, regional, and national levels.

Furthermore, the use of the first logarithmic difference is due to two features: first, the fuels had to be on a logarithmic scale because of the considerable price differences among the fuels; second, the use of the first logarithmic difference allows measurement of the fuels' growth rates. This use shows changes in fuel prices as percentage points, which are useful for analyzing the Brazilian fuel market.

In the figures, the graphs for each state's gasoline, ethanol, diesel, and LPG prices are presented. Figure 3.1 corresponds to gasoline, Figure 3.2 to ethanol, Figure 3.3 to diesel, and Figure 3.4 to LPG.

Figure 3.2 – Ethanol price in Brazilian states



Source: Elaborated by the author based on ANP data. SP = São Paulo; RJ = Rio de Janeiro; MG = Minas Gerais; ES = Espírito Santo; PR = Paraná; SC = Santa Catarina; RGS = Rio Grande do Sul; MT = Mato Grosso; MS = Mato Grosso do Sul; GO = Goiás; DF = Distrito Federal; AM = Amazonas; RR = Roraima; AP = Amapá; RO = Rondônia; AC = Acre; TO = Tocantins; PB = Paraíba; BA = Bahia; MA = Maranhão; PI = Piauí; PE = Pernambuco; AL = Alagoas; SE = Sergipe; RGN = Rio Grande do Norte; CE = Ceará.

Figure 3.3 – Diesel price in Brazilian states

Source: Elaborated by the author based on ANP data. SP = São Paulo; RJ = Rio de Janeiro; MG = Minas Gerais; ES = Espírito Santo; PR = Paraná; SC = Santa Catarina; RGS = Rio Grande do Sul; MT = Mato Grosso; MS = Mato Grosso do Sul; GO = Goiás; DF = Distrito Federal; AM = Amazonas; RR = Roraima; AP = Amapá; RO = Rondônia; AC = Acre; TO = Tocantins; PB = Paraíba; BA = Bahia; MA = Maranhão; PI = Piauí; PE = Pernambuco; AL = Alagoas; SE = Sergipe; RGN = Rio Grande do Norte; CE = Ceará.

Figure 3.4 – LGP price in Brazilian states

Source: Elaborated by the author based on ANP data. SP = São Paulo; RJ = Rio de Janeiro; MG = Minas Gerais; ES = Espírito Santo; PR = Paraná; SC = Santa Catarina; RGS = Rio Grande do Sul; MT = Mato Grosso; MS = Mato Grosso do Sul; GO = Goiás; DF = Distrito Federal; AM = Amazonas; RR = Roraima; AP = Amapá; RO = Rondônia; AC = Acre; TO = Tocantins; PB = Paraíba; BA = Bahia; MA = Maranhão; PI = Piauí; PE = Pernambuco; AL = Alagoas; SE = Sergipe; RGN = Rio Grande do Norte; CE = Ceará.

Considering the graphs presented for each fuel, an upward trend can be observed across all series, with fuel prices reaching their peak during the COVID-19 pandemic. Beyond the COVID-19 pandemic, which significantly affected fuel prices in Brazil, other events also had substantial impacts. These include the changes in fuel pricing introduced in 2016 with the adoption of the Import Parity Price (IPP) policy, as well as the Russian invasion of Ukraine in 2022.

The IPP policy was an internal pricing rule set by Petrobras (the state-owned Brazilian oil company) that relied on several main factors: 1) the international price of crude oil; 2) the exchange rate between the Brazilian Real and US dollars; 3) import costs such as freight, insurance, harbor fees, and related expenses; and 4) a margin determined by Petrobras itself, serving as a safeguard to protect the company from financial losses when fluctuations in international oil prices were not reflected in domestic prices.

Meanwhile, the Russian invasion of Ukraine reduced the global oil supply. Following the imposition of several international sanctions, Russia was restricted from exporting oil and its derivatives to many countries. These measures had a significant impact on the Brazilian economy, given that Brazil relies on imported diesel from Russia to meet domestic demand Miller, Kumleben and Nowak (2024).

Finally, the descriptive statistics for gasoline, ethanol, diesel, and LPG are displayed in Table 3.2-3.5.

Table 3.2 – Descriptive statistics of the first log-differences gasoline prices by state

State	Mean	Median	Std_Dev	Skewness	Kurtosis	JB p-value
SP	0.0045	0.0012	0.0270	-0.3838	11.6861	<0.0001
RJ	0.0044	0.0015	0.0265	-2.2931	28.9186	<0.0001
MG	0.0045	0.0008	0.0284	-1.6822	20.7211	<0.0001
ES	0.0044	0.0004	0.0312	-0.8353	11.2023	<0.0001
PR	0.0047	0.0023	0.0314	-0.7513	11.2280	<0.0001
SC	0.0045	0.0008	0.0290	-0.6141	11.1946	<0.0001
RS	0.0044	0.0009	0.0306	-0.1595	8.0106	<0.0001
MT	0.0044	0.0016	0.0285	-0.4587	9.9947	<0.0001
MS	0.0042	0.0011	0.0287	-1.2861	15.8003	<0.0001
GO	0.0046	0.0031	0.0341	-1.2942	13.7557	<0.0001
DF	0.0047	0.0003	0.0358	-1.1123	11.3999	<0.0001
AM	0.0046	0.0001	0.0397	-0.1196	5.8684	<0.0001
RR	0.0047	0.0003	0.0282	-0.5480	9.0010	<0.0001
RO	0.0045	0.0007	0.0272	-0.7774	15.5046	<0.0001
TO	0.0043	0.0003	0.0282	-0.9306	12.4488	<0.0001
AC	0.0049	0.0003	0.0256	-0.9860	14.5170	<0.0001
PE	0.0045	0.0006	0.0321	-0.7381	9.5636	<0.0001
PB	0.0042	0.0005	0.0294	-0.7009	8.3499	<0.0001
RN	0.0044	0.0006	0.0349	-0.5579	7.7345	<0.0001
AL	0.0044	0.0000	0.0278	-0.8283	11.6394	<0.0001
SE	0.0048	0.0003	0.0315	-1.0017	11.8209	<0.0001
MA	0.0043	-0.0002	0.0290	-1.0045	11.1976	<0.0001
BA	0.0044	0.0009	0.0351	-0.4181	8.7914	<0.0001
CE	0.0045	0.0000	0.0358	-0.5969	7.4086	<0.0001
PI	0.0040	0.0000	0.0355	-0.5535	10.4281	<0.0001

Source: Own elaboration based on ANP data. JB means Jarque-Bera test.

Table 3.3 – Descriptive statistics of the first log-differences ethanol prices by state

State	Mean	Median	Std_Dev	Skewness	Kurtosis	JB p-value
SP	0.0053	0.0049	0.0589	-0.0845	5.1327	<0.0001
RJ	0.0051	0.0031	0.0428	0.5320	8.3904	<0.0001
MG	0.0049	0.0019	0.0456	0.6023	7.5550	<0.0001
ES	0.0046	0.0031	0.0400	0.9864	10.7066	<0.0001
PR	0.0055	0.0041	0.0550	-0.0671	5.6188	<0.0001
SC	0.0050	0.0018	0.0392	0.8735	8.5795	<0.0001
RS	0.0049	0.0010	0.0451	0.6232	8.5224	<0.0001
MT	0.0049	0.0026	0.0574	-0.2430	4.7194	<0.0001
MS	0.0043	0.0004	0.0420	0.6746	7.6610	<0.0001
GO	0.0052	0.0047	0.0593	0.0460	6.2805	<0.0001
DF	0.0047	0.0005	0.0535	0.2506	7.6926	<0.0001
AM	0.0052	0.0004	0.0333	0.9434	7.7970	<0.0001
RR	0.0048	0.0005	0.0312	1.0484	10.5232	<0.0001
RO	0.0051	0.0000	0.0292	1.4610	12.3770	<0.0001
TO	0.0048	0.0023	0.0347	0.5229	8.3163	<0.0001
AC	0.0048	0.0008	0.0270	1.1233	14.3298	<0.0001
PE	0.0053	0.0013	0.0343	-0.1279	12.3368	<0.0001
PB	0.0048	0.0013	0.0341	-0.2093	11.9467	<0.0001
RN	0.0051	0.0013	0.0384	1.9754	18.3300	<0.0001
AL	0.0054	0.0017	0.0352	0.3047	11.9690	<0.0001
SE	0.0051	0.0016	0.0350	0.8504	10.2293	<0.0001
MA	0.0047	0.0015	0.0295	1.0563	9.8141	<0.0001
BA	0.0048	0.0000	0.0401	0.8499	9.6446	<0.0001
CE	0.0050	0.0010	0.0327	1.0722	12.3408	<0.0001
PI	0.0044	0.0002	0.0356	0.9977	8.9143	<0.0001

Source: Own elaboration based on ANP data. JB means Jarque-Bera test.

Table 3.4 – Descriptive statistics of the first log-differences diesel prices by state

State	Mean	Median	Std_Dev	Skewness	Kurtosis	JB p-value
SP	0.0069	0.0005	0.0278	1.4470	9.2755	<0.0001
RJ	0.0070	0.0013	0.0261	1.6618	10.8198	<0.0001
MG	0.0067	0.0005	0.0284	1.2468	9.1697	<0.0001
ES	0.0067	0.0010	0.0281	1.1940	7.8744	<0.0001
PR	0.0069	0.0005	0.0299	1.3333	9.0333	<0.0001
SC	0.0071	0.0013	0.0281	1.2291	9.2978	<0.0001
RS	0.0071	0.0005	0.0295	1.4823	10.2076	<0.0001
MT	0.0067	0.0012	0.0280	1.3463	10.2424	<0.0001
MS	0.0067	0.0009	0.0279	1.2930	8.8352	<0.0001
GO	0.0066	0.0009	0.0295	1.3422	8.8422	<0.0001
DF	0.0069	0.0005	0.0305	1.2957	7.9848	<0.0001
AM	0.0071	0.0005	0.0285	0.9828	11.5801	<0.0001
RR	0.0070	0.0004	0.0285	0.7024	7.3575	<0.0001
RO	0.0069	0.0014	0.0273	1.4215	10.2219	<0.0001
TO	0.0064	0.0010	0.0286	1.5575	10.3616	<0.0001
AC	0.0074	0.0010	0.0261	1.3610	9.6836	<0.0001
PE	0.0068	0.0005	0.0288	1.4270	9.0986	<0.0001
PB	0.0065	0.0010	0.0287	1.4833	10.1214	<0.0001
RN	0.0068	0.0009	0.0296	0.9885	8.3383	<0.0001
AL	0.0067	0.0008	0.0285	0.9485	9.9793	<0.0001
SE	0.0066	0.0009	0.0319	0.6164	9.0922	<0.0001
MA	0.0068	0.0007	0.0292	1.4538	10.0020	<0.0001
BA	0.0067	0.0000	0.0322	1.2238	8.2974	<0.0001
CE	0.0065	0.0005	0.0297	1.2707	9.0188	<0.0001
PI	0.0066	0.0006	0.0299	1.2054	9.1023	<0.0001

Source: Own elaboration based on ANP data. JB means Jarque-Bera test.

Table 3.5 – Descriptive statistics of the first log differences of LPG prices by state

State	Mean	Median	Std_Dev	Skewness	Kurtosis	JB p-value
SP	0.0062	0.0008	0.0212	3.2541	21.9606	<0.0001
RJ	0.0058	0.0005	0.0203	3.0937	19.3227	<0.0001
MG	0.0061	0.0010	0.0220	3.6247	22.7392	<0.0001
ES	0.0056	0.0012	0.0274	2.7589	20.0620	<0.0001
PR	0.0059	0.0007	0.0239	1.9824	17.1330	<0.0001
SC	0.0062	0.0018	0.0202	1.0718	18.6832	<0.0001
RS	0.0062	0.0017	0.0226	3.5840	27.2726	<0.0001
MT	0.0058	0.0010	0.0229	2.7479	20.1271	<0.0001
MS	0.0060	0.0019	0.0221	2.4586	16.7425	<0.0001
GO	0.0064	0.0009	0.0268	3.6188	23.9484	<0.0001
DF	0.0060	0.0004	0.0260	2.3283	17.6820	<0.0001
AM	0.0074	0.0020	0.0223	1.9187	12.8905	<0.0001
RR	0.0073	0.0018	0.0173	2.5369	11.8790	<0.0001
RO	0.0071	0.0017	0.0223	1.7702	11.5041	<0.0001
TO	0.0066	0.0010	0.0233	3.2620	21.1965	<0.0001
AC	0.0066	0.0023	0.0189	2.0372	15.1051	<0.0001
PE	0.0062	0.0008	0.0298	1.7303	11.5828	<0.0001
PB	0.0061	0.0010	0.0265	1.2795	16.0962	<0.0001
RN	0.0062	0.0007	0.0301	2.2748	15.0382	<0.0001
AL	0.0062	0.0007	0.0265	2.9657	19.7745	<0.0001
SE	0.0064	0.0013	0.0226	2.5491	14.4626	<0.0001
MA	0.0070	0.0015	0.0243	2.7914	16.0457	<0.0001
BA	0.0071	0.0014	0.0248	3.1079	20.6139	<0.0001
CE	0.0070	0.0019	0.0262	4.0673	29.8398	<0.0001
PI	0.0064	0.0017	0.0304	3.3195	26.9381	<0.0001

Source: Own elaboration based on ANP data. JB means Jarque-Bera test.

Based on all the results, there is asymmetry in all growth rates of the fuel price series presented in this study, as indicated by the skewness values. The kurtosis values indicated that all fuel price series have heavy tails. Finally, according to the Jarque-Bera test proposed by Jarque and Bera (1987), which tests whether the data follow a normal distribution. The results indicated that none of the growth rates of the fuel price series follow a normal distribution. These results were expected due to asymmetry and heavy tails, and, given the absence of normality in the price time series, they are well-known facts in the literature.

3.5 Results

3.5.1 Estimation for Brazil

This study presents an application of the VAR, MAR, and EMAR models to Brazilian fuel price forecasting. First, set the parameter values for these models: $\hat{\rho}$, $\hat{u_1}$, and $\hat{u_2}$. Thereby, $\hat{\rho}$ is the lag order for the models VAR, MAR, and EMAR, while $\hat{u_1}$ is the number of common factors for the rows (types of fuel) in the EMAR model, and $\hat{u_2}$ is the number of common factors for the columns in the EMAR (states).

According to Samadi and Alwis (2025), the values for $\hat{\rho}$, $\hat{\mu_1}$, and $\hat{\mu_2}$ can be determined using the BIC and AIC criteria. However, this study predicted fuel price growth rates using models that employ a rolling-window scheme, resulting in different numbers of parameters for $\hat{\mu_1}$ and $\hat{\mu_2}$ in each window predicted. As a consequence, the number of parameters will change in each window until the end of the forecast. Therefore, if the BIC and AIC criteria are used to determine parameter values, different parameter values will be obtained for each predicted window.

Moreover, when BIC fails, it tends to overestimate the number of $\hat{\mu_1}$ and $\hat{\mu_2}$, which may slightly reduce estimation efficiency, though it avoids the risk of omitting key subspaces and introducing bias (Cook; Forzani; Zhang, 2015; Forzani; Su, 2021). Finally, another approach to this question is to fix the parameter values.

Using the parameters in fixed mode raises one concern: how to decide which combination of those parameters is best. One possibility is to use trial and error to identify the combination with the lowest Mean Squared Error (MSE) and to respect the parsimony principle. The parsimony principle is useful for solving other estimation problems. One of these concerns the best parameter combination, the MSE value, and the degrees of freedom. Because it may happen that the best combination that minimizes the MSE consumes many degrees of freedom as the parameter values increase.

The parameter values were selected to minimize the in-sample MSE, under the following assumptions: 1) The parameter $\hat{\rho}$ was set to 1, as higher values would result in insufficient observations for estimating the baseline VAR model. 2) The parameter $\hat{u_1}$ was fixed at 2, given that the database contains only four fuel types corresponding to the matrix rows. 3) The parameter $\hat{u_2}$ was also set to 2, since higher values produced comparable MSE results but significantly reduced the degrees of freedom. The different values for $\hat{u_2}$ are present in Tables C.1-C.3 in Appendix C.

Therefore, in this study, the following parameters were considered for the models: the lag order for the VAR, MAR, and EMAR models was set to $\hat{\rho} = 1$, and the envelope dimensions were set to $\hat{u_1} = 2$ and $\hat{u_2} = 2$.

Table 3.6 displays the MSE values for all models estimated, where NOP (number of parameters) and h = 1, 3, 6, and 12 represent the step-ahead.

Table 3.6 – Forecasting performance for the models considering four fuels and twenty-five Brazilian states

Model	**NOP**	**MSE in sample**	**MSFE out of sample**			
			h = 1	**h = 3**	**h = 6**	**h = 12**
VAR	15150	0.1085	0.1014	0.0943	0.0836	0.0621
MAR	1074	0.0867	0.0822	0.0766	0.0677	0.0504
EMAR	491	0.0855	0.0816	0.0759	0.0672	0.0499

Source: Elaborated by the author based on results by models. Being $\hat{u_1} = 2$ and $\hat{u_2} = 2$

Table 3.6 displays the mean squared error (MSE) values for three models across a range of out-of-sample forecast horizons (h). The NOP column represents the number of parameters estimated for each model. The results indicate that the VAR model consistently produces higher MSE values for all out-of-sample horizons analyzed. By contrast, both the MAR and EMAR models achieve significant reductions in MSE. Furthermore, the MAR model decreases the number of required parameters by approximately 93%, while the EMAR model achieves a reduction of about 97%. The EMAR model also demonstrates a slightly lower MSE compared to the MAR model. These outcomes align with the findings of Samadi and Billard (2025), who observed that the EMAR model not only delivers lower MSE in forecasting applications but also utilizes fewer parameters. In summary, given the increased degrees of freedom afforded by the EMAR approach, its use is recommended on the grounds of parsimony.

In addition, although the difference between the MAR and EMAR in MSE values is small, this study addresses two main features of the dataset: first, the data are high-dimensional in the columns (25 states are being analysed in the estimation), and second, the data exhibit non-normal errors as seen in the database section. According to Samadi and Alwis (2025), the EMAR model performs well with these features, as demonstrated by its ability to reduce the average estimation error using simulated data with non-normal errors and to reduce the number of columns during estimation. Given that, the model handles high-dimensional data and non-normal errors better than other models, such as MAR and VAR. Finally, as this study uses time-series price data (non-normal errors) and analyses for each Brazilian state (25 states, high-dimensional estimation), the best option is to use the EMAR model to forecast fuel prices in Brazil.

3.5.2 Estimation for Brazilian regions

One issue with the EMAR estimation using the Brazilian dataset is the disparity between the number of rows (types of fuels) and the number of columns (number of states). The last estimation had 4 rows and 25 columns, which having these numbers, is far from the normal ensemble presented in Samadi and Billard (2025). In the original paper, the authors used two real datasets with 5x4 (five rows and four columns) and 9x8 (nine rows and eight columns), noting that these datasets are nearly equal in row and column counts. From this perspective, we decided to estimate for each Brazilian region using all fuels and to assess whether the EMAR-based estimation is the best for forecasting fuel prices in Brazil.

Considering the Brazilian region, the estimate was made for the Southeast, Midwest, North, and Northeast. Unfortunately, due to the number of states in the South region of Brazil (three states), the model was not able to estimate using only three states. Thus, it has the following arrangements: Southeast with 4 fuels and 4 states (4x4); North with 4 fuels and 5 states (4x5); Midwest with 4 fuels and 4 states (4x4); and Northeast with 4 fuels and 9 states (4x9). The results are in Tables 3.7-3.10 below.

Table 3.7 – Forecasting performance for the models considering four fuels and the southeast region

Model	**NOP**	**MSE in sample**	**MSE out of sample**			
			h = 1	**h = 3**	**h = 6**	**h = 12**
VAR	408	0.0188	0.0173	0.0161	0.0143	0.0106
MAR	66	0.0150	0.0138	0.0128	0.0114	0.0085
EMAR	50	0.0147	0.0137	0.0127	0.0113	0.0084

Source: Elaborated by the author based on results by models, $\hat{u_1} = 2$ and $\hat{u_2} = 2$. (NOP) number of parameters. $\hat{\rho} = 1$.

Table 3.8 – Forecasting performance for the models considering four fuels and the Midwest region

Model	**NOP**	**MSE in sample**	**MSE out of sample**			
			h = 1	**h = 3**	**h = 6**	**h = 12**
VAR	408	0.0224	0.0208	0.0194	0.0172	0.0127
MAR	66	0.0191	0.0178	0.0166	0.0147	0.0109
EMAR	50	0.0210	0.0195	0.0181	0.0160	0.0119

Source: Elaborated by the author based on results by models, $\hat{u_1} = 2$ and $\hat{u_2} = 2$. (NOP) number of parameters. $\hat{\rho} = 1$.

Table 3.9 – Forecasting performance for the models considering four fuels and the Northeast region

Model	**NOP**	**MSE in sample**	**MSE out of sample**			
			h = 1	**h = 3**	**h = 6**	**h = 12**
VAR	408	0.0306	0.0345	0.0321	0.0284	0.0210
MAR	66	0.0307	0.0288	0.0268	0.0237	0.0175
EMAR	50	0.0309	0.0293	0.0279	0.0241	0.0178

Source: Elaborated by the author based on results by models, $\hat{u_1} = 2$ and $\hat{u_2} = 2$. (NOP) number of parameters. $\hat{\rho} = 1$.

Table 3.10 – Forecasting performance for the models considering four fuels and the North region

Model	**NOP**	**MSE in sample**	**MSE out of sample**			
			h = 1	**h = 3**	**h = 6**	**h = 12**
VAR	408	0.0165	0.0155	0.0144	0.0128	0.0095
MAR	66	0.0130	0.0122	0.0114	0.0101	0.0075
EMAR	50	0.0133	0.0126	0.0117	0.0104	0.0077

Source: Elaborated by the author based on results by models, $\hat{u_1} = 2$ and $\hat{u_2} = 2$. (NOP) number of parameters. $\hat{\rho} = 1$.

Considering the analyses presented in Tables 3.7-3.10, for regions of Brazil: Southeast, Midwest, Northeast, and North, respectively. The MSE column in each table represents the Mean Squared Error (MSE) for three models across a range of out-of-sample forecast horizons (h). The NOP column represents the number of parameters estimated in each model. The results indicated that the MSE values were practically equal across the estimated models. Moreover, the MAR model decreases the number of parameters required for estimation by approximately 83%, as the EMAR model reduces it by approximately 87%. Therefore, it is correct to affirm that, according to the Brazilian region's estimates, no model was superior in terms of predictive performance.

Although, the EMAR and MAR models have yield the same predictions; the option for forecasting fuel prices in each region should use the EMAR model. The EMAR model estimates 50 parameters per region, whereas the MAR model estimates 66 parameters per region. Therefore, the EMAR model is preferable to the MAR model, as it is more parsimonious. Lastly, it is important to note that the EMAR model performs better in high-dimensional predictive estimation than in low-dimensional estimation.

3.6 Conclusion

This study analyzed the prices of four fuels (gasoline, ethanol, diesel, and LPG) from July 2001 to September 2025, considering 25 Brazilian states. The main goal was to forecast these prices, using the following models: VAR, MAR, and EMAR, with the VAR model as the benchmark.

The novelty of this study was the application of the EMAR model to forecast fuel prices in Brazil. There are some reasons for that: 1) The model uses the prices of all fuels and states to produce the forecasts; 2) the EMAR model can separate information into useful and non-useful components by using subspaces of the sample to forecast fuel prices, with each subspace containing only useful information; 3) finally, , by using subspaces, the model requires fewer parameters for estimation compared to other methodologies, such as MAR and VAR.

Overall, the study sought to explore how relationships among fuel prices across different states in Brazil are connected. Given this, the price of specific fuel in a state can help forecast the price of the same fuel or other fuels in other states. Thus, analyzing the behaviour of different fuels across many states is beneficial for forecasting fuel prices in general.

Based on the results from all models, it is noted that the MAR and EMAR models have the same MSE values both in-sample and out-of-sample. However, the EMAR model reduced the number of required parameters by approximately 97%, as MAR was reduced by 93% compared to the VAR model. Therefore, Therefore, even if the EMAR model has not achieved the lowest MSE, as seen in Samadi and Alwis (2025), the parsimony criterion suggests that the best model is the EMAR model.

Lastly, the EMAR model's forecasts yielded new insights into retail fuel price forecasting in Brazil. These insights may prove useful for policymakers and transport sector authorities in the future. Furthermore, the results present a new approach to the fuel forecasting literature, as this study extends the international literature by using data from a large country that heavily relies on road transportation.

Final Considerations

This study comprises three essays on the dynamics of fuel prices in Brazil.

The first essay forecasts gasoline and diesel prices for each state in Brazil. Beyond this, the first essay sought to present a general panorama of the forecast for the Brazilian state. To obtain this panorama, three forecasting models were used: ARIMA, Exponential Smoothing, and Singular Spectrum Analysis, from July 2001 to August 2023. Thus, this study offered a new perspective on fuel forecasting in Brazil, filling a gap in national literature.

The second essay aimed to understand how oil price volatility affects domestic fuel price volatility from July 2001 to August 2023. This essay explored the volatility transmission between oil prices and Brazilian gasoline and diesel prices. The application of BEKK-MGARCH models yielded results that explained this dynamic relationship between oil/domestic fuels. Moreover, it was considered in two periods: the entire period (July 2001 to August 2023) and the IPP period (October 2016 to April 2023). Once that, during these periods, the fuel pricing policy in Brazil varied, which influenced the impact of oil volatility on domestic fuel prices.

The third essay offered a new perspective on fuel forecasting in Brazil. Considering Brazilian states and different types of fuel (gasoline, ethanol, diesel, and LPG) from July 2001 to September 2023. This essay sought to conduct a multivariate forecast of fuel prices in Brazil. Moreover, this essay introduces the national and international literature to a new forecasting methodology (EMAR), which has advantages over other multivariate methodologies. These advantages are: 1) it allows us to consider different types of fuels across states to do a forecast for a specific fuel in a certain state; 2) the methodology can separate the useful from the useless information. Thus, the model focuses on useful information, which, in turn, requires fewer parameters than traditional methodologies (e.g., VAR and EMAR). Finally, to assess accuracy, MSE was used.

For the first essay, it was expected that the SSA model would be the best-performing model based on the international literature. However, the results showed that the ARIMA model performed best in terms of RMSE. Nevertheless, the Modified Diabold-Mariano Test, which tests for statistical significance between two forecasting models, showed that there was statistical significance only for horizon h=1; in other forecast horizons, there were few forecasts with statistical significance. Therefore, it was not possible to claim that a single model was best at forecasting fuel prices in Brazil.

In the second essay, the BEKK-MGARCH results indicate that gasoline is most sensitive to oil price volatility, whereas diesel prices are relatively stable. Moreover, considering only the IPP period, the results showed that the base volatility during this period is higher than that for the entire period, and that oil shocks did not affect gasoline and diesel prices. Once that, during the IPP period, Brazilian fuel prices were adjusted to international oil prices, thus the oil shocks were softened, and the base volatility was higher than during the entire period. One reason for this difference in volatility is that, throughout the period, the Brazilian government froze fuel prices at certain times to control inflation and to comply with the interests of third parties.

The third essay demonstrated that the EMAR model had a slightly slower MSE than the MAR model. It was expected to see a larger difference in MSE values between the EMAR and MAR, as noted by Samadi and Alwis (2025). However, the EMAR model estimated the same results of MAR using 97% less parameters than VAR model, as the MAR model reduced the parameters in 93%. So, considering these reduction in terms of parameters, the best choice was to use the EMAR. Furthermore, this study explored other estimations, considering fewer states for four types of fuels, and found that estimates from each Brazilian region were equal for both models. Despite having the same MSE value, it is preferable to consider the EMAR estimation due to its use of fewer parameters, thereby addressing a parsimony issue.

Finally, all these studies offered a new perspective on the dynamics of domestic fuel prices, using different approaches. Once that, each approach presented in the essays sought to understand the behavior of fuel prices in Brazil. Therefore, this study filled gaps in the international and national literature on univariate and multivariate forecasting, as well as volatility.

REFERENCES

AASTVEIT, K. A.; BJØRNLAND, H. C.; THORSRUD, L. A. What drives oil prices? emerging versus developed economies. *Journal of Applied Econometrics*, v. 30, n. 7, p. 1013–1028, 2015.

ABRAMSON, B. Using belief networks to forecast oil prices. *International Journal of Forecasting*, v. 7, p. 299–315, 1991.

AKçELIK, F.; ÖGüNç, F. Pass-through of crude oil prices at different stages in turkey. *Central Bank Review*, v. 16, p. 41–51, 2016.

ALMEIDA, E. L. F.; OLIVEIRA, P. V.; LOSEKANN, L. Impacts of fuel price controls in brazil and pricing mechanism options. *Revista de Economia Política*, v. 35, n. 3, p. 531–556, 2015.

ALVES, D. H.; ASSIS, L. A.; SANTOS, V. E. S.; TAVARES, F. B. R. Impactos da greve dos caminhoneiros à luz do código de defesa do consumidor. *JURIS – Revista da Faculdade de Direito*, v. 28, n. 2, p. 155–166, 2018.

ANJOS, C. N. D.; BESARRIA, C. d. N.; MEDEIROS, R. K. d.; JESUS, D. P. de; ARAúJO, H. A. C. de. Oil price forecast: A comparison between arima, lstm and arima-lstm models. In: *Anais do Congresso de Séries Temporais e Econometria*. Florianópolis: , 2023. Apresentado em 2023.

ASUAMAH, S.; OHENE-MANU, J. An econometric investigation of forecasting premium fuel. *International Journal of Energy Economics and Policy*, v. 5, n. 3, p. 716–724, 2015.

BACON, R. W. Rockets and feathers: The asymmetric speed of adjustment of uk retail gasoline prices to cost changes. *Energy Economics*, v. 13, n. 3, p. 211–218, 1991.

BAGHESTANI, H. Predicting gasoline prices using michigan survey data. *Energy Economics*, v. 50, p. 27–32, 2015.

BARBAGLIA, L.; CROUX, C.; WILMS, I. Volatility spillovers in commodity markets: A large t-vector autoregressive approach. *Energy Economics*, v. 85, p. 104554, 2020.

BARBOSA, G.; SOUSA, D.; BESARRIA, C. N.; LIMA, R.; JESUS, D. P. Asymmetry in the prices of crude oil and diesel and gasoline prices in brazil. *Journal of Economic Studies*, 2023.

BARRA, G. M. J.; SILVA, R. O.; SILVEIRA, R. I. M. Impacto da greve dos caminhoneiros na gestão de risco em cadeias de suprimentos: o caso de um hospital da zona da mata mineira. *Brazilian Journal of Production Engineering (BJPE)*, v. 6, n. 6, p. 11–28, 2020.

BAUMEISTER, C.; KILIAN, L.; LEE, T. Inside the crystal ball: New approaches to predicting the gasoline price at the pump. *Journal of Applied Econometrics*, v. 32, p. 275–295, 2017.

BAYARRI, M. J.; BERGER, J. O. The interplay of bayesian and frequentist analysis. 2004.

BERRISCH, J.; ZIEL, F. Distributional modeling and forecasting of natural gas prices. *Journal of Forecasting*, v. 41, p. 1065–1086, 2022.

BOLLERSLEV, T. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, v. 31, n. 3, p. 307–327, 1986.

BOLLERSLEV, T. Modelling the coherence in short-run nominal exchange rates: A multivariate generalized arch model. *Review of Economics and Statistics*, v. 72, p. 498–505, 1990.

BOLLERSLEV, T.; ENGLE, R. F.; WOOLDRIDGE, J. M. A capital asset pricing model with time-varying covariances. *The Journal of Political Economy*, v. 96, p. 116–131, 1988.

BOROUMAND, R. H.; GOUTTE, S.; PORCHER, S.; PORCHER, T. Asymmetric evidence of gasoline price responses in france: A markov-switching approach. *Economic Modelling*, v. 51, p. 206–214, 2015.

BOX, G.; JENKINS, G. *Time series analysis: Forecasting and control*. San Francisco, CA: Holden-Day, 1970.

Brazilian Chamber of Deputies. *Lei Complementar nº 194, de 23 de junho de 2022*. 2022. Acesso em: 16 maio 2026. Available on: <https://www2.camara.leg.br/legin/fed/leicom/2022/leicomplementar-194-23-junho-2022-792887-norma-pl.html>.

Brazilian Institute of Geography and Statistics. *National Consumer Price Index*. 2024. Accessed on: September 3, 2024. Available on: <https://www.ibge.gov.br/estatisticas/economicas/precos-e-custos/9256-indice-nacional-de-precos-ao-consumidor-amplo.html?=&t=resultados>.

CAGNIN, R. F.; PRATES, D. M.; FREITAS, M. C. P.; NOVAIS, L. F. The macroeconomic management of the dilma government (2011 and 2012). *Novos Estudos*, v. 97, nov 2013.

CARPIO, L. G. T. The effects of oil price volatility on ethanol, gasoline, and sugar price forecasts. *Energy*, v. 181, p. 1012–1022, 2019.

Central Bank of Brazil. *Relatório de Inflação: Efeito da paralisação no setor de transporte de cargas sobre a inflação ao consumidor*. 2018. Acessado em: 01 out. 2023. Available on: <https://www.bcb.gov.br/publicacoes/boxrelatoriosinf>.

CHANG, D.; SERLETIS, A. Oil, uncertainty, and gasoline prices. *Macroeconomic Dynamics*, v. 22, p. 546–561, 2018.

CHEN, R.; XIAO, H.; YANG, D. Autoregressive models for matrix-valued time series. *Journal of Econometrics*, v. 222, p. 539–560, 2021.

CHEN, Y.; LU, J.; MA, M. How does oil future price imply bunker price—cointegration and prediction analysis. *Energies*, v. 15, n. 10, p. 3630, 2022. Artigo de acesso aberto.

CHEVALLIER, J. Time-varying correlations in oil, gas and co prices: An application using bekk, ccc and dcc-mgarch models. *Applied Economics*, v. 44, n. 32, p. 4257–4274, 2012.

Civil House. *Government Announces the Increase of Ethanol Blending in Gasoline to 27%*. 2015. Accessed on: December 14, 2024. Available on: <https://www.gov.br/casacivil/pt-br/assuntos/noticias/2015/marco/mistura-de-etanol-na-gasolina-sera-de-27-a-partir-de-16-de-marco>.

CONWAY, J. *A Course in Functional Analysis*. Second. New York: Springer, 1990.

COOK, R. D.; FORZANI, L.; ZHANG, X. Envelopes and reduced-rank regression. *Biometrika*, v. 102, n. 2, p. 439–456, 2015.

COOK, R. D.; LI, B.; CHIAROMONTE, F. Envelope models for parsimonious and efficient multivariate linear regression. *Statistica Sinica*, p. 927–960, 2010.

COOK, R. D.; ZHANG, X. Algorithms for envelope estimation. *Journal of Computational and Graphical Statistics*, v. 25, n. 1, p. 284–300, 2016.

DADA, J. T.; AKINLO, T. Asymmetric relationship between oil prices and remittances in oil-importing and-exporting countries: Evidence from sub-saharan africa. *OPEC Energy Review*, Wiley Online Library, v. 47, n. 2, p. 118–133, 2023.

DELGADO, F.; GAUTO, M. *Composição dos Preços de Paridade dos Combustíveis no Brasil*. 2021. Conjuntura Econômica | FGV Energia. Accessed on: February 19, 2026. Available on: <https://fgvenergia.fgv.br/publicacao/petroleo-qualidade-fisico-quimicas-precos-emercados-o-caso-das-correntes-nacionais>.

DIEBOLD, F. X.; MARIANO, R. S. Comparing predictive accuracy. *Journal of Business and Economic Statistics*, v. 13, n. 3, p. 253–263, 1995.

DING, S.; COOK, R. D. Matrix variate regressions and envelope models. *Journal of the Royal Statistical Society: Series B*, v. 80, n. 2, p. 387–408, 2018.

Economática. *Petrobras' Loss Reaches the Second Highest Level in the History of Publicly Traded Brazilian Companies*. 2015. Accessed on: February 3, 2015. Available on: <https://insight.economatica.com/prejuizo-da-petrobras-registra-o-segundo-maior-nivel-da-historia/#:~:text=O%20preju%C3%ADzo%20da%20Petrobr%C3%A1s%20no,R%24%2044%2C2%20bilh%C3%B5es>.

ELDER, J. Another perspective on the effects of inflation volatility. *Journal of Money, Credit and Banking*, v. 36, p. 911–928, 2004.

ELDER, J.; SERLETIS, A. Oil price uncertainty. *Journal of Money, Credit and Banking*, v. 42, p. 1137–1159, 2010.

ELSNER, J. B.; TSONIS, A. A. *Singular spectral analysis: A new tool in time series analysis*. 1st. ed. New York, NY: Plenum Press, 1996.

Energy Research Company. *Behavior of international oil prices in the 2nd half of 2017 (Informe EPE-DPG-SPT No. 02)*. Brasília, DF, Brasil, 2017.

Energy Research Company. *International oil prices: Studies for the ten-year energy expansion plan 2031*. Brasília, DF, Brasil, 2021.

ENGLE, R. F. Dynamic conditional correlation. *Journal of Business & Economic Statistics*, v. 20, n. 3, p. 339–350, 2002.

ENGLE, R. F.; KRONER, K. F. Multivariate simultaneous generalized arch. *Econometric Theory*, v. 11, p. 122–150, 1995.

FGV Europe Projects. *The truck drivers' strike: Origins and impacts on the Brazilian economy*. 2018.

FORZANI, L.; SU, Z. Envelopes for elliptical multivariate linear regression. *Statistica Sinica*, v. 31, p. 301–332, 2021.

GARCíA-MARTOS, C.; RODRíGUEZ, J.; SáNCHEZ, M. Modelling and forecasting fossil fuels, co and electricity prices and their volatilities. *Applied Energy*, v. 101, p. 363–375, 2013.

GAUTO, M.; DELGADO, F.; COUTO, M. The paradox of fuel prices. *Conjuntura Econômica*, 2021.

GELMAN, A. et al. *Bayesian Data Analysis*. 3. ed. : CRC Press, 2013.

GELMAN, A.; RUBIN, D. B. Inference from iterative simulation using multiple sequences. *Statistical Science*, v. 7, n. 4, p. 457–511, 1992.

GOLYANDINA, N.; KOROBEYNIKOV, A. Basic singular spectrum analysis and forecasting with r. *Computational Statistics & Data Analysis*, v. 71, p. 934–954, 2014.

GOLYANDINA, N.; NEKRUTKIN, V.; ZHIGLJAVSKY, A. *Analysis of time series structure: SSA and related techniques*. Boca Raton, FL: Chapman & Hall/CRC, 2001.

GOLYANDINA, N.; ZHIGLJAVSKY, A. *Singular spectrum analysis for time series*. Berlin, Germany: Springer, 2013. (Springer Briefs in Statistics).

GUO, L.; HUANG, X.; LI, Y.; LI, H. Forecasting crude oil futures price using machine learning methods: Evidence from china. *Energy Economics*, v. 127, p. 107089, 2023.

HAMILTON, J. D. Oil and the macroeconomy since world war ii. *Journal of Political Economy*, v. 91, n. 2, p. 228–248, 1983.

HARVEY, D. I.; LEYBOURNE, S. J.; NEWBOLD, P. Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, v. 13, p. 281–291, 1997.

HARVEY, D. I.; LEYBOURNE, S. J.; NEWBOLD, P. Tests for forecast encompassing. *Journal of Business and Economic Statistics*, v. 16, p. 254–259, 1998.

HASSANI, H. Singular spectrum analysis: Methodology and comparison. *Journal of Data Science*, v. 5, n. 2, p. 239–257, 2007.

HASSANI, H.; HERAVI, S.; ZHIGLJAVSKY, A. Forecasting european industrial production with singular spectrum analysis. *International Journal of Forecasting*, v. 25, p. 103–118, 2009.

HASSANI, H.; SILVA, E. S.; ANTONAKAKIS, N.; FILIS, G.; GUPTA, R. Forecasting accuracy evaluation of tourist arrivals. *Annals of Tourism Research*, v. 63, p. 112–127, 2017.

HASSANI, H.; WEBSTER, A.; SILVA, E. S.; HERAVI, S. Forecasting u.s. tourist arrivals using optimal singular spectrum analysis. *Tourism Management*, v. 46, p. 332–335, 2015.

HORáK, J.; JANNOVá, M. Predicting the oil price movement in commodity markets in global economic meltdowns. *Forecasting*, v. 5, p. 374–389, 2023.

HYNDMAN, R. J.; ATHANASOPOULOS, G. *Forecasting: Principles and Practice*. : OTexts, 2018.

Instituto Brasileiro de Geografia e Estatística. *Áreas Territoriais 2025*. 2025. Acesso em: 28 jun. 2026. Available on: <https://www.ibge.gov.br/geociencias/organizacao-do-territorio/estrutura-territorial/15761-areas-dos-municipios.html>.

Intercontinental Exchange. *Global oil markets*. 2024. <https://www.ice.com/oil>. Acesso em: 2 mar. 2024.

IPEAData. *IPC (FIPE) – Taxa de variação*. 2024. <http://www.ipeadata.gov.br/Default.aspx>. Acesso em: 2 mar. 2024.

JARQUE, C. M.; BERA, A. K. A test for normality of observations and regression residuals. *International Statistical Review*, v. 55, n. 2, p. 163–172, 1987.

KANCZUK, F. Brazil through the eyes of chorinho. *Brazilian Review of Econometrics*, v. 35, n. 2, p. 171–195, 2015.

KILIAN, L.; LüTKEPOHL, H. *Structural Vector Autoregressive Analysis*. 1. ed. Cambridge, MA: Cambridge University Press, 2017.

KLEIN, C. E.; SILVA, W. V. D.; VEIGA, C. P. D.; MARIANI, V. C.; COELHO, L. D. S. Fuel price forecasting combining wavelet neural network and adaptive differential evolution. *International Journal of Business Forecasting and Marketing Intelligence*, v. 6, n. 3, p. 167–185, 2020.

KUSCHNIG, N.; VASHOLD, L. Bvar: Bayesian vector autoregressions with hierarchical prior selection in r. *Journal of Statistical Software*, v. 100, n. 14, nov. 2021.

LAMOTTE, O.; PORCHER, T.; SCHALCK, C.; SILVESTRE, S. Asymmetric gasoline price responses in france. *Applied Economics Letters*, v. 20, n. 5, p. 457–461, 2013.

MENSI, W.; REHMAN, M. U.; VO, X. V. Dynamic frequency relationships and volatility spillovers in natural gas, crude oil, gas oil, gasoline, and heating oil markets: Implications for portfolio management. *Resources Policy*, Elsevier, v. 73, p. 102172, 2021.

METROPOLIS, N.; ROSENBLUTH, A. W.; ROSENBLUTH, M. N.; TELLER, A. H.; TELLER, E. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, v. 21, n. 6, p. 1087–1092, 1953.

MILLER, C.; KUMLEBEN, N.; NOWAK, C. *Sanctions on Russia and the Splintering of the World Oil Market*. Washington, DC, 2024.

Ministry of Mines and Energy. *Secretariat of Oil, Natural Gas, and Biofuels*. 2025. Accessed on: May 26, 2025. Available on: <https://antigo.mme.gov.br/web/guest/secretarias/petroleo-gas-natural-e-biocombustiveis/acoes-e-programas/programas/biodiesel/o-programa#:~:text=Principais%20diretrizes%20do%20PNPB%3A&text=A%20Lei%20n%C2%BA%2011.097%2C%20de,qualquer%20parte%20do%20territ%C3%B3rio%20nacional>.

Ministry of Mines and Energy. *Secretariat of Oil, Natural Gas, and Biofuels*. 2025. <https://antigo.mme.gov.br/web/guest/secretarias/petroleo-gas-natural-e-biocombustiveis/acoes-e-programas/programas/biodiesel/o-programa#:~:text=Principais%20diretrizes%20do%20PNPB%3A&text=A%20Lei%20n%C2%BA%2011.097%2C%20de,qualquer%20parte%20do%20territ%C3%B3rio%20nacional>. Accessed: 26 May 2025.

Ministry of transport. *Frota de veículos 2025*. 2025. Acesso em: 6 mar. 2026. Available on: <https://www.gov.br/transportes/pt-br/assuntos/transito/conteudo-Senatran/frota-de-veiculos-2025>.

MORAES, M. L.; BACCHI, M. R. P. Ethanol: From the beginning to the current stages of production. *Revista de Política Agrícola*, v. 4, oct–dec 2014.

MORETTIN, P.; TOLOI, C. *Analise de séries temporais*. 2nd. ed. São Paulo, Brazil: Editora Edgar Blucher, 2006.

National Agency of Petroleum, Natural Gas and Biofuels. *Fuels in Brazil: Price policies and tax structure*. 2001.

National Agency of Petroleum, Natural Gas and Biofuels. *Nota Técnica SDR/ANP nº 068/2018*. 2018.

National Agency of Petroleum, Natural Gas and Biofuels. *Historical Series of Price Surveys*. 2024. Accessed on: August 15, 2023. Available on: <https://www.gov.br/anp/pt-br/assuntos/precos-e-defesa-da-concorrencia/precos/precos-revenda-e-de-distribuicao-combustiveis/serie-historica-do-levantamento-de-precos>.

National Association of Automobile Manufacturers. *History of the Automotive Industry*. 2024. Accessed on: May 26, 2025. Available on: <https://anfavea.com.br/site/historia-da-industria-automotiva/>.

National Association of Automobile Manufacturers. *What Was the Proálcool Program?* 2024. Accessed on: January 20, 2024. Available on: <https://anfavea.com.br/site/o-que-foi-o-proalcool/>.

National Association of Railway Transporters. *General Information*. 2024. Accessed on: September 30, 2024. Available on: <https://www.antf.org.br/informacoes-gerais/>.

National Transport Confederation. *Pesquisa CNT – Perfil Empresarial*. 2021. <https://www.cnt.org.br/painel-empresarial>. Acesso em: 30 set. 2023.

NEAL, R. M. Mcmc using hamiltonian dynamics. In: BROOKS, S.; GELMAN, A.; JONES, G. L.; MENG, X. L. (Ed.). *Handbook of Markov Chain Monte Carlo*. Boca Raton, FL: Chapman & Hall/CRC, 2011. p. 113–162.

Official Gazette of the Union. *Decree No. 10.638, No. 39-A*. 2021. Accessed on: May 26, 2025. Available on: <https://pesquisa.in.gov.br/imprensa/jsp/visualiza/index.jsp?data=01/03/2021&jornal=600&pagina=2&totalArquivos=2>.

Official Gazette of the Union. *Complementary Law No. 194, Year CLX, No. 117-B, Brasília*. 2022. Accessed on: May 26, 2025. Available on: <https://pesquisa.in.gov.br/imprensa/jsp/visualiza/index.jsp?data=23/06/2022&jornal=601&pagina=1&totalArquivos=3>.

PEI, Y.; HUANG, C.-J.; SHEN, Y.; WANG, M. A novel model for spot price forecast of natural gas based on temporal convolutional network. *Energies*, v. 16, p. 2321, 2023.

PINCHEIRA-BROWN, P.; BENTANCOR, A.; HARDY, N.; JARSUN, N. Forecasting fuel prices with the chilean exchange rate: Going beyond the commodity currency hypothesis. *Energy Economics*, v. 106, p. 105802, 2022.

QUEIROZ, E.; ALVES, K.; OLIVEIRA, F.; AGUIAR, E. Variable step size evolving participatory learning with kernel recursive least squares applied to gas prices forecasting in brazil. *Evolving Systems*, v. 13, p. 297–306, 2022.

RAHMAN, S. Another perspective on gasoline price responses to crude oil price changes. *Energy Economics*, v. 55, p. 10–18, 2016.

RAST, P.; MARTIN, S. R.; LIU, S.; WILLIAMS, D. R. *A New Frontier for Studying Within-Person Variability: Bayesian Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models.* : American Psychological Association, 2020.

RIZVI, S. M. H.; SYED, T.; QURESHI, J. Real-time forecasting of petrol retail using dilated causal cnns. *Journal of Ambient Intelligence and Humanized Computing*, v. 13, p. 989–1000, 2022.

SAFARI, A.; DAVALLOU, M. Oil price forecasting using a hybrid model. *Energy*, v. 148, p. 49–58, 2018.

SAMADI, S. Y.; ALWIS, T. P. D. Envelope matrix autoregressive models. *Journal of Business & Economic Statistics*, p. 1–16, 2025. Available on: <https://doi.org/10.1080/07350015.2025.2537404>.

SAMADI, S. Y.; BILLARD, L. On a matrix-valued autoregressive model. *Journal of Time Series Analysis*, v. 46, n. 1, p. 3–32, 2025.

SANTOS, G. C.; BARBOZA, F.; VEIGA, A. C. P.; SILVA, M. F. Forecasting brazilian ethanol spot prices using lstm. *Energies*, v. 14, p. 7987, 2021.

SILVA, C. Análise da previsão do preço do etanol hidratado no estado de são paulo: uma aplicação do modelo arima. *Brazilian Journal of Development*, v. 5, n. 10, p. 17763–17778, 2019.

SILVA, L. L.; CHAVES, E. P. S. Análise da variação dos preços da cesta básica antes e depois da greve dos caminhoneiros no brasil em 2018. *Revista de Empreendedorismo e Gestão de Micro e Pequenas Empresas*, v. 5, n. 1, p. 1–15, 2020.

SIMS, C. A. Macroeconomics and reality. *Econometrica*, v. 48, n. 1, p. 1–48, jan. 1980.

Sindicato Nacional das Empresas Distribuidoras de Gás Liquefeito de Petróleo . *Utilidades do LPG: Industry – Energy in Factories.* 2023. Magazine, Volume 4. Accessed: 25 Nov. 2025. Available on: <sandbox:/mnt/data/58e34e78-3e00-4e20-9c49-a079c5d8ee78.png>.

Statista. *Oil consumption share worldwide in 2010 and 2024, by country.* 2024. <https://www.statista.com/statistics/>. Accessed November 25, 2025. Screenshot available at </mnt/data/c94561f2-8568-4464-8044-3b12a60dca39.png>.

STEFANAKOS, C. N.; SCHINAS, O. Forecasting bunker prices: A nonstationary, multivariate methodology. *Transportation Research Part C*, v. 38, p. 177–194, 2014.

SU, M.; ZHANG, Z.; ZHU, Y.; ZHA, D. Data-driven natural gas spot price forecasting with least squares regression boosting algorithm. *Energies*, v. 12, p. 1094, 2019.

The Global Economy. *Gasoline Consumption - Country Rankings, 2022.* 2022. <https://www.theglobaleconomy.com/>. Accessed November 25, 2025. Data in thousands of barrels per day; Source: World Bank.

The Global Economy. *Liquefied Petroleum Gas (LPG) Consumption - Country Rankings, 2022*. 2022. <https://www.theglobaleconomy.com/>. Accessed November 25, 2025. Data in thousands of barrels per day; Source: World Bank.

TIPYAN, J.; LEENAWONG, C. Quantitative models for forecasting vehicle fuel prices in thailand. In: FACULTY OF SCIENCE, KING MONGKUT INSTITUTE OF TECHNOLOGY LADKRABANG. *2009 World Congress on Computer Science and Information Engineering*. Bangkok, Thailand, 2009.

WU, H.; LI, S. Volatility spillovers in china's crude oil, corn and fuel ethanol markets. *Energy Policy*, v. 62, p. 878–886, 2013.

YE, M.; ZYREN, J.; SHORE, J. A monthly crude oil spot price forecasting model using relative inventories. *International Journal of Forecasting*, v. 21, p. 491–501, 2005.

YOSTHONGNGAM, S.; TANSUCHAT, R.; YAMAKA, W. Volatility spillovers between ethanol and corn prices: A bayesian analysis. *Energy Reports*, v. 8, p. 1030–1037, 2022.

ZHANG, G.; PATUWO, B. E.; HU, M. Y. Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, v. 24, p. 35–62, 1998.

ZHANG, G. P. Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, v. 50, p. 159–175, 2003.

APPENDIX A – Tables A

1.1 Appendix

Table A.1 – Parameters of the Gasoline Forecasting Models

Series	SSA (L, r)	ARIMA	ES (α, β)
SP	(72,1-6)	(0,1,1) (0,0,0)	(1;0.0303)
RJ	(72,1-6)	(1,1,0) (0,0,0)	(1;0.0627)
MG	(72,1-7)	(1,1,0) (0,0,0)	(1;0.0627)
ES	(72,1-8)	(0,1,1) (0,0,0)	(1;0.0066)
PR	(72,1-7)	(0,1,1) (0,0,0)	(1;0.0000)
SC	(72,1-7)	(0,1,1) (0,0,0)	(1;0.0268)
RS	(72,1-5)	(0,1,1) (0,0,0)	(1;0.0451)
MT	(72,1-7)	(1,1,0) (0,0,0)	(1;0.0487)
MS	(72,1-5)	(0,1,1) (0,0,0)	(1;0.0640)
GO	(72,1-12)	(0,1,1) (0,0,0)	(1;0.0407)
DF	(72,1-7)	(0,1,1) (0,0,0)	(1;0.0137)
AM	(72,1-7)	(0,1,1) (0,0,0)	(1;0.0375)
PA	(72,1-5)	(0,1,3) (0,0,0)	(1;0.0421)
RR	(60,1-6)	(1,1,0) (0,0,0)	(1;0.0261)
AP	(72,1-12)	(0,1,3) (0,0,0)	(1;0.0244)
RO	(72,1-6)	(2,1,1) (0,0,0)	(1;0.0621)
AC	(72,1-5)	(2,1,2) (0,0,0)	(1;0.1900)
TO	(72,1-5)	(0,1,1) (0,0,0)	(1;0.0182)
PB	(72,1-7)	(1,1,2) (0,0,0)	(1;0.0000)
BA	(72,1-7)	(0,1,1) (0,0,0)	(1;0.0271)
MA	(72,1-7)	(0,1,3) (0,0,0)	(1;0.0452)
PI	(72,1-9)	(0,1,1) (0,0,0)	(1;0.0117)
PE	(72,1-5)	(0,1,1) (0,0,0)	(0.8591;0)
AL	(72,1-5)	(3,1,0) (0,0,0)	(1;0.0354)
SE	(72,1-6)	(3,1,0) (0,0,0)	(1;0.0336)
RN	(72,1-7)	(0,1,1) (0,0,0)	(1;0.0270)
CE	(72,1-6)	(0,1,1) (0,0,0)	(1;0.0203)

Source: prepared based on the research results. All states estimated with the ES model were considered to have additive seasonality. L is the window length and r is the number of eigenvalues.

Table A.2 – Parameters of the diesel Forecasting Models

Series	SSA (L, r)	ARIMA	ES (α, β, γ)
SP	(80,1-8)	(1,1,0) (0,0,0)	(1;0.0455)
RJ	(80,1-8)	(1,1,0) (0,0,0)	(1;0.0505)
MG	(80,1-8)	(1,1,0) (0,0,0)	(1;0.0610)
ES	(80,1-9)	(1,1,0) (0,0,0)	(1;0.0483)
PR	(80,1-9)	(1,1,0) (0,0,0)	(1;0.0495)
SC	(80,1-9)	(1,1,0) (0,0,0)	(1;0.0589)
RS	(72,1-9)	(1,1,0) (0,0,0)	(1;0.0537)
MT	(80,1-8)	(1,1,0) (0,0,0)	(1;0.0640)
MS	(80,1-8)	(1,1,0) (0,0,0)	(1;0.0635)
GO	(80,1-8)	(1,1,0) (0,0,0)	(1;0.0543)
DF	(72,1-10)	(1,1,0) (0,0,0)	(1;0.0474)
AM	(80,1-8)	(1,1,0) (0,0,0)	(1;0.0847)
PA	-	-	-
RR	(60,1-7)	(1,1,0) (0,0,0)	(1;0.2284)
AP	(80,1-8)	(1,1,0) (0,0,0)	(1;0.2030)
RO	(80,1-8)	(1,1,0) (0,0,0)	(1;0.0686)
AC	(80,1-7)	(1,1,0) (0,0,0)	(1;0.0648)
TO	(80,1-10)	(1,1,0) (0,0,0)	(1;0.2054)
PB	(60,1-8)	(1,1,0) (0,0,0)	(1;0.0479)
BA	(80,1-10)	(1,1,0) (0,0,0)	(1;0.0552)
MA	(80,1-9)	(1,1,0) (0,0,0)	(1;0.0620)
PI	(60,1-7)	(0,1,1) (0,0,0)	(1;0.0602)
PE	(80,1-9)	(1,1,0) (0,0,0)	(1;0.0556)
AL	(80,1-8)	(1,1,0) (0,0,0)	(1;0.0671)
SE	(80,1-9)	(1,1,0) (0,0,0)	(1;0.0461)
RN	(72,1-8)	(1,1,0) (0,0,0)	(1;0.0557)
CE	(80,1-9)	(1,1,0) (0,0,0)	(1;0.0668)

Source: prepared based on the research results. All states estimated with the AE model were considered to have additive seasonality. L is the window length and r is the number of eigenvalues.

APPENDIX B – Tables B

Table B.1 – Matrix C Coefficients for Imported Gasoline

			Credible Interval			
Coefficient	Mean	Std. Dev.	Median	2.5%	97.5%	$\hat{R}$
C_{11}	0.03	0.02	0.02	0.00	0.08	1
C_{22}	51.54	15.32	52.29	20.69	80.29	1

Source: prepared based on the research results.

Table B.2 – Matrix A Coefficients for Imported Gasoline

			Credible Interval			
Coefficient	Mean	Std. Dev.	Median	2.5%	97.5%	$\hat{R}$
a_{11}	0.28	0.07	0.27	0.17	0.46	1
a_{12}	0.19	0.34	0.19	-0.51	0.87	1
a_{21}	0.00	0.01	0.00	-0.01	0.02	1
a_{22}	0.07	0.24	0.16	-0.31	0.42	1

Source: prepared based on the research results.

Table B.3 – Matrix B Coefficients for Imported Gasoline

			Credible Interval			
Coefficient	Mean	Std. Dev.	Median	2.5%	97.5%	$\hat{R}$
b_{11}	0.85	0.06	0.86	0.69	0.94	1
b_{12}	0.48	0.58	0.43	-0.49	1.74	1
b_{21}	-0.01	0.01	-0.01	-0.03	0.02	1
b_{22}	0.04	0.46	0.09	-0.79	0.75	1

Source: prepared based on the research results.

Table B.4 – Forecasting performance for the models considering four fuels and twenty-five Brazilian states, $\hat{u_1} = 2$ and $\hat{u_2} = 3$

Model	**NOP**	**MSE in sample**	**MSE out of sample**			
			h = 1	**h = 3**	**h = 6**	**h = 12**
VAR	15150	0.1085	0.1014	0.0943	0.0836	0.0621
MAR	1074	0.0867	0.0822	0.0764	0.0678	0.0503
EMAR	516	0.0855	0.0816	0.0758	0.0672	0.0498

Source: Elaborated by the author based on results by models.

Table B.5 – Forecasting performance for the models considering four fuels and twenty-five Brazilian states, $\hat{u_1} = 2$ and $\hat{u_2} = 4$

Model	**NOP**	**MSE in sample**	**MSE out of sample**			
			h = 1	**h = 3**	**h = 6**	**h = 12**
VAR	15150	0.1085	0.1014	0.0943	0.0836	0.0621
MAR	1074	0.0867	0.0822	0.0766	0.0677	0.0504
EMAR	541	0.0857	0.0816	0.0760	0.0672	0.0499

Source: Elaborated by the author based on results by models.

Table B.6 – Forecasting performance for the models considering four fuels and twenty-five Brazilian states, $\hat{u}_1 = 2$ and $\hat{u}_2 = 5$

Model	**NOP**	**MSE in sample**	**MSE out of sample**			
			h = 1	**h = 3**	**h = 6**	**h = 12**
VAR	15150	0.1085	0.1014	0.0943	0.0836	0.0621
MAR	1074	0.0867	0.0822	0.0766	0.0677	0.0504
EMAR	566	0.0857	0.0816	0.0759	0.0672	0.0499

Source: Elaborated by the author based on results by models.

APPENDIX C – Table C

Table C.1 – Forecasting performance for the models considering four fuels and twenty-five Brazilian states, $\hat{u}_1 = 2$ and $\hat{u}_2 = 3$

Model	**NOP**	**MSE in sample**	**MSFE out of sample**			
			h = 1	**h = 3**	**h = 6**	**h = 12**
VAR	15150	0.1085	0.1014	0.0943	0.0836	0.0621
MAR	1074	0.0867	0.0822	0.0764	0.0678	0.0503
EMAR	516	0.0855	0.0816	0.0758	0.0672	0.0498

Source: Elaborated by the author based on results by models.

Table C.2 – Forecasting performance for the models considering four fuels and twenty-five Brazilian states, $\hat{u}_1 = 2$ and $\hat{u}_2 = 4$

Model	**NOP**	**MSE in sample**	**MSFE out of sample**			
			h = 1	**h = 3**	**h = 6**	**h = 12**
VAR	15150	0.1085	0.1014	0.0943	0.0836	0.0621
MAR	1074	0.0867	0.0822	0.0766	0.0677	0.0504
EMAR	541	0.0857	0.0816	0.0760	0.0672	0.0499

Source: Elaborated by the author based on results by models.

Table C.3 – Forecasting performance for the models considering four fuels and twenty-five Brazilian states, $\hat{u}_1 = 2$ and $\hat{u}_2 = 5$

Model	**NOP**	**MSE in sample**	**MSFE out of sample**			
			h = 1	**h = 3**	**h = 6**	**h = 12**
VAR	15150	0.1085	0.1014	0.0943	0.0836	0.0621
MAR	1074	0.0867	0.0822	0.0766	0.0677	0.0504
EMAR	566	0.0857	0.0816	0.0759	0.0672	0.0499

Source: Elaborated by the author based on results by models.