

Article

# Application of Non-Parametric and Forecasting Models for the Sustainable Development of Energy Resources in Brazil

Gabriela Mayumi Saiki <sup>1,\*</sup>, André Luiz Marques Serrano <sup>1,\*</sup>, Gabriel Arquelau Pimenta Rodrigues <sup>1</sup>,  
Guilherme Dantas Bispo <sup>1</sup>, Vinícius Pereira Gonçalves <sup>1</sup>, Clóvis Neumann <sup>1</sup>,  
Robson de Oliveira Albuquerque <sup>1</sup> and Carlos Alberto Schuch Bork <sup>2</sup>

<sup>1</sup> Professional Post-Graduate Program in Electrical Engineering (PPEE), Department of Electrical Engineering (ENE), Technology Faculty, University of Brasilia (UnB), Brasilia 70910-900, Brazil; gabriel.arquelau@redes.unb.br (G.A.P.R.); guilherme.bispo@redes.unb.br (G.D.B.); vpgvinicius@unb.br (V.P.G.); clovisneumann@unb.br (C.N.); robson@redes.unb.br (R.d.O.A.)  
<sup>2</sup> Brazilian National Confederation of Industry (CNI), Brasilia 70040-903, Brazil; carlos.bork@cni.com.br  
\* Correspondence: mayumisaiiki.gms@gmail.com (G.M.S.); andrelms@unb.br (A.L.M.S.)

**Abstract:** To achieve Sustainable Development Goal 7 (SDG7) and improve energy management efficiency, it is essential to develop models and methods to forecast and enhance the process accurately. These tools are crucial in shaping the national policymakers' strategies and planning decisions. This study utilizes data envelopment analysis (DEA) and bootstrap computational methods to evaluate Brazil's energy efficiency from 2004 to 2023. Additionally, it compares seasonal autoregressive integrated moving average (SARIMA) models and autoregressive integrated moving average (ARIMA) forecasting models to predict the variables' trends for 2030. One significant contribution of this study is the development of a methodology to assess Brazil's energy efficiency, considering environmental and economic factors to formulate results. These results can help create policies to make SDG7 a reality and advance Brazil's energy strategies. According to the study results, the annual energy consumption rate is projected to increase by an average of 2.1% by 2030, which is accompanied by a trend of GDP growth. By utilizing existing technologies in the country, it is possible to reduce electricity consumption costs by an average of 30.58% while still maintaining the same GDP value. This demonstrates that sustainable development and adopting alternatives to minimize the increase in energy consumption can substantially impact Brazil's energy sector, improving process efficiency and the profitability of the Brazilian industry.

**Keywords:** data envelopment analysis (DEA); energy; forecasting; supply; time series



**Citation:** Saiki, G.M.; Serrano, A.L.M.; Rodrigues, G.A.P.; Bispo, G.D.; Gonçalves, V.P.; Neumann, C.; Albuquerque, R.d.O.; Bork, C.A.S. Application of Non-Parametric and Forecasting Models for the Sustainable Development of Energy Resources in Brazil. *Resources* **2024**, *13*, 150. <https://doi.org/10.3390/resources13110150>

Academic Editor: Eva Pongrácz

Received: 8 September 2024

Revised: 10 October 2024

Accepted: 16 October 2024

Published: 23 October 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Integrating renewable energy sources is a global trend in energy distribution systems. It enables them to (i) address the unresolved energy challenges posed by traditional centralized power plants, (ii) reduce global CO<sub>2</sub> emissions, and (iii) increase the long-term supply of sustainable energy [1]. Distributed energy resources (DERs), particularly wind and photovoltaic solar energy, play an increasingly important role in the energy sector [2].

In Brazil, a large part of the energy matrix comprises renewable sources, especially energy of hydraulic origin. In 2022, hydraulic sources were responsible for the production of 427.1 TWh, representing around 61.9% of the total production in the country, while photovoltaic solar energy sources had a production value of 30.1 TWh (4.4%), thermal solar energy of 11.6 TWh (1.7%), and wind energy source of 81.6 TWh (11.8%). Of all end consumers, the industry is responsible for electricity consumption, which was 218.7 TWh (31.7%) in 2022. If operational losses are ignored, the industry, in general, is responsible for the consumption of 37.3% of the country's electrical energy [3]. This shows that making the energy matrix efficient encompasses sustainable factors to achieve SDG 7 and economic factors to make the Brazilian industry more competitive internationally.

Understanding the demand for electrical energy is crucial to efficiently allocate and manage energy resources within a system with limited production capacity. This understanding can also help to achieve sustainable development goals, particularly Sustainable Development Goal 7 (SDG 7) and related goals. This goal ensures everyone has reliable, sustainable, modern, and affordable access to energy. It is important to note that energy represents a significant economic cost for a country and must be considered when measuring the gross domestic product (GDP). This highlights the importance of efficient management in this sector [4].

Sustainable Development Goals encompass a wide range of targets and guidelines, which include the following: ensuring universal access to modern, viable, and affordable energy services (7.1), increasing the share of renewable energy in the global energy mix (7.2), doubling the global rate of improvement in energy efficiency (7.3), strengthening measures of international cooperation for clean energy research and technology (7.4), and improving the infrastructure for sustainable energy services in developing countries (7.5) [5]. In addition, Table A1 (Appendix A) shows Brazil's action to reach these targets. An understanding of energy production, as shown in the various matrices that comprise the national electricity grid of Brazil, as well as an ability to forecast future energy demand, are critical features for public services, policymakers, and stakeholders, as they allow them to make informed decisions about sustainable development objectives [6].

The use of forecasting techniques has attracted significant interest in the literature and has been explored in several contexts. These include their practical applicability, effectiveness in different industrial sectors, suitability to assist organizations of various sizes, use for temporal horizons, and degrees of precision [7].

Considering this, it is important to understand how the country is progressing toward achieving sustainable development goals. However, the lack of planning and knowledge about the current state of the energy sector is a gap in the literature on the development of the Brazilian industry. Therefore, the study aims to (i) measure the efficiency of energy resources in Brazil concerning the country's gross domestic product (GDP), (ii) compare the suitability of ARIMA and SARIMA models for forecasting energy demand and GDP using the Akaike and Bayesian information criteria and check the models' forecast error metrics (RMSE, MAE, MAPE, MASE), (iii) verify the time-series trends in a forecast scenario up to 2030, and (iv) propose a step-by-step approach to analyzing the Brazilian energy scenario to support the proposal of public policies in line with SDG7, serving as a basis for strategic planning.

The study aims to utilize data envelopment analysis (DEA) to analyze the energy efficiency in Brazil and identify opportunities for enhancing efficiency in the country's energy resources. It also seeks to compare the performance of ARIMA and SARIMA models using the structured dataset to determine the most suitable model for this case study. In addition, the study aims to utilize the chosen model to make predictions and provide valuable information on the future behavior of the variables. Furthermore, the study proposes a methodology for evaluating Brazilian energy efficiency and supporting strategic planning in achieving Sustainable Development Goal 7.

This study proposes a methodology to assess a country's energy efficiency, combining environmental and economic factors. We apply this methodology to the Brazilian context, thus revealing the energy characteristics of a developing country with abundant natural resources.

The study employs data envelopment analysis and bootstrap computational methods to evaluate energy usage efficiency statistically, providing valuable findings for decision-makers and the academic community. Also, the results provide a basis for future research, which can further explore alternative methodologies and expand the analysis to other regions or sectors.

Additionally, it contributes to the forecasting domain by comparing SARIMA and ARIMA models to predict future trends, projecting a 2.1% annual increase in energy consumption by 2030. The study also identifies a potential 30.58% reduction in electricity

consumption costs through existing technologies without compromising GDP growth, emphasizing the role of sustainable practices in advancing Brazil's energy strategies. These findings have significant implications for achieving Sustainable Development Goal 7.

The structure of the paper is organized as follows. Section 2 reviews the literature on time-series forecasting models and their applicability in energy resources. This section gathers the most recent studies on the subject and critically analyzes the findings. The following Section 3 will provide an overview of the materials and methods used in this study, detailing the methodology utilized, and explaining the structure of Sections 4–7. These sections offer a detailed explanation of the mathematical models and procedures applied, including the implementation of the bootstrap-DEA and the use of ARIMA and SARIMA forecasts to analyze the energy resources in Brazil. Section 8 discusses the main findings and results of our application of bootstrap-DEA. The calculations of the efficiency index and their respective 95% confidence intervals are also presented. In addition, the application of the SARIMA model is discussed to forecast the time series of energy demand and gross domestic product (GDP) in Brazil. Finally, Section 9 concludes the paper by summarizing the main contributions of the research.

## 2. Literature Review

The need to develop a sustainable energy plan and the emphasis placed on forecasting energy demand emphasize the importance of effectively managing this demand, which is aligned with the long-term objectives of the environmental, social, and economic sustainability goals [8]. By integrating energy demand forecasting into sustainable energy planning, societies can move toward a future that is both sustainable and equitable—combining knowledge about productivity, sustainability, and industry [9].

Therefore, forecasting energy demand is fundamental for energy planning and management, enabling stakeholders to make informed decisions regarding resource allocation, infrastructure development, policy formulation, and its impact on GDP. Energy demand forecasting methods can be broadly categorized into qualitative and quantitative approaches. As documented by [10], time-series models are widely used for energy demand forecasting because they capture the temporal patterns and trends inherent in energy consumption data.

Table 1 presents a comparison between this paper and previous works. This work updates the data envelopment analysis on Brazil's energy consumption by incorporating more recent data and advanced forecasting techniques. While a previous study from 2012 applied DEA [11], and another from 2024 utilized SARIMA for forecasting [8], our research combines both DEA and SARIMA models to provide a more comprehensive analysis of energy efficiency and future consumption trends. Additionally, we update a 2022 study that applied DEA using data from 2015 [12].

**Table 1.** Comparison between works that use DEA or ARIMA/SARIMA applied to energy consumption.

Reference	Year	Targeted Country	ARIMA/SARIMA	DEA
[11]	2012	Brazil	✗	✓
[13]	2015	BRICS	✗	✓
[14]	2022	India	✓	✗
[8]	2024	Brazil	✓	✗
This work	2024	Brazil	✓	✓

### 2.1. Predictive Models Using Machine Learning Techniques

Several studies have compared the performance of time-series models in different contexts and for various applications. A comprehensive review of forecasting methods, including classical and modern approaches, was conducted [15]. The study showed that the most generic deep learning models, such as LSTM and GRU, performed better than simpler models, such as DeepAR and DeepState. However, these gains were not significant, and the result is worse when looking at more extended scenarios. Similarly, Ref. [16] used

a non-linear autoregressive neural network (NAR) to predict the next decade of energy demand based on a publicly available dataset for global energy consumption; however, the model that performed best (FB prophet) is challenging to apply [8].

Energy demand forecasting has been a significant area of research with several methods proposed for accurate forecasting. Neural networks, such as non-linear autoregressive neural networks, can deal with statistical, empirical and theoretical problems effectively, as presented by [17]; however, similarly to the study of [16], the best forecasting model, which in this case will be NAR, also proves to have limitations in terms of application due to the complexity of the model, resulting in various errors. Convolution neural networks (CNNs) and conditional random fields (CRFs) have been used to predict energy consumption, achieving high accuracy, according to [18], although it is highly accurate that in a real scenario where there is a vast amount of data, a considerable amount of GPUs will be needed to run the models. The forecast of electricity prices is also relevant in this context [19].

Developing countries face challenges in achieving sustainable economic growth with low carbon emissions. Still, for [10], hybridized artificial neural networks (ANNs) with metaheuristic techniques are superior at load prediction, although they are naturally more complex. Probabilistic forecasting models, such as DeepAR and deep state space, perform better for longer forecast horizons, as conducted by [15].

Machine learning (ML) techniques are widely used, especially for short-term electricity forecasts, while engineering-based models cover long time horizons and household appliance consumption; according to [20], essential issues raised by the study point to the need to have industrial energy demand models, as well as the fact that different models perform differently when applied to other scenarios. In power generation, forecasting models have also been used to predict electricity supply and production parameters, such as wind speed [21]. However, this approach is interesting to be applied in scenarios considering wind energy; it does not help evaluate an energy matrix with a predominance of hydro and solar generation, as in Brazil [8].

## 2.2. Econometric Models: ARIMA and SARIMA

In this way, it is understood that the most recent studies dealing with energy demand forecasting have focused on using machine learning techniques. Despite presenting good results compared to more conventional methods, they have come up against the fact that they are complex for practical application. Therefore, traditional models found in the literature, like ARIMA and SARIMA, are more suitable for countries such as Brazil because they are simple to apply and interpret.

In contrast, the ARIMA model is based on adjusting the observed values to reduce the difference between the values produced in the model and the observed values to near zero. For the construction of ARIMA models, Box–Jenkins suggested the following iterative steps; according to Bayer, the ARIMA model  $(p, d, q)$  consists of combining an autoregressive (AR) model of order  $p$  with a time series differenced  $d$  times (number of differences needed to make the series stationary) and a moving average (MA) model of order  $q$ . Thus, it has the following form:  $\epsilon$  is white noise;  $f(B)$  and  $q(B)$  are the autoregressive polynomial and moving average polynomial, respectively. The seasonal ARIMA model, or SARIMA, follows this presentation format: SARIMA  $(p, d, q)(P, D, Q)$ , where the parameters  $(P, D, Q)$  are the seasonal equivalents of  $(p, d, q)$ . The seasonal part is represented by three additional similar parameters; thus, it is called SARIMA. The SARIMA model is defined by a simple part of parameters  $(p, d, q)$  and a seasonal part composed of parameters  $(P, D, Q)$  [22]. Table 2 presents a brief conceptual comparison between these two models [23].

**Table 2.** Strengths and limitations of predicting models.

Model	Strengths	Limitations	Gaps in the Literature
ARIMA	Enables the measurement of promising results in the forecasted values of the variable ( $y_t$ ). It outperforms the autoregressive and moving average (ARMA) model. Considers autoregressive (AR), integrated (I), and moving average (MA) terms to account for lags in non-stationary time series.	ARIMA models will only deal with non-stationary time series through differentiation. Does not consider seasonal factors in time series.	Explain the future with past knowledge and are subject to inaccuracies caused by events outside the norm, which have a strong impact on the values ( $y_t$ ).
SARIMA	Considers not only the previous period for future forecasting but also a seasonality term ( $s$ ). Considers autoregressive (AR), integrated (I), and moving average (MA) terms with seasonality ( $P, D, Q$ ). It outperforms ARIMA when seasonality is present.	Challenges with irregular seasonal components.	Instability of seasonal measures of supply and demand, as the concept of seasonality assumes that the seasonality term ( $s$ ) is standardized and represents a specific time period.

### 2.3. Comparative of Machine Learning and Classical Econometric Forecasting Models

Several studies have thoroughly explored the conceptual and technical aspects of weather forecasting models, such as ARIMA and SARIMA, highlighting their reliability and robustness across various forecasting scenarios [24]. These models are beneficial for time series analysis, as they can effectively model temporal dependencies and seasonality. By capturing patterns in historical data, they provide accurate short- and medium-term forecasts. Researchers have applied these models to various weather-related datasets, validating their performance in predicting temperature, precipitation, and other meteorological variables [25]. In addition, these models are known for their interpretability, making them accessible and widely used in academic research and practical forecasting applications [26]. Furthermore, they are a benchmark for comparing newer machine-learning techniques in weather forecasting.

Many studies compared traditional and deep learning models. Some studies have indicated the superiority of conventional statistical models over artificial intelligence (AI), which states that traditional models have interpretable parameters and vast theoretical and empirical literature and are usually suitable for smaller datasets with relatively simple relationships [24]. They do not require large volumes of data to generate effective predictions, and finally, there is an ease of diagnosis and adjustments that facilitate model validation. In contrast, others have demonstrated the better performance of AI-based forecasting models, which can capture complex non-linear patterns and interactions between variables that traditional models cannot identify [27]. These models are also suitable for working with large volumes of data (big data), including structured and unstructured data, such as images, text, or audio, which makes them ideal for modern applications in fields such as marketing, finance, and demand forecasting [28]. Their high precision in predicting different values also stands out [29].

Several studies have demonstrated that there are no significant performance differences between traditional statistical models, such as ARIMA and SARIMA, and AI-based models in certain forecasting scenarios [30]. This indicates that despite the advancements in machine learning techniques, traditional models continue to perform competitively, particularly in structured data environments with well-defined temporal patterns [31]. Furthermore, the choice between these approaches often depends on the specific characteristics of the dataset and the complexity of the relationships between variables [31].

Section 3 delves deeper into the mathematical frameworks underpinning these models, providing detailed explanations of the statistical tests and assumptions involved. It also elaborates on how ARIMA and SARIMA models are constructed and applied to time series data, focusing on their capacity to model seasonal and non-seasonal components. The section compares these traditional methods with more recent AI-driven approaches,

evaluating their effectiveness in various forecasting contexts and providing insights into their advantages and limitations.

### 3. Materials and Methods

This section outlines the methodology employed for this research, detailing the statistical tools and the dataset used. The structure used to implement the research methodology is illustrated in Figure 1.

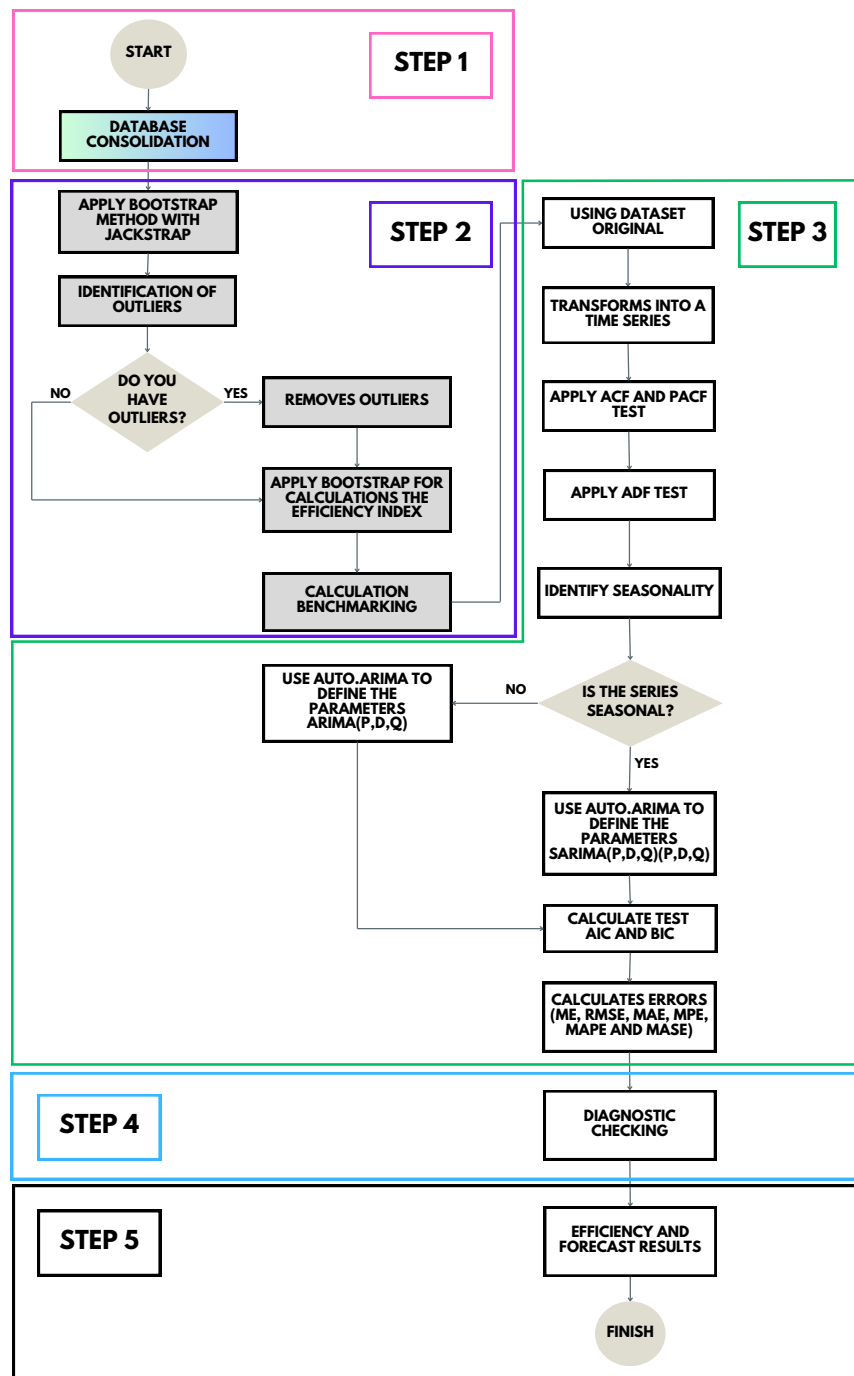


Figure 1. Structure for executing the research.

Figure 1 presents the methodology used during the research.

The first step (Step 1) involves consolidating the database, which includes pre-processing and structuring the data for subsequent analysis. The information is obtained from two primary sources: electricity demand data, provided by the Energy Research Company (EPE) [3], and GDP data, provided by the Institute of Applied Economic Research (IPEA). The dataset covers the period from 2004 to 2023 with monthly granularity. Section 4 describes the complete database consolidation process.

The second step (Step 2) aims to measure Brazil's energy efficiency between 2004 and 2023. Data envelopment analysis (DEA) is used, which is a non-parametric methodology commonly applied to measure the relative efficiency of decision-making units (DMUs) based on multiple inputs and outputs [32].

Thus, data envelopment analysis (DEA) compares the efficiency of DMUs using inputs and outputs [33]. The steps and principles adopted in DEA are as follows. First, selecting input and output variables is crucial to assessing efficiency. Typical inputs may include resources such as energy consumed or capital invested, while outputs may be the value of output or service generated. These indicators are selected according to their relevance to the energy sector and economy. Next, the Efficiency Frontier is constructed, in which the DEA model identifies the most efficient DMUs and constructs an "efficiency frontier", representing the benchmark for best practices [34]. DMUs operating on the frontier are considered 100% efficient (score 1), while those below the frontier receive scores less than 1, indicating inefficiency. Next, the Relative Efficiency is calculated, in which the efficiency of each DMU is calculated by solving a series of linear programming models, which seek to maximize the ratio between outputs and inputs. DMUs that reach the frontier are considered efficient, and those that fall below are compared with efficient units to identify room for improvement. Finally, there is identification of inefficiencies, which will occur after calculating the efficiency scores: the model allows the identification of the specific causes of inefficiency in each DMU [35]. Based on this, suggestions can be made for adjustments in inputs or outputs, allowing DMUs to improve their performance.

To increase the robustness of the analysis and ensure that the results are not affected by outliers, Jackknife [36] is used, which is a statistical technique that consists of recalculating the efficiency by iteratively removing one DMU at a time, helping to identify outliers that may disproportionately influence the results. After applying DEA, the bootstrap method is used to improve the reliability of the results and estimate confidence intervals for the efficiency scores. Bootstrapping is a resampling technique widely used to assess the variability of estimators. In bootstrapping, a series of new samples (simulations) are created from the original sample of DMUs with replacement. This means some DMUs may appear more than once in a sample, while others may be excluded. This process is repeated thousands of times (1000 or 10,000 times, for example), generating efficiency distributions for each DMU.

By applying DEA to each bootstrap sample generated via resampling, DEA is reapplied to calculate the corresponding efficiency scores, resulting in a distribution of efficiency scores for each DMU rather than a single value. Confidence interval construction is also performed, which takes these efficiency score distributions and constructs confidence intervals for the original scores, allowing the uncertainty associated with the efficiency estimates to be quantified, providing a more accurate and robust assessment of the performance of the DMUs. Confidence intervals indicate the extent to which the efficiency scores vary, providing greater confidence in interpreting the results. Thus, applying bootstrapping makes it possible to check the stability of the efficiency scores. DMUs with large bootstrap score variations indicate low confidence in the original efficiency estimate, while smaller variations suggest that the efficiency is robust. These details about DEA and the bootstrap process help ensure the methodology is transparent and replicable, allowing other researchers to reproduce the analysis and check the robustness of the results. The steps and rationale behind applying these techniques are described in Section 5.

The third step (Step 3) applies the first two phases of the Box–Jenkins method to time series modeling and energy demand forecasting. This involves identifying the parameters

of an ARIMA (autoregressive integrated moving average) model using the autocorrelation function (ACF) and partial autocorrelation function (PACF), along with stationarity tests such as ADF, KPSS, and Phillips–Perron. The methodology is described in Section 6.

The fourth step (Step 4) involves checking the ARIMA modeling results, ensuring that the identified parameters are adequate. If not, the parameter identification process is repeated until the results are acceptable. This analysis includes the evaluation of the model residuals, as described in Section 7.

Finally, the fifth step (Step 5) presents and discusses the results of the energy efficiency indices and the chosen prediction model. The conclusions are presented in Section 8, emphasizing the practical application of the results.

#### 4. Database Consolidation

This section details Step 1 of the methodology shown in Figure 1.

The research database comprises 240 rows with two columns representing variables (demand and GDP) for each month from 2004 to 2023.

For the efficiency analysis, consider the value of  $x_1$ , which represents energy demand as input, and  $y_1$ , which means the gross domestic product as output.

Similarly, energy demand and gross domestic product are also considered in the forecast analysis. To evaluate whether the variables are subject to seasonal behavior during forecasts, the values would have to be every month, which is why monthly values were chosen. The complete database is available on figshare [37].

Table 3 shows the statistical summaries of the variables in the study, including the minimum, 1st quartile, median, mean, 3rd quartile, and maximum for demand and GDP.

**Table 3.** Statistical summary of electricity supply and demand and GDP. The considered time range is between 2004 and 2023.

	Demand (GWh)	GDP (Millions of R\$)
Min	26,508	142,861
1st Qu.	32,234	267,691
Median	37,866	459,337
mean	36,622	457,192
3rd Qu	40,269	582,831
Max	46,407	954,063

The R 4.3.3 software was used for the database structure, outlier removal, efficiency calculation, and forecasting. In this way, R software offered several advantages for conducting simulations and data analysis, making it an ideal tool for complex research. First, R has a wide range of specialized packages, such as *deaR* and *forecast*, which facilitate the implementation of advanced methodologies, such as data envelopment analysis (DEA) and ARIMA/SARIMA modeling, with high flexibility and accuracy. In addition, the ability to perform bootstrap resampling with the *boot* package ensures the robustness of the results, generating reliable confidence intervals. Another important advantage is the ease of automation and reproducibility through scripts and automatic reports with R Markdown, which increases the transparency and credibility of the research. R also excels in data visualization with packages such as *ggplot2*, allowing effective communication of results. Finally, its computational efficiency, support for big data, and ability to integrate with statistical methods and machine learning make R a powerful and versatile tool for large-scale simulations.

#### 5. Estimation of Energy Efficiency

This section explains the DEA methodology, the outlier removal procedures, and the calculation of bootstrap confidence intervals for the energy efficiency index. Therefore, this section presents the methods and procedures related to Step 2, as shown in Figure 1.



### 5.1. Data Envelopment Analysis (DEA)

Data envelopment analysis (DEA) is a mathematical model based on econometrics and operations research. The model originates from the productivity concepts introduced by [38], where productivity is defined as the relationship between products and inputs of a specific production process. Previous measures and metrics were overly specific and impractical, especially when dealing with numerous variables. Building upon this productivity concept, Ref. [39] proposed the relative efficiency model, aiming to optimally utilize both outputs and inputs while considering the constant returns to scale of the decision-making units ( $DMU_i$ ) through mathematical programming problems [40].  $DMU_i$  refers to the units being compared, such as hospitals, schools, organizations, companies, and countries, among others [41].

The DEA model is a non-parametric approach, which does not require specific knowledge of the problem's nature. This inherent flexibility makes it more applicable and robust than parametric models like stochastic frontier analysis (SFA), simplifying its application in various scenarios. This advantage is particularly evident in the business sector, where non-parametric models outperform parametric ones. As a result, the DEA model has been widely applied in the literature in different fields, including industrial [42], energy [43], economic [44], mining, environmental, water industry [45], public policies [46], and agribusiness [47].

Measuring energy productivity in underdeveloped regions can be challenging due to various factors, such as limited data availability. This study, however, uses data published by governmental institutions within Brazil to conduct the analysis.

### 5.2. Bootstrap Method for Outlier Removal

Outliers are data points that are considered abnormal within a dataset because they significantly differ from the other observations. Several factors can cause this behavior. The main factors identified as causing outliers in the data are (i) errors in the database that occur during data collection and organization of the data frame, (ii) accurate and highly atypical data, and (iii) actual and deficient data. Depending on the model's sensitivity to outliers, these factors can cause significant distortions and bias the results, making their removal necessary [41].

The presence of outliers in the DEA model creates a problematic situation when dealing with the frontier of a set of efficient production units, as deterministic measures and values are sensitive to errors. Therefore, removing outliers is necessary to avoid biasing the analysis and to produce more robust results for the model [32].

One approach is to use the Jack-knife leverage technique [36]. This technique assesses the impact of removing each data point from the dataset. In this application, the effect of the removed data on the efficiency values of other  $DMU_i$  is observed in combination with the bootstrap resampling technique to remove outliers.

The procedure is as follows:

- i Calculate the efficiency scores of all  $DMU_i$  by the classical model [39], generating a set of efficiencies given by  $\{\theta_i \mid i = 1, \dots, n\}$ ;
- ii Randomly select a subset  $K$  with  $(K = 1, \dots, k)$  corresponding to 10% of the original sample of  $DMU_i$ , which results in a set of  $DMU_K$ ;
- iii Calculate the efficiency scores of all selected  $DMU_K \{\theta_K \mid K = 1, \dots, k\}$  by bootstrap resampling  $B$  times, where  $B$  takes values from  $(B = 1, \dots, b)$ ;
- iv Assesses the impact using a statistical measure to analyze if there were significant changes in the efficiency scores through the leverage of each selected  $DMU_K$  in  $B$ , storing the leverage information in  $l_{KS}$ ;
- v Repeat Steps 2, 3, and 4  $S$  times with  $(S = 1, \dots, s)$ ;
- vi Calculate the local leverage  $l_K$  from the sum of the leverages of  $l_{KS}$  divided by  $n_K$ , which corresponds to approximately  $n_K \approx \frac{S \times B}{K}$ ;

- vii Calculate the global leverage through the standard deviation of the efficiency measures before and after removing the data; for more details on calculating leverage, please refer to the provided reference [36].

Using leverage minimizes the probability of selecting outliers for random resampling. This probability is determined using the Heaviside function, which is used in outlier removal based on a step function that assumes binary values (0 or 1). This function is used to identify and remove outliers that distort the efficiency analysis results. The main reason for the effectiveness of the Heaviside pattern lies in its simplicity and ability to detect extreme deviations in the behavior of the data more sensitively and directly compared to other techniques [48]. Therefore, any DMU with a leverage value significantly greater than the overall leverage will be eliminated [36].

Two methods were used to detect outliers in the data. The first method, the Kolmogorov–Smirnov (K-S) test, identified 2.08% of the data as anomalous. The second method, the Heaviside criterion, detected 5.42% outliers. Using the Heaviside standard was more effective in this study because it was more sensitive to significant and extreme deviations in the data, which is crucial for a robust energy efficiency analysis. It offers a simple and effective approach to detect and remove outliers, thus contributing to the reliability and accuracy of the DEA results. Therefore, the authors continued with the Heaviside standard instead of the Kolmogorov–Smirnov test. We proceeded with this method and removed the outliers from the original dataset, resulting in a consolidated data frame of 227 observations. The statistical summary of the model input and output data without outliers can be found in Table 4.

**Table 4.** Statistical summary of demand and GDP data without outliers between 2004 and 2023.

	Demand (GWh)	GDP (Millions of R\$)
Min	27,657	156,954
1st Qu.	32,480	273,200
Median	37,867	458,517
mean	36,722	452,982
3rd Qu	40,078	573,219
Max	46,407	950,791

When comparing the values in Tables 3 and 4, notice a difference in the mean values of demand and GDP. While the demand values are close to the original values (36.622 to 36.722), there is a more noticeable difference in gross domestic product (GDP) (457.192 to 452.982). To better understand the potential negative impact of these observations, the efficiency measure for each month was calculated before and after the outlier removal process. This information can be found in Figures 2 and 3.

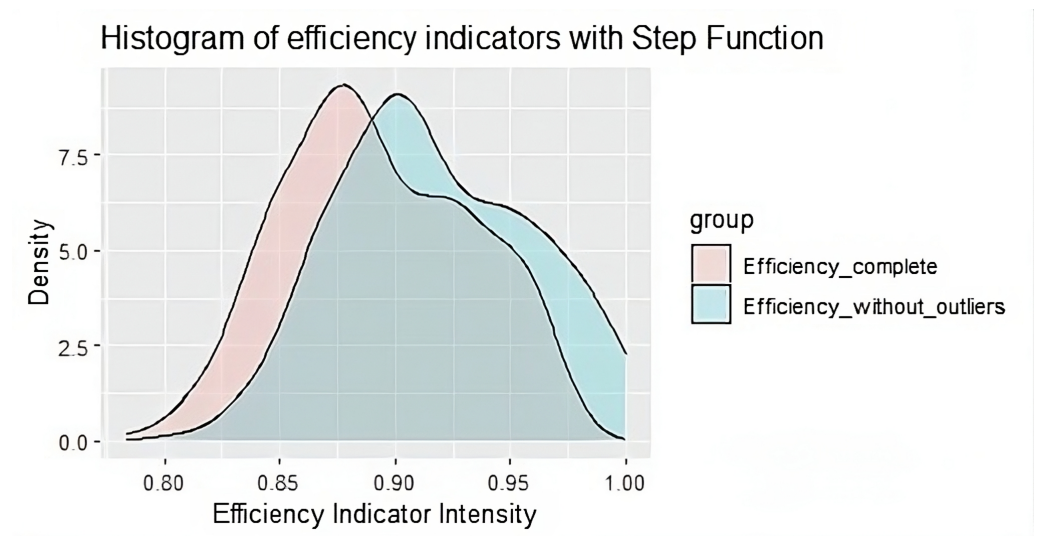


**Figure 2.** Box plot for the efficiency index for the data with and without outliers.

Figure 2 depicts the shift in the median value, and Figure 3 illustrates the density distribution of the data. When the database is cleaned, the values shift toward higher values. The complete dataset had a median value of 0.888, while the values without outliers have a median value close to 0.912, representing a difference of 0.024.

A Wilcoxon test was conducted to determine if this difference is statistically significant.  $V = 25.878$  and  $p\text{-value} < 2.2e^{-16}$ , indicating that this difference is substantial and should

be considered relevant. The data were then parameterized. This is because R tends to produce errors with null and non-parameterized data.



**Figure 3.** Density of the efficiency index for the data with and without outliers.

### 5.3. Confidence Interval Construction with Bootstrap

Due to the absence of statistical inference tools for classical non-parametric models, we can use bootstrap-based methods to compute DEA efficiency, including statistical inferences, hypothesis tests, and confidence intervals.

This model is a stochastic DEA model. It introduces the bootstrap concept and, later, introduces descriptive methods to identify influential data for non-parametric calculations [35]. These methods allow the use of statistical inferences without compromising the non-parametric nature of the problem [49].

The authors [35] have applied bootstrapping to estimate the confidence intervals of the efficiency measures given by the DEA. Bootstrap simulates a sample by using the original estimator and making the simulation results replicate the original through a resampling process, which is repeated  $W$  times (usually  $W = 2000$ ). This process can be described in four steps:

- i For each observation  $(x_i, y_i), i = 1, \dots, n$ , calculate the corresponding DEA efficiency score  $\hat{\theta}_i$  using linear programming;
- ii Draw a dataset from the original sample randomly using bootstrap, generating a random sample of size  $P$ , where  $P$  is the same size as the original sample. For this sample, obtain  $\hat{\theta}_r^*, r = 1, \dots, p$ ;
- iii From this random sample  $\hat{\theta}_r^*$ , construct  $\{x_r^* = [x_1^*, \dots, x_p^*], x_r^* = [x_1^*, \dots, x_p^*]$ , where for  $\hat{\theta}_i$  with orientation input  $x_r^* = \left(\frac{\hat{\theta}_i}{\hat{\theta}_r^*}\right)x_i$  and for  $\hat{\theta}_i$  with orientation output  $y_r^* = \left(\frac{\hat{\theta}_r^*}{\hat{\theta}_i}\right)y_i$  with  $r = 1, \dots, p$ ;
- iv Calculate the estimated bootstrap  $\hat{\theta}_{(r,w)}^*$  by solving the linear problem with DEA constraints using  $\hat{\theta}_{(r,w)}^*$ , which is given by Equation (1) [32].

$$\hat{\theta}_{(r,w)}^* = \min \left\{ \hat{\theta} \mid \sum_{r=1}^p \lambda_r x_r^* \leq \hat{\theta} x_0^*, \sum_{r=1}^p \lambda_r y_r^* \geq y_0^*, \sum_{r=1}^p \lambda_r = 1, \lambda_r \geq 0, \forall r \right\} \quad (1)$$

Here,  $x_0^*$  is the virtual input, and  $y_0^*$  is the virtual output.  $\lambda_r$  represents the weights. Then, we repeated Steps 2, 3, and 4  $W$  times to obtain the result for each observation  $r = 1, \dots, p$ , resulting in a set of bootstrap estimates  $\{\hat{\theta}_{r,w}^*, w = 1, \dots, W\}$ .

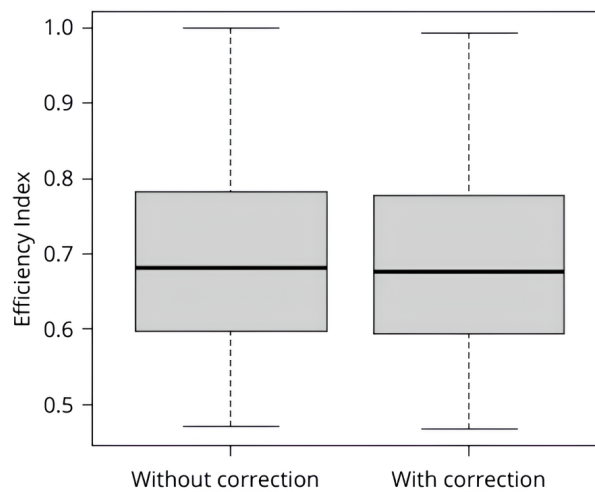
With the bootstrap number of repetitions  $W$ , we obtained the result with a 95% confidence interval for each observation from the set of estimates [32].

All of these values are calculated using a function provided by the *FEAR* package in R with the function `boot.sw98`

After applying the bootstrap–DEA computational method, the adjusted values were compared to analyze the data’s return to scale. The estimated result is  $\theta_E = 1.010459$ , and the critical value is  $C_\alpha = 0.9350187$ . As a result, no significant differences are observed in the technological frontiers (T), indicating a consistent scaling behavior.

This aligns with economic theory because the analysis examines Brazil at various times. In theory, the output–input relationship’s return should remain constant since Brazil’s size does not change over time, preventing cost variations due to size. As a result, a continual return to the scale can be assumed.

Thus, the values given by the CRS model using the bootstrap computational method with 2000 replications were used. A box plot graph is plotted in Figure 4 to visualize the values with and without correction.



**Figure 4.** Box plot for the efficiency index for the data without and with correction.

Visually, from Figure 4, small changes in efficiency scores can be observed. Therefore, it was decided to evaluate the results using the Wilcoxon test to verify whether the differences in scores with and without correction are significant. The null hypothesis ( $H_0$ ) stated no significant differences in the medians. The test returned a  $p$ -value  $< 2.2e^{-16}$ ; thus,  $H_0$  is rejected.

The statistical summary of the data is presented in Table 5.

**Table 5.** Statistical summary of the data without and with correction with a 95% confidence interval.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
No correction	0.4703	0.5971	0.6813	0.6987	0.7823	1.0000
With correction	0.4672	0.5933	0.6770	0.6942	0.7773	0.9936
Max. 95%	0.4702	0.5970	0.6813	0.6986	0.7822	0.9999
Min. 95%	0.4593	0.5832	0.6655	0.6825	0.7641	0.9768

Table 5 shows that the efficiency scores without correction resulted in higher efficiency values (0.6813 and 0.6770). Hence, the difference between the models without and with correction is significant and should be considered.

### 6. Identification of Parameters ( $p, d, q$ )

This section explains the functioning of the ARIMA and SARIMA models and the procedures used to carry out the initial stage of the Box–Jenkins methodology to identify the parameters ( $p, d, q$ ). This section refers to Step 3 of the methodology presented in Figure 1.

### 6.1. Autoregressive Integrated Moving Average (ARIMA) Models and Seasonal (SARIMA)

A time series is determined by a variable ( $y_t$ ) with quantitative values over a period ( $t$ ). Therefore, a time series analysis aims to explain future behaviors based on the variable's past behaviors. It is a logical procedure based on historical facts, and it is essential to understand how stochastic processes work because variables are subject to random variables, which can arise from points, events, or phenomena observed over time. Time series forecasting involves various areas of knowledge (multidisciplinary), and its implementation is not complex, mainly because it only requires the variable under analysis with its historical data [50]. For this purpose, there are autoregressive (AR), moving average (MA), autoregressive moving average (ARMA), and autoregressive integrated moving average (ARIMA) models, which will be described in detail in Section 6.1.

The autoregressive model is based on the concept that the values ( $y_t$ ) are explained by the ( $m$ ) previous values given by ( $y_{t-1}, y_{t-2}, y_{t-3}, \dots, y_{t-m}$ ). Thus, for an autoregressive model of order ( $p$ ) AR( $p$ ), it will be given by Equation (2) [22].

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + e_t \quad (2)$$

Here,  $\mu$  is the mean of the variable ( $y$ ),  $\alpha_t$  is the proportion given the period ( $t - 1$ ), and ( $e_t$ ) represents the random error of values that are uncorrelated with the mean ( $\mu$ ) but cause changes due to unknown and uncontrollable reasons in the values of ( $y_t - \mu$ ), being random disturbances of the period ( $t$ ). Therefore, the values involved in the preceding models consider their current and previous values, demonstrating that the AR( $p$ ) model is explained by its values and independent of its regressors.

The moving average model considers the values of the error term ( $e_t$ ) in terms of the present and past as well as a constant ( $c$ ) to explain ( $y_t$ ). Therefore, for a moving average process of order  $q$  MA( $q$ ), it can be expressed by Equation (3) in a general way [22].

$$y_t = c + \beta_0 e_t + \beta_1 e_{t-1} + \beta_2 e_{t-2} + \dots + \beta_q e_{t-q} \quad (3)$$

Therefore, the moving average model will be a forecast based on the linear combination of white noise. This is feasible because the AR and MA models assume they are based on a series of linear systems with random error, a mean of zero, a constant standard deviation, and no autocorrelation.

The autoregressive and moving average models, which result in the ARMA model, are the autoregressive terms and the moving average term ARMA gives ( $p, q$ ). For simplicity, the ARMA(1,1) model is presented in Equation (4) [22].

$$y_t = w + \alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p} + e_t + \beta_0 e_t + \dots + \beta_q e_{t-q} \quad (4)$$

Here,  $w$  represents a constant term.

The AR, MA, and ARMA models described assume that the time series involved in the analysis are (weakly) stationary, meaning a series with constant mean, variance, and autocovariance, not varying over time, where the measurements given by the variable ( $y_t$ ) will have variations with more or less constant amplitudes, and their values tend to return to their mean. However, many time series are non-stationary and do not exhibit this continuous behavior. One way to deal with these cases is by differencing ( $d$ ) the time series according to the order of differencing, usually with a value of ( $d$ ) being 1 or 2 [50]. Thus, a non-stationary time series of order ( $d = 1$ ) must be divided once to become stationary. Therefore, a time series ARIMA (1,2,1), differenced twice ( $d = 2$ ), can be analyzed by the ARMA (1,0,1) model, as already presented by Equation (4) of the ARMA model [22].

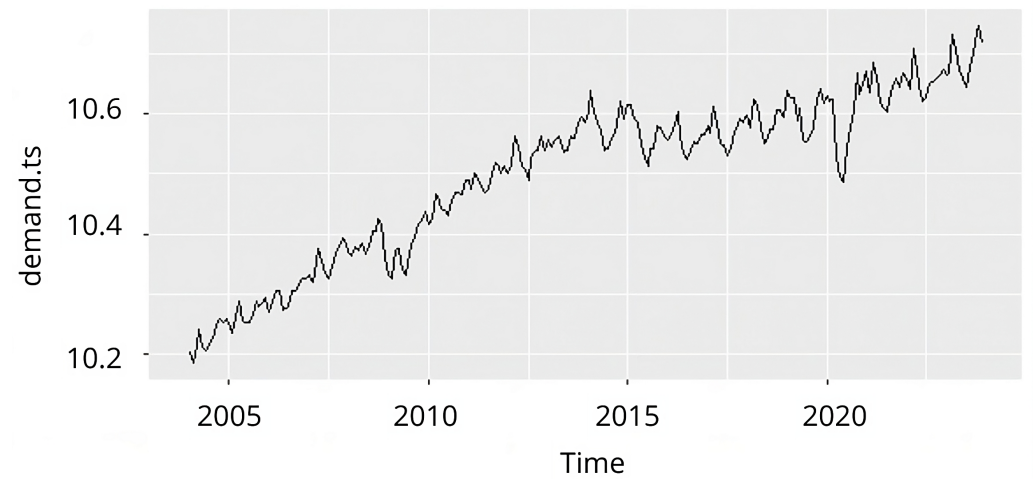
In cases with a seasonal behavior of the variable ( $y_t$ ), the seasonal autoregressive integrated moving average model, also known as SARIMA, is used. The SARIMA model can be represented by Equation (5).

$$y_t = c + \sum_{n=1}^p \alpha_n y_{t-n} + \sum_{n=1}^q \beta_n e_{t-n} + e_{t-1} \sum_{n=1}^P \alpha_n y_{t-sn} + \sum_{n=1}^Q \beta_n e_{t-sn} + e_t \quad (5)$$

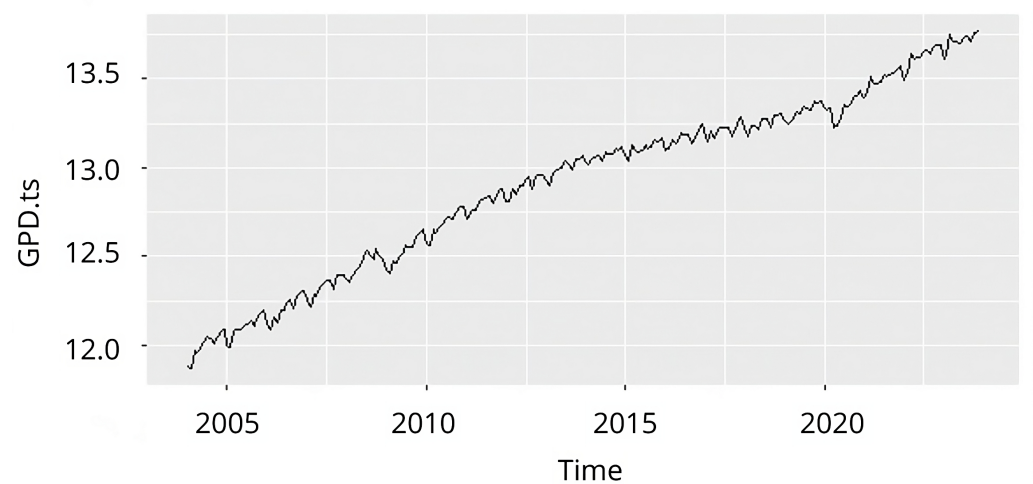
It consists of including seasonal autoregressive terms ( $P$ ), seasonal differencing ( $D$ ), a seasonal moving average ( $Q$ ), and the seasonal component ( $S$ ).

### 6.2. Transforming Variables into Time Series

For data analysis on demand and GDP, the data are transformed into a time series from 2004 to 2023 using the logarithm of the variables. These historical data series can be viewed in Figures 5 and 6, respectively. Their respective data are summarized in Table 6.



**Figure 5.** Energy demand time series from 2004 to 2023.



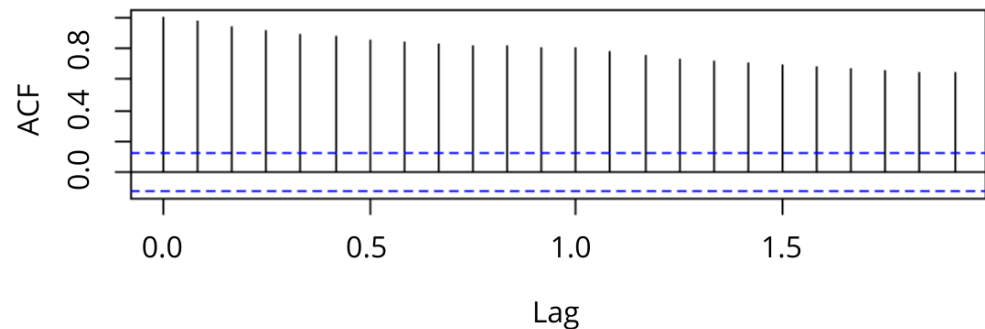
**Figure 6.** GDP time series from 2004 to 2023.

**Table 6.** Statistical summary for the electric energy demand and GDP time series.

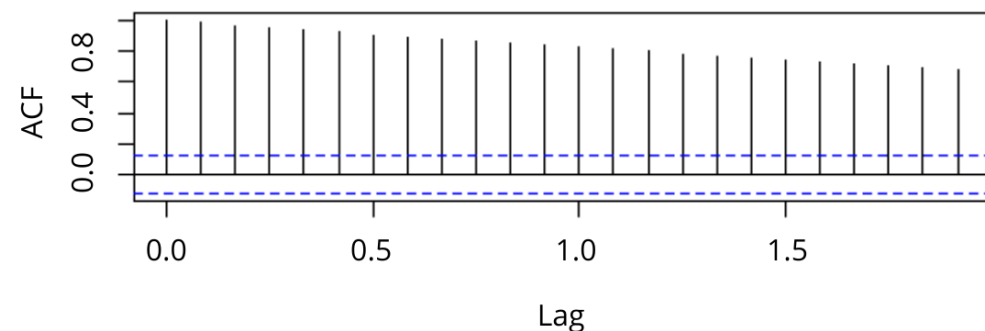
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Demand	10.19	10.38	10.54	10.50	10.60	10.75
GDP	11.87	12.50	13.04	12.92	13.28	13.77

Figures 5 and 6 do not visually demonstrate stationary behavior, which is a prerequisite for using AR, MA, or ARIMA models. These models are based on dealing with data that have a constant mean and variance over time. Therefore, the trend of the data in a preliminary analysis tends to be non-stationary, and the volatility of the data suggests a seasonal behavior, indicating that the SARIMA model may be appropriate for analyzing data forecasts. However, these conclusions must undergo statistical testing to obtain reliable results and projections corresponding to reality.

The autocorrelation function is used to identify non-stationary time series and observe the presence of unit roots and trends in the data. The autocorrelation function can be visualized in Figures 7 and 8.



**Figure 7.** Autocorrelation function of energy demand, illustrating the correlation of the time series with its own past values.



**Figure 8.** GDP autocorrelation function, illustrating the correlation of the GDP time series with its own past values.

From the ACF plots, both the demand and GDP exhibit trends in the data and the presence of unit roots, and there is no evidence of ergodicity in the models. In the ACF test, if the lags decay very slowly, it indicates non-stationarity in the model's data. The blue dotted lines indicate the significance limits for the autocorrelation coefficients, showing whether the autocorrelation values at different lags are significantly different from zero.

For a model to be considered stationary, it must exhibit constant mean, variance, and autocorrelation, thus showing no changes over time. However, values may fluctuate above and below but converge to the same mean. One of the stationarity tests is the augmented Dickey–Fuller (ADF) test, where the null hypothesis ( $H_0$ ) indicates the presence of a unit root, suggesting that the model is not stationary.

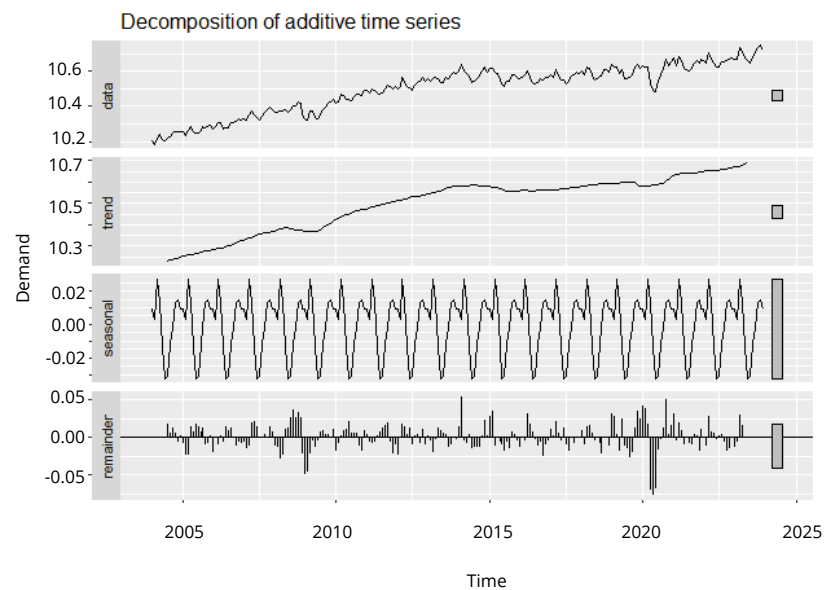
The ADF test on demand yielded a  $p$ -value of 0.2599, which suggests an acceptance of  $H_0$ , indicating that the market is not stationary. The ADF test in GDP returned a  $p$ -value of 0.7107, which suggests that the GDP time series is also not stationary.

### 6.3. Time Series Seasonality Test

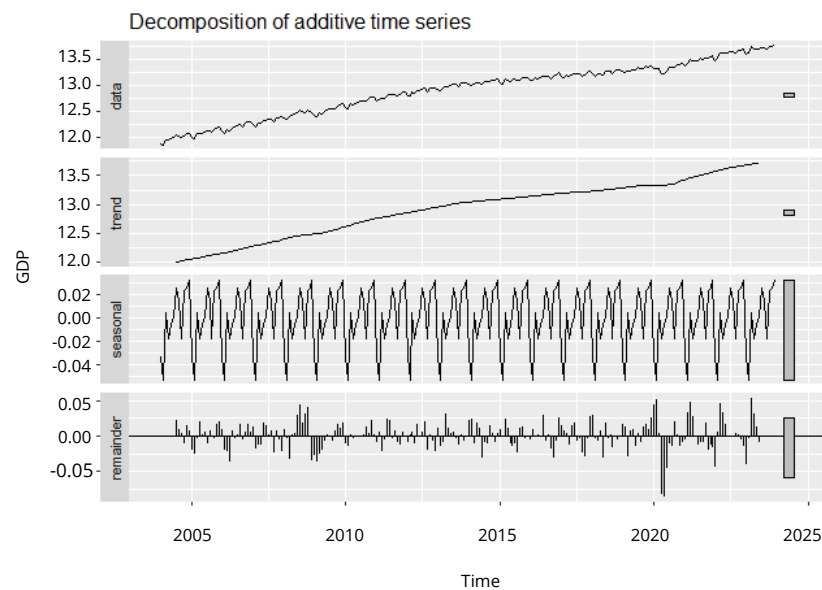
To visualize the seasonal series, the series is decomposed to identify patterns through trend components (analysis of the series plot), seasonality (analysis of the seasonal index plot), cycles (analysis of the seasonal index plot and irregularities), and residuals (analysis of the irregularities plot). This advances the study of time series since it isolates a specific term determined by the series. The decomposition of the series can be seen in Figures 9 and 10.

The energy demand and GDP data show a growth trend over time, requiring one difference to achieve stationarity. A seasonal variation behavior caused by external factors is observed concerning the seasonal index. To understand whether this seasonal variation is significant and should be considered, the range of variation of the seasonal component

was examined. In both cases, it is important to indicate that the seasonal component is present in the series and is substantial enough to be considered in the forecast.



**Figure 9.** Decomposition of the energy demand time series, revealing its trend, seasonal, and residuals components.

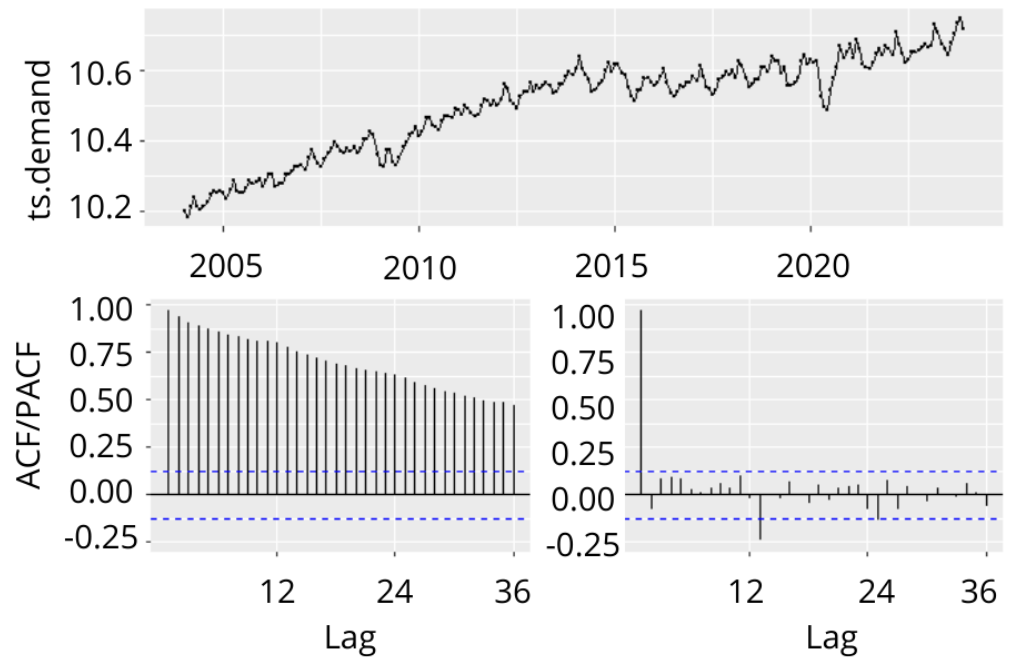


**Figure 10.** Decomposition of the GDP time series, highlighting its trend, seasonal, and residuals components.

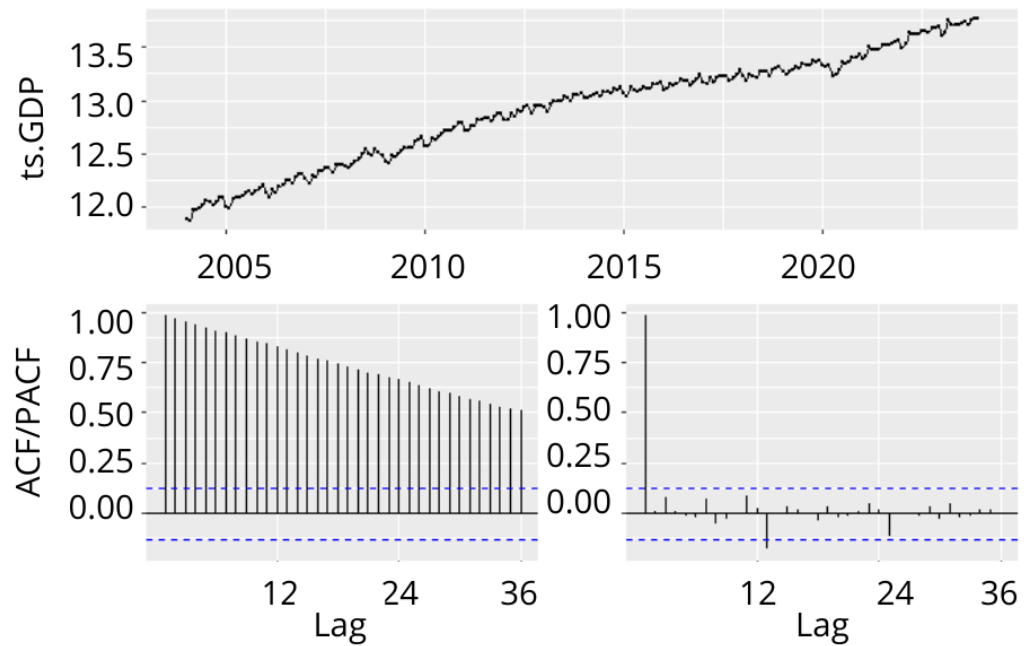
There are no irregular cycles with systematic alternation or long low and high values periods. Instead, there is regularity in the data distribution. Thus, to confirm the analyses made using the decomposition graphs, autocorrelation and partial autocorrelation graphs of the series were created, as shown in Figures 11 and 12.

The ACF graph also shows demand and GDP trends over time, and the PACF graph demonstrates the absence of data seasonality, as most of the partial autocorrelation values vary within the range represented by the blue dashed line. However, as the series exhibits a trend and non-stationarity, logarithmized and differenced series were used to re-examine the ACF and PACF functions. The series decomposition indicated a possible seasonality, which can be visualized in Figures 13 and 14.





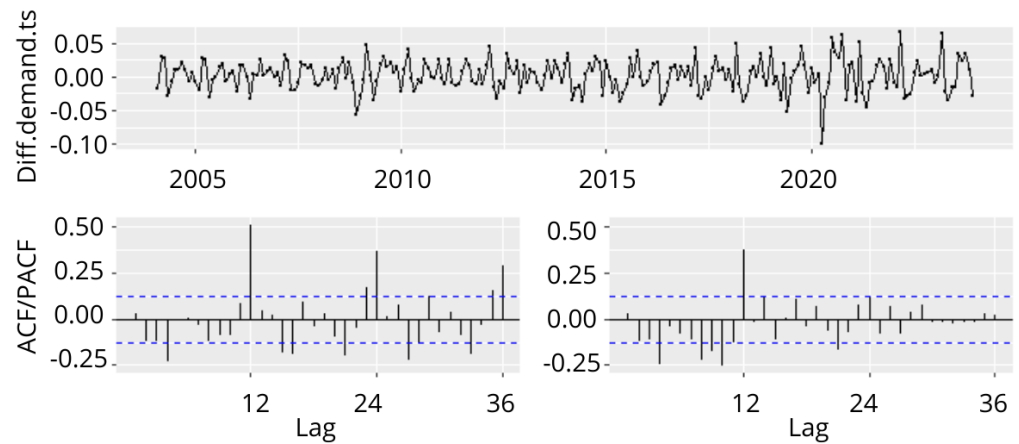
**Figure 11.** Energy demand first autocorrelation (FAC) and partial autocorrelation (FACP), showing immediate past correlations.



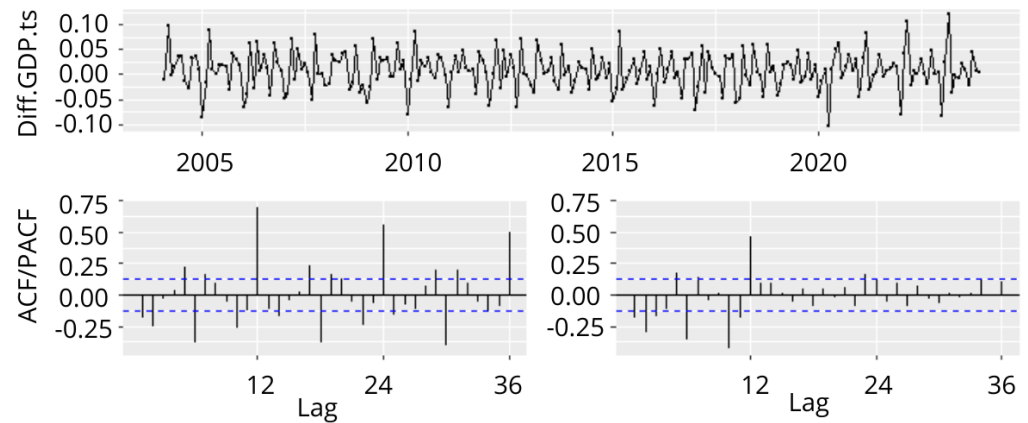
**Figure 12.** GDP first autocorrelation (FAC) and partial autocorrelation (FACP), showing immediate past correlations.

Thus, by removing the data trends through the first differencing, it is possible to identify a seasonal behavior given by the ACF and PACF functions, as there are several periods where the autocorrelation values exceed the intervals given by the dashed blue line. Significant intensity extrapolations are observed every 12 lags in demand and GDP with moderate intensity extrapolations every four lags in demand and every six lags in GDP.

Therefore, the SARIMA model is more suitable for applying forecasts than the ARIMA model because it considers the seasonal factors present in these series.



**Figure 13.** Energy demand first autocorrelation (FAC) and partial autocorrelation (FACP) with one difference, illustrating correlations with the first lag after differencing.



**Figure 14.** GDP first autocorrelation (FAC) and partial autocorrelation (FACP) with one difference, illustrating correlations with the first lag after differencing.

One option is to use the `auto.arima` function to determine the best model by adjusting the parameters and evaluating the model’s performance with the AICc indices. This dynamic process involves several dependencies and can help select parameters.

#### 6.4. Parameters Test with Auto Arima

The Box–Jenkins methodology will define the parameters  $(p, d, q)$ . This methodology involves fitting autoregressive integrated moving average models,  $ARIMA(p, d, q)$ , to a dataset. For model construction, an algorithm is structured in which the choice of model structure is based on the data itself.

The R function `auto.arima` performs most of this process by testing different parameter combinations  $(p, 1, q)$  and comparing AICc metrics to determine the best model. The `auto.arima` algorithm is an automated technique for identifying the most suitable ARIMA model for a time series. It consists of three key components. The autoregressive (AR) component uses past values of the time series to predict future values, the integrated component (I) represents the number of differences needed to make the time series stationary by removing trends, and the moving average (MA) component uses forecast errors of past values to correct future forecasts.

The `auto.arima` model offers several advantages, including automation, which reduces the need for manual intervention in selecting model parameters and increases efficiency, saving time and effort by automating the model search process. Additionally, the model provides flexibility to handle complex time series data, including trends and seasonality. However, the `auto.arima` algorithm has some limitations, such as the default presence of a

stepwise approach to improve the model's potential range, so the stepwise function can be disabled to explore a more expansive model space. In addition, there is the option of deactivating the approximation function to improve model accuracy, although this may slow down the adjustment process.

The models defined as the best parameters by auto.arima for demand were SARIMA (1,1,1) (0,1,2), and for the adjusted model, it was SARIMA (1,1,1) (1,1,1). For GDP, the parameters were SARIMA (1,1,3) (0,1,2), and for the adjusted model, it was SARIMA (0,1,2) (2,1,1). The AIC, AICc and BIC test values, as well as the error parameters, can be seen in Tables 7 and 8.

**Table 7.** AIC, AICc, and BIC values for different parameters used in auto.arima.

Variable	SARIMA	AIC	AICc	BIC
Demand	(1,1,1) (0,1,2)	−1164.68	−1164.41	−1147.56
	(1,1,1) (1,1,1)	−1164.68	−1164.41	−1147.56
GDP	(1,1,3) (0,1,2)	−1079.2	−1078.69	−1055.23
	(0,1,2) (2,1,1)	−1078.71	−1078.33	−1058.16

**Table 8.** RMSE, MAE, MAPE, and MASE values for different parameters used in auto.arima.

Variable	SARIMA	RMSE	MAE	MAPE	MASE	ACF1
Demand	(1,1,1) (0,1,2)	0.0174	0.0132	0.1253	0.3928	0.0041
	(1,1,1) (1,1,1)	0.0173	0.0132	0.1250	0.3917	0.0062
GDP	(1,1,3) (0,1,2)	0.0206	0.0154	0.1192	0.1686	−0.0011
	(0,1,2) (2,1,1)	0.0208	0.0157	0.1209	0.1710	−0.0076

Table 7 presents the values for the AIC, AICc, and BIC tests provided by the auto.arima models with the lowest values describing the best model. However, the default and adjusted models for energy demand showed the same values, and for the GDP models, the first had a lower BIC value while the adjusted model had a lower AIC. Following the principle of parsimony, the simpler model was chosen, thus proceeding with the unadjusted model.

This is confirmed when observing the values given by RMSE, MAE, MAPE, and MASE, which are measures of error evaluation. The closer the value is to zero, the better the model. Therefore, the SARIMA(1, 1, 1)(0, 1, 2) performed better for demand than SARIMA(1, 1, 3)(0, 1, 2), and both variables performed better with the adjusted model.

Adjusting the parameters to apply the ARIMA model shows a drop in yield. The RMSE, MAE, MAPE, and MASE values are also presented in Table 9.

**Table 9.** ME, RMSE, MAE, MAPE, and MASE values for different parameters used in ARIMA.

Variable	ARIMA	RMSE	MAE	MAPE	MASE	ACF1
Demand	(1,1,1)	0.0231	0.0184	0.1749	0.5478	−0.0404
GDP	(1,1,3)	0.0359	0.0287	0.2235	0.3136	−0.1028
	(0,1,2)	0.0372	0.0302	0.2346	0.3292	−0.1340

Table 9 contains the parameters of the ARIMA model and the adjusted models for GDP.

Comparing the SARIMA (Table 8) model with the ARIMA (Table 9) model shows that the SARIMA models perform better. Therefore, we will analyze the residuals of the SARIMA model to determine if they exhibit characteristics of white noise, which is essential for applying the forecasts.

## 7. Diagnostic Checking

This section presents the statistical tests applied to analyze the residuals. This section refers to Step 4 of the methodology, which can be seen in Figure 1.

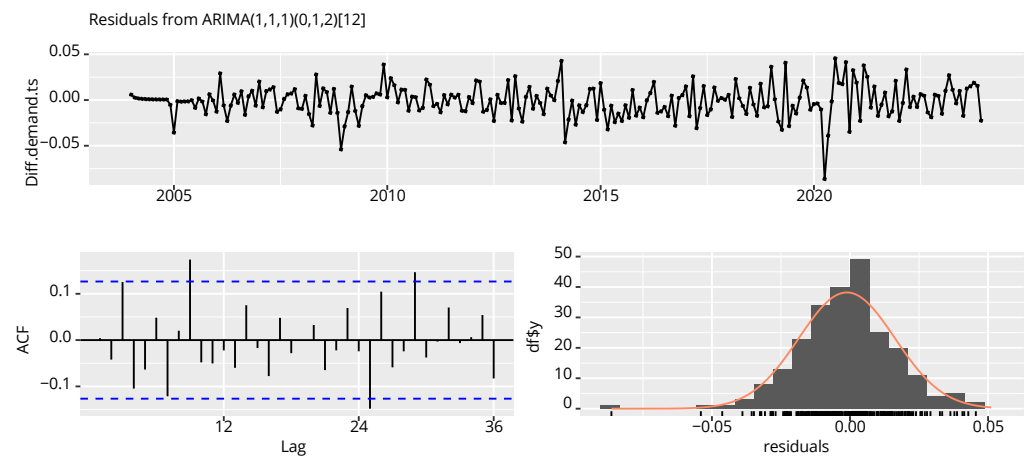
The Ljung–Box test was applied to the models to diagnose the parameters used. The results presented show that the models do not fail. The values of the x-squared test show no autocorrelation in the residuals up to lag 1. The values of the  $p$ -value indicate that the

null hypothesis cannot be rejected and that there is no autocorrelation in the residuals, showing that the residuals behave like white noise (Table 10).

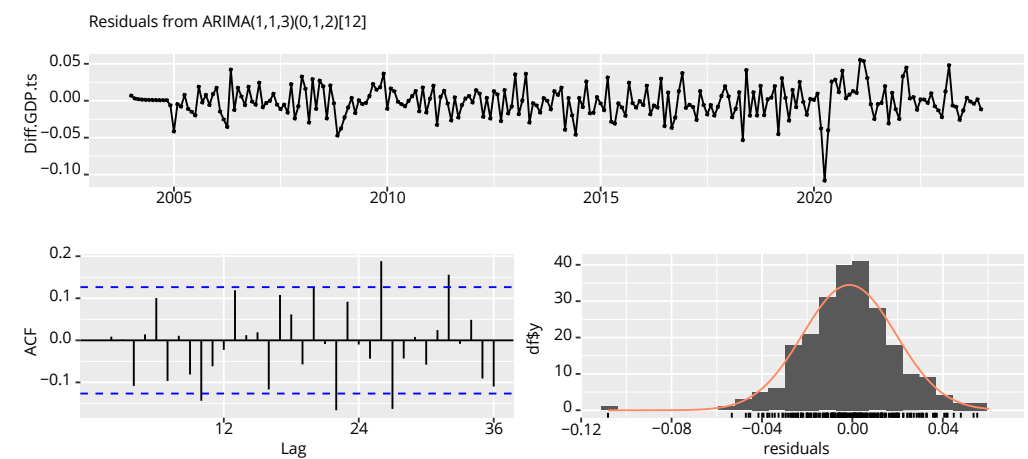
**Table 10.** Ljung–Box test.

SARIMA	$\chi$ -Squared	df	<i>P</i> value
(1,1,1) (0,1,2)	0.0040	1	0.9494
(1,1,3) (0,1,2)	0.0003	1	0.9869

Observing the waste graph in Figures 15 and 16, it is clear that waste presents homoscedasticity over time except for 2020 (COVID-19).



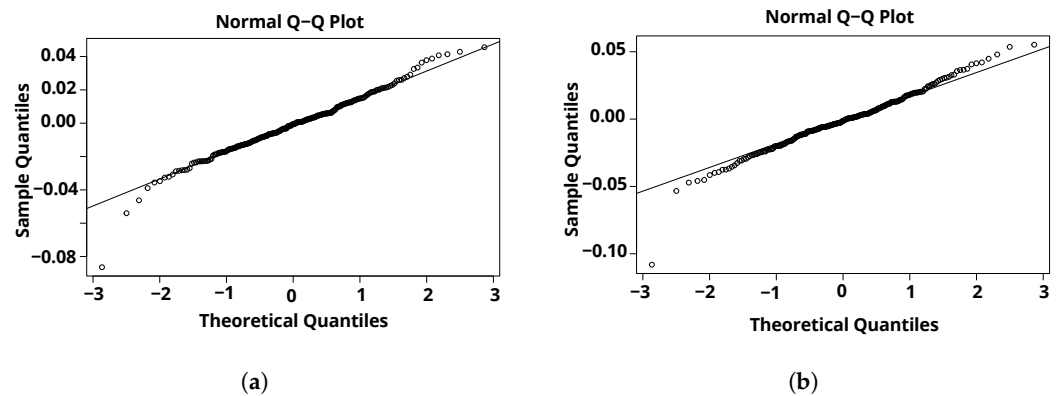
**Figure 15.** Analysis of residuals from the SARIMA (1,1,1) (0,1,2) model for the energy demand time series, assessing the differences between observed and predicted values.



**Figure 16.** Analysis of the residuals from the SARIMA(1,1,3) (0,1,2) model applied to the GDP time series, evaluating the differences between observed and predicted values.

Blue lines indicate the significance limits for autocorrelation coefficients, while the red line approximates a normally distributed curve.

To confirm the normality of the data, the normal Q-Q plot of the residuals was plotted, as shown in Figure 17a,b.



**Figure 17.** Normal Q-Q plots of the residuals from the SARIMA models, assessing the normality of the residuals by comparing their quantiles against a theoretical normal distribution. (a) Normal Q-Q plot of the residuals from the SARIMA(1,1,1) (0,1,2) model for energy demand. (b) Normal Q-Q plot of the residuals from the SARIMA(1,1,3) (0,1,2) model for GDP.

Therefore, we decided to use the parameters based on the first test (Table 10) and the observations in Figures 15–17. Therefore, it is accepted that the residuals behave like white noise.

## 8. Results and Discussion

This section presents an analysis and comparison of the results obtained in this study to highlight the innovative contributions of the research. One of the main aspects that was compared was the performance of ARIMA/SARIMA models applied to forecasting energy demand in Brazil. Previous studies, such as [27,51], demonstrated the effectiveness of these models in time series scenarios. However, this study differentiated itself by using a combination of advanced forecasting techniques, combining classical modeling with bootstrap resampling methods to estimate confidence intervals, increasing the results' robustness.

In addition, using the Heaviside pattern for detecting and removing outliers proved to be more effective than traditional methods, such as the Kolmogorov–Smirnov test, as reported in previous research. The Heaviside pattern, when implemented in the context of data envelopment analysis (DEA), showed greater sensitivity to identifying extreme deviations, resulting in a more accurate assessment of the energy efficiency of DMUs (decision-making units). Compared to other studies, such as [25], which used more traditional outlier detection approaches, the innovative application of this technique in this study strengthens the conclusions about energy efficiency in Brazil.

Another innovative point refers to the use of the Jackknife leverage technique, which, by iteratively removing DMUs and recalculating efficiency scores, provided a robust and detailed analysis, minimizing the influence of outliers on the final assessment. This method has been underused in the existing literature, making its application here an important methodological contribution.

When comparing the results of this study with existing research, it is clear that despite the widespread use of ARIMA and SARIMA models in energy time series forecasts, the combination of these models with bootstrap reanalysis techniques and the focus on more sensitive outlier detection improve the quality and accuracy of the forecasts. In addition, incorporating more robust methods, such as Jackknife leverage and the Heaviside pattern, represents an innovation in relation to previous studies, which typically focus on simpler statistical methods.

This section is Step 5 of the methodology visualized in Figure 1, and it presents the consolidation of the results found in Steps 2, 3, and 4. In summary, this research advances existing knowledge by efficiently integrating classical and innovative techniques, standing out for the robust application of methodologies that improve the analysis of energy efficiency and demand predictability in the Brazilian energy sector.

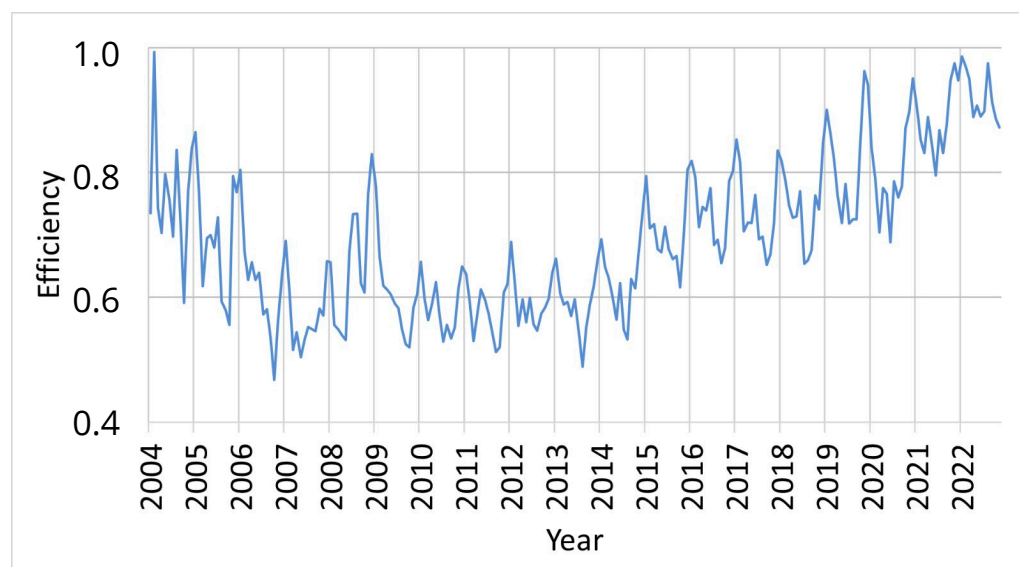
### 8.1. Energy Efficiency Analysis

The efficiency scores for the dataset are shown in Table A2 (Appendix A). Blank spaces indicate outliers for the month.

The DMU with the highest efficiency score was in August 2004, while the lowest efficiency score was observed in April 2006. On average, Brazil can reduce energy expenditures by 30.58% given the same GDP value, considering the productive frontier of 2004 and 2023. The mean value of 0.6942 for the efficiency index is relatively low, indicating that in most months, Brazil operates well below what can be considered efficient.

The DEA highlights the necessity for substantial changes to minimize energy usage. It also suggests that such reductions are achievable, as there were times when energy efficiency values were nearing 1.

The trend of efficiency indices from 2004 to 2023 is presented in Figure 18.



**Figure 18.** Efficiency index from 2004 to 2023 calculated using data envelopment analysis (DEA), reflecting the relative performance and productivity of energy demand during this period.

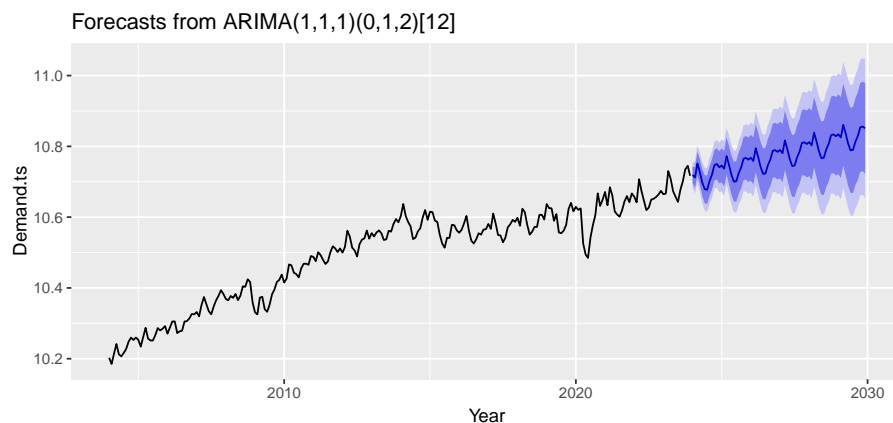
Despite Brazil presenting an inefficient index on average, Figure 18 shows a decline initially followed by a recovery over time, indicating that with technological advancements, operations are becoming more efficient. There is, however, a significant variation from month to month, demonstrating that there are periods where operations become efficient and periods where they do not [52].

### 8.2. SARIMA Forecasting

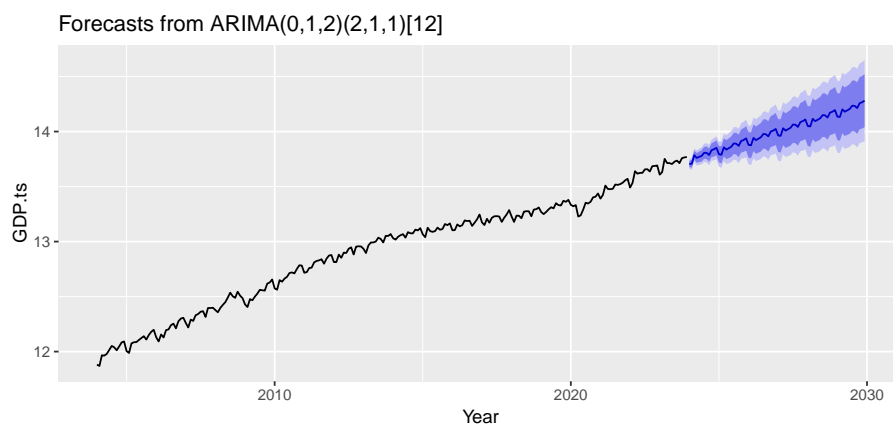
With the  $(p, d, q)$  parameters defined, SARIMA forecasts were made. The models were forecasted for 72 months with December 2023 as the reference month.

The results with 95% confidence intervals from the SARIMA(1,1,1) (0,1,2) model for demand and SARIMA(1,1,3) (0,1,2) model for GDP are presented, respectively, in Tables A3 and A4 (Appendix A).

The times series data have been plotted in Figures 19 and 20 to improve clarity. In the plots, the blue region indicates the forecast made by SARIMA with the darker line representing the forecast and the surrounding area representing the model confidence interval.



**Figure 19.** Forecasting energy demand with the SARIMA(1,1,1) (0,1,2) model, extending predictions through 2030 based on historical data and identified seasonal patterns.



**Figure 20.** SARIMA (0,1,2) (2,1,1) forecasting for GDP projected through 2030, providing estimates based on historical data and seasonal patterns.

The forecast data indicate a consistent upward trend for the next two years. Anticipated annual demand growth is 2.1% from 2023 to 2030. These increases are below the average the International Energy Agency (IEA) gives, which predicts an average annual growth of 2.5% through 2026 [53].

The rise in energy consumption is thought to be caused by increased economic activity and residential energy use due to the high temperatures Brazil has experienced in recent years. SARIMA predicts that this increase will be more gradual compared to the energy matrix in 2023.

According to projections, GDP is on an upward trend. In 2024, it should close at close to 11.7 trillion and could reach 19.5 trillion in 2030.

### 8.3. Implications

In recent years, Brazil has been seeking alternatives to mitigate the impacts caused by the increase in greenhouse gas (GHG) emissions. The country aims to address the effects of climate change by proposing public policies until 2050. These policies are fundamental in developing strategies to mitigate changes and reduce energy use. Brazil also focuses on making the energy sector more efficient by expanding renewable sources such as wind, solar, and biomass energy [54].

Some policy proposals are being considered for 2050 to reduce carbon emissions and make the Brazilian energy transition more efficient [55]. These proposals include (i) implementing a flexible energy policy to support decarbonization, better use of energy resources, and alignment with climate targets (NDC); (ii) encouraging technological neutrality through market opening, promoting greater competitiveness, and providing incentives for combining different

technologies; (iii) exploring Brazil's advantages in oil and gas, biofuels, and renewables to facilitate the energy transition and utilize new decarbonized energy sources; (iv) ensuring an inclusive and economically viable energy transition without imposing higher costs on society; (v) improving legal and regulatory frameworks to promote emission reduction technologies, advanced biofuels, and hydrogen; and (vi) studying the climate resilience of energy solutions, such as hydroelectric, wind, solar, and biofuels [56]. The country has implemented key policies in the energy sector, including the Ten-Year Energy Expansion Plan 2030 and Law 13,576/2017 National Policy for Biofuels (RenovaBio). These policies aim to improve the energy sector's efficiency and sustainability [57].

Brazil has implemented studies and policies and developed a diverse renewable energy matrix. In 2022, 48% of the energy used in the country comes from renewable sources, which is higher than the global average of 14% [57]. However, compared to other countries, Germany and Great Britain are pursuing more ambitious goals for the sustainability of the energy sector. They seek to reduce CO<sub>2</sub> emissions through carbon pricing in this sector [58]. Although, in practical terms, OECD countries have only 11% of the energy matrix from renewable sources, Brazil has a competitive advantage for sustainable energy transition [57], and it still has room for improvement in terms of efficiency.

The procedure adopted, which involves DEA combined with computational methods like bootstrap, has proved to be effective in measuring energy efficiency and assessing the frontiers of the country's productive opportunities. Based on energy demand and GDP variables, the conclusions of its effectiveness show that Brazil has become increasingly energy efficient over time due to its technological advances. However, there is still significant room for improvement to achieve actual efficiency since the potential reduction of the profit margin in energy consumption costs is approximately 30.58%. As shown in Section 8.1, there are periods of highs and lows throughout the months under consideration, suggesting that seasonal factors impact energy efficiency.

Although Brazil has improved its efficiency over the years, there has been considerable variation from month to month, highlighting the need for further improvements. There is still significant potential for improvement by introducing new technologies that can optimize production costs or maintain productivity while reducing energy costs. One potential solution is to invest in sustainable energy technologies to increase the amount of renewable energy in the country's energy matrix. By doing so, Brazil can increase its efficiency and competitiveness and reduce electricity costs for households and industries. This highlights the importance of the 2030 agenda targets for the industrial sector, particularly SDG 7, which aims to ensure access to reliable, sustainable, renewable, and affordable energy.

## 9. Conclusions

This study used data envelopment analysis (DEA) to evaluate Brazilian energy efficiency and applied SARIMA models to forecast trend demand and GDP variables, using monthly data from 2004 to 2023 and computer simulations using R software.

The study accomplished its objectives. (i) It utilized non-parametric models from the bootstrap-DEA computational methods to analyze the country's energy resources. The study revealed inefficiencies (30.58%) in the energy matrix over a significant period, indicating a substantial opportunity for improvement to enhance the country's efficiency. (ii) SARIMA outperformed the ARIMA model in forecasting demand and GDP series, which was primarily due to the seasonal behavior of the series. (iii) The study provided a comprehensive understanding of electricity demand by utilizing SARIMA models to predict the future behavior of the energy matrix. (iv) The study proposed a methodology for evaluating Brazilian energy efficiency, which can aid in developing public policies and initiatives to ensure the country's alignment with SDG7, thus increasing awareness of energy strategies.

A suggestion for further investigation would be to include more variables such as climate, energy imports and exports, energy prices, and the effects of CO<sub>2</sub> pollution resulting from electricity generation in the country. According to the Electricity 2024 Analysis and Forecast for 2026 by the International Energy Agency (IEA) [53], energy production significantly contributes



to CO<sub>2</sub> emissions worldwide. Additionally, incorporating other variables correlated with energy consumption and GDP in the energy sector could enhance the analysis, providing more robust results for efficiency evaluation in future scenarios using forecasting. Furthermore, the research could benefit from testing a wider range of seasonal models to assess likely trends in the future. An approach that could also be applied would be cointegration to find a linear combination between two variables I(d), which leads to a variable of lower order of integration, as well as integrating machine learning models into smart grids for the better management of energy resources in energy generation, distribution, and supply [59].

**Author Contributions:** Conceptualization, data curation, writing, research and methodology, G.M.S. and A.L.M.S.; software, G.A.P.R.; validation, G.D.B.; Review, V.P.G.; project administration, C.N.; supervision, resource acquisition and funding, R.d.O.A. and C.A.S.B. All authors have read and agreed with the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** The data analyzed or generated supporting the results reported during the study can be found at <https://doi.org/10.6084/m9.figshare.25714932.v2> (accessed on 7 June 2024).

**Acknowledgments:** The authors would like to thank the Brazilian National Confederation of Industry (CNI) for partially supporting this project and for their support and collaboration throughout this research project.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

AR	autoregressive
MA	moving average
I	integrated
ARIMA	autoregressive integrated moving average
SARIMA	seasonal autoregressive integrated moving average
GDP	gross domestic product
ACF	autocorrelation function
PACF	partial autocorrelation function
ML	machine learning
CNNs	convolutional neural networks
CRFs	conditional random fields
ANNs	artificial neural networks
SG	smart grids
SMI	smart metering infrastructure
DEA	data envelopment analysis
DMUs	decision-making units
CRS	constant return scale
VRS	variable return scale
sfa	stochastic frontier analysis
KPSS	Kwiatkowski Phillips Schmidt Shin
DF	Dickey–Fuller
MAE	mean absolute error
MAPE	mean absolute percentage error
MSE	mean square error
RMSE	root mean square error
MASE	mean absolute scaled error
AIC	Akaike information criterion
BIC	Bayesian information criterion
AIC <sub>c</sub>	corrected Akaike information criterion
EPE	Energy Research Company
IPEA	Institute of Applied Economic Research

## Appendix A

**Table A1.** Actions for the SDG7 targets.

SDG Target	Description	Actions Taken by Brazil
<b>7.1: Universal access to affordable, reliable, and modern energy services</b>	Ensure everyone has access to clean and sustainable energy.	<ul style="list-style-type: none"> <li>National program for universal access to electricity (Light for All): expands access to electricity to remote and low-income areas</li> <li>Social tariff programs: provide subsidies for low-income households</li> </ul>
<b>7.2: Increase the share of renewable energy in the global energy mix</b>	Promote clean energy sources.	<ul style="list-style-type: none"> <li>Dominant hydropower: Brazil uses hydropower as its main source of electricity generation (around 60%)</li> <li>Biofuels program: encourages the production and use of biofuels, such as sugarcane ethanol, reducing dependence on fossil fuels</li> <li>Investment in wind and solar energy: government incentives and auctions promote the development of wind and solar farms</li> </ul>
<b>7.3: Double the global rate of improvement in energy efficiency</b>	Reduce energy consumption without compromising economic activity.	<ul style="list-style-type: none"> <li>Brazilian labeling program (PBE): labeling program that promotes energy-efficient appliances</li> <li>National plan for energy efficiency (PNEF): sets targets and strategies for various sectors to improve energy efficiency</li> <li>Industrial energy efficiency programs: offer incentives and technical assistance to industries for adopting energy-saving practices</li> </ul>
<b>7.4: Enhance international cooperation for clean energy research and technology</b>	Collaborate with other countries on clean energy development.	<ul style="list-style-type: none"> <li>Participation in international initiatives: Brazil actively participates in forums like the International Renewable Energy Agency (IRENA) and Mission Innovation.</li> <li>Bilateral cooperation: engages in joint research and development projects with other countries focusing on clean energy technologies</li> </ul>
<b>7.5: Expand infrastructure for sustainable energy services in developing countries</b>	Assist developing nations in accessing clean energy solutions.	<ul style="list-style-type: none"> <li>South-South cooperation: Brazil shares its expertise and technology in renewable energy with other developing countries</li> <li>Technology transfer initiatives: provides technical assistance and training programs for capacity building in clean energy technologies</li> </ul>

**Table A2.** Energy efficiency scores.

Date	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
January	-	0.697	0.728	0.573	0.553	0.733	0.583	0.529	0.573	0.556
February	-	0.837	0.594	0.581	0.549	0.734	0.548	0.556	0.543	0.547
March	-	0.727	0.579	0.535	0.546	0.623	0.525	0.534	0.512	0.574
April	0.735	0.591	0.556	0.467	0.582	0.608	0.520	0.552	0.519	0.583
May	-	0.771	0.794	0.560	0.571	0.766	0.584	0.614	0.608	0.598
June	-	0.840	0.769	0.636	0.658	0.830	0.605	0.650	0.621	0.640
July	-	0.865	0.805	0.691	0.656	0.777	0.657	0.637	0.689	0.663
August	0.994	0.765	0.674	0.611	0.555	0.664	0.599	0.590	0.624	0.607
September	0.744	0.618	0.628	0.516	0.548	0.619	0.563	0.530	0.554	0.589
October	0.703	0.695	0.656	0.544	0.539	0.613	0.592	0.572	0.597	0.593
November	0.798	0.700	0.628	0.504	0.531	0.606	0.625	0.613	0.560	0.570
December	0.757	0.680	0.639	0.532	0.674	0.590	0.570	0.596	0.599	0.597
Date	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
January	0.543	0.549	0.677	0.684	0.693	0.654	0.718	0.688	0.795	0.889
February	0.489	0.533	0.662	0.692	0.698	0.660	0.725	0.786	0.868	0.908
March	0.552	0.629	0.666	0.655	0.652	0.676	0.725	0.760	0.831	0.890
April	0.588	0.614	0.616	0.679	0.669	0.764	0.858	0.778	0.878	0.899
May	0.617	0.677	0.701	0.787	0.720	0.741	0.963	0.872	0.947	0.976
June	0.661	0.737	0.805	0.802	0.836	0.848	-	0.897	-	-
July	0.693	0.795	0.819	0.853	0.820	0.901	0.940	0.951	-	-
August	0.649	0.711	0.793	0.816	0.786	0.863	0.839	0.903	0.975	-
September	0.631	0.718	0.712	0.706	0.747	0.821	0.790	0.852	0.948	0.912
October	0.602	0.677	0.745	0.720	0.728	0.762	0.704	0.831	0.986	0.887
November	0.564	0.672	0.739	0.719	0.730	0.719	0.775	0.889	0.971	0.873
December	0.623	0.713	0.776	0.764	0.770	0.782	0.766	0.849	0.951	-

**Table A3.** Forecast the demand energy (MWh).

Month	2024	2025	2026	2027	2028	2029	2030
January	45,223.43	46,442.76	47,464.77	48,527.26	49,613.77	50,724.61	51,860.32
February	44,885.43	46,019.91	47,037.97	48,090.97	49,167.71	50,268.56	51,394.06
March	46,702.57	47,694.36	48,753.33	49,844.79	50,960.80	52,101.80	53,268.34
April	45,577.83	46,511.62	47,546.96	48,611.44	49,699.84	50,812.61	51,950.29
May	44,254.19	45,233.10	46,241.76	47,277.05	48,335.57	49,417.79	50,524.24
June	43,446.27	44,366.31	45,356.86	46,372.36	47,410.62	48,472.13	49,557.41
July	43,354.12	44,422.76	45,415.42	46,432.24	47,471.84	48,534.72	49,621.40
August	44,459.96	45,466.72	46,483.31	47,524.05	48,588.10	49,675.97	50,788.20
September	45,265.66	46,200.20	47,233.62	48,291.16	49,372.38	50,477.82	51,608.00
October	46,496.98	47,352.26	48,411.76	49,495.68	50,603.87	51,736.88	52,895.26
November	46,648.16	47,442.90	48,504.64	49,590.64	50,700.96	51,836.14	52,996.74
December	46,200.65	47,209.60	48,266.27	49,346.94	50,451.80	51,581.41	52,736.30

**Table A4.** Forecast the GDP (thousand of R\$).

Month	2024	2025	2026	2027	2028	2029	2030
January	894,130.7	976,779.9	1,065,062.3	1,160,704.3	1,264,826.2	1,378,243.9	1,501,813.5
February	898,594.8	975,965.6	1,063,101.4	1,158,260.6	1,262,037.6	1,375,153.7	1,498,425.1
March	972,345.1	1,045,097.5	1,139,144.1	1,241,412.9	1,352,764.2	1,474,063.1	1,606,221.8
April	949,898.1	1,024,401.4	1,115,918.2	1,215,828.3	1,324,772.7	1,443,515.7	1,572,916.8
May	956,952.5	1,035,981.0	1,129,154.4	1,230,504.7	1,340,868.8	1,461,097.3	1,592,092.1
June	966,061.6	1,050,497.1	1,144,394.5	1,246,874.4	1,358,608.9	1,480,388.1	1,613,096.0
July	991,411.8	1,080,793.6	1,177,950.5	1,283,661.4	1,398,785.3	1,524,203.6	1,660,854.9
August	993,631.4	1,079,827.0	1,176,389.2	1,281,751.8	1,396,619.0	1,521,808.2	1,658,230.4
September	972,186.6	1,057,416.6	1,152,433.2	1,255,838.2	1,368,460.2	1,491,156.9	1,624,844.3
October	1,016,733.6	1,106,198.2	1,205,156.1	1,313,110.6	1,430,794.5	1,559,049.6	1,698,811.3
November	1,024,990.2	1,115,216.2	1,215,391.6	1,324,431.4	1,443,198.9	1,572,594.3	1,713,581.9
December	1,038,563.0	1,132,211.8	1,233,529.5	1,344,038.9	1,464,500.1	1,595,778.8	1,738,834.1

## References

- Ang, T.Z.; Salem, M.; Kamarol, M.; Das, H.S.; Nazari, M.A.; Prabakaran, N. A comprehensive study of renewable energy sources: Classifications, challenges and suggestions. *Energy Strategy Rev.* **2022**, *43*, 100939. [\[CrossRef\]](#)
- Weschenfelder, F.; de Novaes Pires Leite, G.; Araújo da Costa, A.C.; de Castro Vilela, O.; Ribeiro, C.M.; Villa Ochoa, A.A.; Araújo, A.M. A review on the complementarity between grid-connected solar and wind power systems. *J. Clean. Prod.* **2020**, *257*, 120617. [\[CrossRef\]](#)
- Brazilian Energy Research Company National Energy Balance. 2023. Available online: <https://www.epe.gov.br/pt/publicacoes-dados-abertos/publicacoes/balanco-energetico-nacional-2023> (accessed on 1 April 2024).
- Yasmeen, R.; Yao, X.; Ul Haq Padda, I.; Shah, W.U.H.; Jie, W. Exploring the role of solar energy and foreign direct investment for clean environment: Evidence from top 10 solar energy consuming countries. *Renew. Energy* **2022**, *185*, 147–158. [\[CrossRef\]](#)
- Shyu, C.W. A framework for ‘right to energy’ to meet UN SDG7: Policy implications to meet basic human energy needs, eradicate energy poverty, enhance energy justice, and uphold energy democracy. *Energy Res. Soc. Sci.* **2021**, *79*, 102199. [\[CrossRef\]](#)
- Caldeira, A.A.; Wilbert, M.D.; Moreira, T.B.S.; Serrano, A.L.M. Brazilian State debt sustainability: An analysis of net debt and primary balance; [Sustainability of Brazilian state debt: An analysis of the relationship between net debt and primary result]; [Sustainability of the debt of the Brazilian states: An analysis of the net debt and the primary balance]. *Public Adm. Mag.* **2016**, *50*, 285–306. [\[CrossRef\]](#)
- Ahmad, T.; Chen, H. A review on machine learning forecasting growth trends and their real-time applications in different energy systems. *Sustain. Cities Soc.* **2020**, *54*, 102010. [\[CrossRef\]](#)
- Serrano, A.L.M.; Rodrigues, G.A.P.; Martins, P.H.d.S.; Saiki, G.M.; Filho, G.P.R.; Gonçalves, V.P.; Albuquerque, R.d.O. Statistical Comparison of Time Series Models for Forecasting Brazilian Monthly Energy Demand Using Economic, Industrial, and Climatic Exogenous Variables. *Appl. Sci.* **2024**, *14*, 5846. [\[CrossRef\]](#)
- Bispo, G.D.; Vergara, G.F.; Saiki, G.M.; Martins, P.H.d.S.; Coelho, J.G.; Rodrigues, G.A.P.; Oliveira, M.N.d.; Mosquera, L.R.; Gonçalves, V.P.; Neumann, C.; et al. Automatic Literature Mapping Selection: Classification of Papers on Industry Productivity. *Appl. Sci.* **2024**, *14*, 3679. [\[CrossRef\]](#)
- Arnob, S.S.; Arefin, A.I.M.S.; Saber, A.Y.; Mamun, K.A. Energy Demand Forecasting and Optimizing Electric Systems for Developing Countries. *IEEE Access* **2023**, *11*, 39751–39775. [\[CrossRef\]](#)
- Lins, M.E.; Oliveira, L.B.; Da Silva, A.C.M.; Rosa, L.P.; Pereira, A.O., Jr. Performance assessment of alternative energy resources in Brazilian power sector using data envelopment analysis. *Renew. Sustain. Energy Rev.* **2012**, *16*, 898–903. [\[CrossRef\]](#)
- Costa, M.A.; Salvador, C.V.M.; da Silva, A.V. Stochastic data envelopment analysis applied to the 2015 Brazilian energy distribution benchmarking model. *Decis. Anal. J.* **2022**, *3*, 100061. [\[CrossRef\]](#)

13. Camioto, F.d.C.; Rebelatto, D.A.d.N.; Rocha, R.T. Energy efficiency analysis of BRICS countries: A study using Data Envelopment Analysis. *G&P* **2015**, *23*, 192–203.
14. Chaturvedi, S.; Rajasekar, E.; Natarajan, S.; McCullen, N. A comparative assessment of SARIMA, LSTM RNN and Fb Prophet models to forecast total and peak monthly energy demand for India. *Energy Policy* **2022**, *168*, 113097. [[CrossRef](#)]
15. Rafayal, S.; Cevik, M.; Kici, D. An empirical study on probabilistic forecasting for predicting city-wide electricity consumption. In Proceedings of the 35th Canadian Conference on Artificial Intelligence, Toronto, ON, Canada, 30 May–3 June 2022.
16. Riady, S.R.; Apriani, R. Multivariate time series with Prophet Facebook and LSTM algorithm to predict the energy consumption. In Proceedings of the 2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE), Jakarta, Indonesia, 16 February 2023; pp. 805–810. [[CrossRef](#)]
17. Al-Haija, Q.A.; Mohamed, O.; Elhaija, W.A. Predicting global energy demand for the next decade: A time-series model using nonlinear autoregressive neural networks. *Energy Explor. Exploit.* **2023**, *41*, 1884–1898. [[CrossRef](#)]
18. Thangavel, A.; Govindaraj, V. Forecasting Energy Demand Using Conditional Random Field and Convolution Neural Network. *Elektron. Elektrotehnika* **2022**, *28*, 12–22. [[CrossRef](#)]
19. Gundu, V.; Simon, S.P. Pso-lstm for short term forecast of heterogeneous time series electricity price signals. *J. Ambient. Intell. Humaniz. Comput.* **2021**, *12*, 2375–2385. [[CrossRef](#)]
20. Verwiebe, P.A.; Seim, S.; Burges, S.; Schulz, L.; Müller-Kirchenbauer, J. Modeling Energy Demand—A Systematic Literature Review. *Energies* **2021**, *14*, 7859. [[CrossRef](#)]
21. Sengar, S.; Liu, X. Ensemble approach for short term load forecasting in wind energy system using hybrid algorithm. *J. Ambient. Intell. Humaniz. Comput.* **2020**, *11*, 5297–5314. [[CrossRef](#)]
22. Gujarati, D.N.; Porter, D.C. *Econometria Básica-5*; Amgh Editora: Porto Alegre, Brazil, 2011.
23. Martinello, L.M.; Rodrigues, S.B.; Hickmann, T.; Corrêa, J.M.; Teixeira, L.L. Comparative study between ARIMA and ETS forecasting models for temporal data on milk production in Brazil. *J. Inst. Laticín. Cândido Tostes* **2021**, *76*, 12–27. [[CrossRef](#)]
24. Kontopoulou, V.I.; Panagopoulos, A.D.; Kakkos, I.; Matsopoulos, G.K. A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks. *Future Internet* **2023**, *15*, 255. [[CrossRef](#)]
25. Dimri, T.; Ahmad, S.; Sharif, M. Time series analysis of climate variables using seasonal ARIMA approach. *J. Earth Syst. Sci.* **2020**, *129*, 149. [[CrossRef](#)]
26. Shadab, A.; Ahmad, S.; Said, S. Spatial forecasting of solar radiation using ARIMA model. *Remote. Sens. Appl. Soc. Environ.* **2020**, *20*, 100427. [[CrossRef](#)]
27. Yamak, P.T.; Yujian, L.; Gadosey, P.K. A Comparison between ARIMA, LSTM, and GRU for Time Series Forecasting. In Proceedings of the Association for Computing Machinery, ACAI '19, New York, NY, USA, 7 February 2020; pp. 49–55. [[CrossRef](#)]
28. Rundo, F.; Trenta, F.; di Stallo, A.L.; Battiato, S. Machine Learning for Quantitative Finance Applications: A Survey. *Appl. Sci.* **2019**, *9*, 5574. [[CrossRef](#)]
29. Zhu, W.; Zhang, R.; Liu, H.; Xin, L.; Zhong, J.; Zhang, H.; Qi, J.; Wang, Y.; Zhu, Z. Prediction of ionic liquid surface tension via a generalized interpretable Structure-Surface Tension Relationship model. *AIChE J.* **2024**, *11*, e18558. [[CrossRef](#)]
30. Zhou, K.; Wang, W.Y.; Hu, T.; Wu, C.H. Comparison of Time Series Forecasting Based on Statistical ARIMA Model and LSTM with Attention Mechanism. *J. Phys. Conf. Ser.* **2020**, *1631*, 012141. [[CrossRef](#)]
31. Arunkumar, K.; Kalaga, D.V.; Mohan Sai Kumar, C.; Kawaji, M.; Brenza, T.M. Comparative analysis of Gated Recurrent Units (GRU), long Short-Term memory (LSTM) cells, autoregressive Integrated moving average (ARIMA), seasonal autoregressive Integrated moving average (SARIMA) for forecasting COVID-19 trends. *Alex. Eng. J.* **2022**, *61*, 7585–7603. [[CrossRef](#)]
32. Rosano-Pena, C.; De Almeida, C.A.R.; Rodrigues, E.C.C.; Serrano, A.L.M. Spatial dependency of eco-efficiency of agriculture in São Paulo. *Braz. Bus. Rev.* **2020**, *17*, 328–343. [[CrossRef](#)]
33. Marques Serrano, A.L.; Saiki, G.M.; Rosano-Penã, C.; Rodrigues, G.A.P.; Albuquerque, R.d.O.; García Villalba, L.J. Bootstrap Method of Eco-Efficiency in the Brazilian Agricultural Industry. *Systems* **2024**, *12*, 136. [[CrossRef](#)]
34. Saiki, G.; Serrano, A.; Rodrigues, G.; Rosano, C.; Pomper Mayer, F.; Albuquerque, P. An Analysis of the Eco-Efficiency of the Agricultural Industry in the Brazilian Amazon Biome. *Sustainability* **2024**, *16*, 5731. [[CrossRef](#)]
35. Simar, L.; Wilson, P.W. Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Models. *Manag. Sci.* **1998**, *44*, 49–61. [[CrossRef](#)]
36. Stošić, B.D.; de Sousa, M.d.C.S. Jackstrapping DEA scores for robust efficiency measurement. *An. XXV Encontro Bras. Econom. SBE* **2003**, *23*, 1525–1540.
37. Saiki, G.M. Sustainable Development of the Brazilian Energy Sector. 2024. Available online: [https://figshare.com/articles/dataset/Non-Parametric\\_Methods\\_Application\\_and\\_Forecast\\_Model\\_Towards\\_Sustainable\\_Development\\_of\\_the\\_Brazilian\\_Energy\\_Sector/25714932/?file=45984321](https://figshare.com/articles/dataset/Non-Parametric_Methods_Application_and_Forecast_Model_Towards_Sustainable_Development_of_the_Brazilian_Energy_Sector/25714932/?file=45984321) (accessed on 15 October 2024).
38. Farrell, M.J. The Measurement of Productive Efficiency. *J. R. Stat. Soc. Ser. A* **1957**, *120*, 253–281. [[CrossRef](#)]
39. Charnes, A.; Cooper, W.; Rhodes, E. Measuring the efficiency of decision making units. *Eur. J. Oper. Res.* **1978**, *2*, 429–444. [[CrossRef](#)]
40. Chen, Y.; Yin, G.; Liu, K. Regional differences in the industrial water use efficiency of China: The spatial spillover effect and relevant factors. *Resour. Conserv. Recycl.* **2021**, *167*, 105239. [[CrossRef](#)]
41. Bogetoft, P.; Otto, L. *Benchmarking with Dea, Sfa, and R*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2010; Volume 157.

42. Matsumoto, K.; Chen, Y. Industrial eco-efficiency and its determinants in China: A two-stage approach. *Ecol. Indic.* **2021**, *130*, 108072. [CrossRef]
43. Boczar, P.; Błażejczyk-Majka, L. Economic Efficiency versus Energy Efficiency of Selected Crops in EU Farms. *Resources* **2024**, *13*, 123. [CrossRef]
44. Liu, X.; Guo, P.; Guo, S. Assessing the eco-efficiency of a circular economy system in China's coal mining areas: Emery and data envelopment analysis. *J. Clean. Prod.* **2019**, *206*, 1101–1109. [CrossRef]
45. Sala-Garrido, R.; Mocholi-Arce, M.; Maziotis, A.; Molinos-Senante, M. The carbon and production performance of water utilities: Evidence from the English and Welsh water industry. *Struct. Chang. Econ. Dyn.* **2023**, *64*, 292–300. [CrossRef]
46. Antunes, J.J.M.; Neves, J.C.; Elmor, L.R.C.; Araujo, M.F.R.D.; Wanke, P.F.; Tan, Y. A new perspective on the U.S. energy efficiency: The political context. *Technol. Forecast. Soc. Change* **2023**, *186*, 122093. [CrossRef]
47. Martín-Gamboa, M.; Iribarren, D. Dynamic Ecocentric Assessment Combining Emery and Data Envelopment Analysis: Application to Wind Farms. *Resources* **2016**, *5*, 8. [CrossRef]
48. Sousa, M.D.C.D.; Stosic, B. Technical Efficiency of the Brazilian Municipalities. *J. Product. Anal.* **2005**, *24*, 157–181. [CrossRef]
49. Wilson, P.W. FEAR: A software package for frontier efficiency analysis with R. *Socio-Econ. Plan. Sci.* **2008**, *42*, 247–254. [CrossRef]
50. Alsharif, M.H.; Younes, M.K.; Kim, J. Time Series ARIMA Model for Prediction of Daily and Monthly Average Global Solar Radiation: The Case Study of Seoul, South Korea. *Symmetry* **2019**, *11*, 240. [CrossRef]
51. Siami-Namini, S.; Tavakoli, N.; Namin, A.S. A Comparative Analysis of Forecasting Financial Time Series Using ARIMA, LSTM, and BiLSTM. *arXiv* **2019**, arXiv:1911.09512. [CrossRef]
52. Da Silva, A.V.; de Oliveira Ribeiro, C.; Rego, E.E. A data envelopment analysis approach to measuring socio-economic efficiency due to renewable energy sources in Brazilian regions. *Braz. J. Chem. Eng.* **2023**. [CrossRef]
53. Brazilian Institute for Applied Economic Research. Electricity 2024. Analysis and Forecast to 2026. 2023. Available online: <https://www.ipea.gov.br/portal/categorias/45-todas-as-noticias/noticias/14833-ipea-preve-crescimento-de-3-2-do-pib-neste-ano-e-mantem-em-2-0-a-estimativa-para-2024> (accessed on 25 April 2024).
54. Lucena, A.F.; Hejazi, M.; Vasquez-Arroyo, E.; Turner, S.; Köberle, A.C.; Daenzer, K.; Rochedo, P.R.; Kober, T.; Cai, Y.; Beach, R.H.; et al. Interactions between climate change mitigation and adaptation: The case of hydropower in Brazil. *Energy* **2018**, *164*, 1161–1177. [CrossRef]
55. Lucena, A.F.; Clarke, L.; Schaeffer, R.; Szklo, A.; Rochedo, P.R.; Nogueira, L.P.; Daenzer, K.; Gurgel, A.; Kitous, A.; Kober, T. Climate policy scenarios in Brazil: A multi-model comparison for energy. *Energy Econ.* **2016**, *56*, 564–574. [CrossRef]
56. PROGRAM, E.T. Carbon Neutrality 2050: Scenarios for an Efficient Transition in Brazil. 2023. Available online: [https://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-726/PTE\\_RelatorioFinal\\_EN\\_5JUN.pdf](https://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-726/PTE_RelatorioFinal_EN_5JUN.pdf) (accessed on 9 October 2024).
57. of Economy, M. Brazil's Green Monitor. 2022. Available online: [https://www.gov.br/mdic/pt-br/assuntos/assuntos-economicos-internacionais/acompanhamento-economico/brazil-green-monitor/brazil\\_green\\_monitor-2022-04.pdf](https://www.gov.br/mdic/pt-br/assuntos/assuntos-economicos-internacionais/acompanhamento-economico/brazil-green-monitor/brazil_green_monitor-2022-04.pdf) (accessed on 4 September 2024).
58. Gugler, K.; Haxhimusa, A.; Liebensteiner, M. Effectiveness of climate policies: Carbon pricing vs. subsidizing renewables. *J. Environ. Econ. Manag.* **2021**, *106*, 102405. [CrossRef]
59. Rituraj, R.; Ecker, D.; Annamaria, V.K. Smart and Sustainable Grids Using Data-Driven Methods; Considering Artificial Neural Networks and Decision Trees. In Proceedings of the 2022 IEEE 20th Jubilee International Symposium on Intelligent Systems and Informatics (SISY), Subotica, Serbia, 15–17 September 2022; pp. 409–416. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.