# ANALYZING THE DELAYED EFFECTS OF NATURAL GAS PRICES ON ELECTRICITY PRICING IN THE TURKISH DAY-AHEAD MARKET AFTER THE PANDEMIC

A Thesis

by

Betül Elif Özmen

Submitted to the
Graduate School of Sciences and Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Master of Science

in the
Department of Data Science

Özyeğin University
June 2024

# ANALYZING THE DELAYED EFFECTS OF NATURAL GAS PRICES ON ELECTRICITY PRICING IN THE TURKISH DAY-AHEAD MARKET AFTER THE PANDEMIC

Approved by:

Asst. Prof. Göktürk Poyrazoğlu, Advisor
Dept. of Electrical & Electronics Eng.
*Özyeğin University*

Assoc. Prof. Dr. Elvin Çoban Göktürk
Dept. of Industrial Eng.
*Özyeğin University*

Asst. Prof. Emre Çelebi
Dept. of Industrial Eng.
*Yeditepe University*

Date Approved: May 22, 2024

*To my beloved family..*

# ABSTRACT

This study investigates the correlation between natural gas and electricity prices in Türkiye, which has experienced significant spikes in natural gas prices due to the pandemic and regional conflicts. The determination of reference prices for electricity and natural gas in Türkiye is facilitated by transparent markets operated by the Energy Exchange. Using data from the EXIST transparency platform, this research employs regression techniques to determine the temporal relationship between abrupt increases in natural gas prices and their impact on electricity prices. Prior studies have shown that elevated natural gas prices affect electricity prices with a delay, depending on the proportion of natural gas in a country's energy supply. This research applies a similar methodology to assess the time-lag effect of natural gas prices on electricity prices within the Turkish energy market post-COVID-19. Metrics such as mean absolute error, mean square error, mean absolute percent error, and mean percent error validate the findings. The results demonstrate a eightyfive-day time lag in the correlation between natural gas and electricity prices in Türkiye. This delay improves forecasting accuracy by up to 16.17% in terms of error reduction.

# ÖZETÇE

Bu çalışmada, pandemi ve bölgesel çatışmalar nedeniyle doğal gaz fiyatlarında önemli artışlar yaşayan Türkiye'de doğal gaz ve elektrik fiyatları arasındaki ilişki incelenmektedir. Türkiye'de elektrik ve doğal gaz için referans fiyatların belirlenmesi, Enerji Borsası tarafından işletilen şeffaf piyasalar aracılığıyla kolaylaştırılmaktadır. Bu araştırma, EPİAŞ şeffaflık platformundan elde edilen verileri kullanarak, doğal gaz fiyatlarındaki ani artışların elektrik fiyatları üzerindeki etkileri arasındaki zamansal ilişkiyi belirlemek için regresyon teknikleri kullanmaktadır. Daha önce yapılan çalışmalar, doğal gaz fiyatlarındaki artışın, bir ülkenin enerji arzındaki doğal gaz oranına bağlı olarak elektrik fiyatlarını gecikmeli olarak etkilediğini göstermiştir. Bu araştırma, COVID-19 sonrası Türkiye enerji piyasasında doğal gaz fiyatlarının elektrik fiyatları üzerindeki zaman gecikmeli etkisini değerlendirmek için benzer bir metodoloji uygulamaktadır. Ortalama mutlak hata, ortalama kare hata, ortalama mutlak yüzde hata ve ortalama yüzde hata gibi ölçütler, bulguları doğrulamaktadır. Sonuçlar, Türkiye'de doğal gaz ve elektrik fiyatları arasındaki korelasyonda seksen beş günlük bir gecikme olduğunu göstermektedir. Bu gecikme, hata azaltma açısından tahmin doğruluğunu %16,17'ye kadar artırmaktadır.

# ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor, Göktürk Poyrazoğlu, for their invaluable guidance, insightful feedback, and constant encouragement, which shaped this work and helped me navigate challenges. I am deeply grateful to my father Atilla Özmen, my mother İnci Özmen, and my sister Cemre Zeynep Özmen for their unconditional love and patience throughout this journey, motivating me to persevere during difficult moments. I thank my friends, relatives, colleagues, and the people I love who have shared this journey, offering moral support, stimulating discussions, and a sense of community. This work is dedicated to all those affected by the harsh realities of war and the painful losses caused by natural disasters in our country and around the world. It is also devoted to those combating injustices faced by women in the workplace and in daily life. My greatest hope is to see a world where peace prevails, natural and human-made tragedies are overcome, and justice and equality reign supreme.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# CHAPTER I

# INTRODUCTION

Natural gas and electricity play a crucial role in the energy sector of Türkiye. The Türkiye Energy Exchange (EXIST) facilitates transparent marketplaces that create reference prices for these commodities. Türkiye heavily depends on imported natural gas to meet its energy needs due to its limited domestic sources. There are three main techniques used to achieve this importation: long-term contracts for pipeline transportation, prolonged contracts for liquefied natural gas, and purchases from the spot market. The nation currently possesses substantial reserves of imported natural gas and has intentions to substantially increase its storage capacity by the conclusion of 2023.

In 2022, the main sources of electricity generation in Türkiye were coal, natural gas, hydroelectric power, wind, solar, geothermal energy, and other sources. Although coal continues to be a major source of energy, natural gas also plays a vital role as a principal fuel for generating electricity, fulfilling a significant amount of the country's energy requirements. The price of electricity is heavily influenced by the cost of natural gas. The volatility in natural gas prices has a direct impact on the expenses related to power generation, highlighting the interconnected nature of these vital energy resources.

The electrical day-ahead market (DAM) operates via a uniform-price auction mechanism, where bids are submitted for the next day's supply and demand. The participants are required to submit their bids for supply and demand by the next day. The market operator uses advanced optimization models that depend on preset bid structures to achieve a balance between supply and demand. The objective of the auction process is to ascertain the price at which the market is cleared for each hour. The auction method ensures that all offers that are successfully matched are settled at a consistent price, which encourages participants to engage in effective bidding strategies.

Natural gas power facilities frequently exhibit higher marginal costs compared to alternative energy sources, which they leverage to mitigate supply and demand volatility. Therefore, variations in natural gas prices can impact the incremental cost of power plants that employ gas as a fuel source, thereby influencing electricity costs. The recent global pandemic and a regional conflict in Western Europe have significantly impacted the equilibrium between energy supply and demand. Consequently, this has resulted in abrupt and unpredictable surges in natural gas prices worldwide, with Europe experiencing particularly pronounced effects.

The physical transportation of natural gas and electricity in market contexts reveals several notable distinctions due to the unique characteristics of these commodities. The electricity markets operate on abbreviated temporal scales, with day-ahead and intraday markets playing critical roles in maintaining grid stability. Pipelines or liquefied natural gas carriers primarily facilitate the transportation of natural gas, ensuring its efficient delivery. As a result, natural gas market prices may not accurately reflect the commodity's immediate delivery price. The temporal discrepancy between natural gas supply and electricity generation plays a significant role in influencing electricity prices. In response to these challenges, Türkiye has implemented measures to liberalize regulations related to determining reference prices for electricity and natural gas. These regulations are now overseen by EXIST, which operates transparent markets.

A prior study examined the impact of price interdependence, specifically focusing on the disparity between the timing of natural gas delivery to power plants and the determination of its corresponding price. The study revealed that the initial rise in natural gas prices did not immediately impact electricity prices in Türkiye. However, the study noted an upward trend in electricity prices approximately two weeks later.

This study examines the relationship between natural gas and day-ahead market electricity prices in the post-pandemic market using an analogous hypothesis. A comprehensive investigation is performed to investigate the variations in natural gas prices and determine the precise manner and timing in which electricity prices started to be influenced. This fresh examination is based on past scholarly research and incorporates the established parameters identified in that particular study.

The training set for this study uses data from the EXIST data transparency platform, covering the period from 2020 to 2022. Data was collected during the final two weeks of December 2022, and prices for the same period were projected to create the test set. The parameters identified in the previous study, including time, day, week, weekend, holiday, and demand data, serve as the foundation for the training dataset. The accuracy of the model is assessed using error metrics such as Mean Absolute Error (MAE), Mean Square Error (MSE), Mean Absolute Percent Error (MAPE), Mean Percent Error (MPE), $R^2$ Score (R-Kare), and Root Mean Squared Error (RMSE).

## 1.1 Problem Statement

The global surge in natural gas prices following the COVID-19 pandemic and regional geopolitical tensions had a profound impact on numerous nations around the world. As one of the nations in question, Türkiye has encountered a significant phenomenon that necessitates a comprehensive comprehension of the correlation between electricity and natural gas prices. The assessment of the temporal and methodological aspects of the impact of volatile fluctuations in natural gas prices on electricity prices is critical for the energy market's future trajectory.

Prior studies have demonstrated a temporal delay between fluctuations in natural gas prices and their subsequent impact on electricity prices. The duration of this time delay varies based on the proportion of a country's natural gas supply. The issue at hand holds significant implications for energy-importing nations like Türkiye, as the degree of foreign dependency plays a crucial role in shaping the economic repercussions of energy price fluctuations.

## *1.2 Motivation*

The main aim of this study is to thoroughly investigate and evaluate the long-term effect of natural gas prices on electricity costs in the Turkish energy market after the COVID-19 epidemic. The objective of this research is to address deficiencies and constraints in previous studies regarding the topic. The objective of this study is to examine the influence of natural gas prices on electricity costs in the current energy market conditions of Türkiye. The association between natural gas and power prices will be determined by analyzing EXIST's transparency platform data and implementing regression methods.

This research endeavor aims to enhance comprehension of Türkiye's energy policies and market regulations while assessing the impact of energy price fluctuations on economic outcomes. Additionally, this study will serve as a crucial foundation for devising strategies to effectively regulate energy prices and safeguard energy stability in Türkiye, given its status as a nation reliant on energy imports.

## *1.3 Scope of Work*

This research examines the correlation between natural gas and electricity prices within the Turkish energy market, specifically focusing on 2020–2022. The study aims to analyze the influence of the pandemic and regional conflicts on these prices. The present study aims to construct a linear model utilizing the Least Square Error Method, to minimize the total sum of squared deviations between the observed electricity prices and the electricity prices that are to be estimated. The linear model will be employed to conduct multiple linear regression analyses.

This study will primarily utilize the Least Square Error Method and the intricacies of multiple linear regression analysis. This thesis introduces a research framework aimed at comprehending and assessing the consequences of post-pandemic and regional conflict-induced natural gas price hikes on electricity prices within Türkiye's intraday energy market.

The study aims to delve into several key areas: Firstly, it will explore the theoretical basis of the Least Square Error Method and linear regression analysis. Secondly, the process of determining the appropriate linear model and selecting relevant variables will be addressed. Thirdly, it entails data preparation and linear model construction. Finally, the study will assess the model's performance and analyze various error metrics.

The present study aims to investigate the impact of pandemics and regional conflicts on the energy market, as well as the delayed influence of fluctuations in natural gas prices on electricity prices. The thesis encompasses a broader range of subjects than those mentioned and considers various significant factors to ensure the results accurately represent the study's objectives and scope. Its primary objective is to enhance understanding of the correlation between natural gas and electricity prices within Türkiye's energy market while providing valuable data to inform future energy policies.

# CHAPTER II

# LITERATURE REVIEW

The literature review in this chapter reveals the use of regression, multiple regression methods, time series, artificial neural networks, and deep learning in examining the impact of natural gas or any commodity on electricity prices. Studies in the energy markets of other countries were also examined, with the exception of the Turkish market. It has yet to be determined whether the effect of natural gas prices on electricity prices is directly examined. This study examined the continuation of the previous study and the effect of affected natural gas prices on electricity prices after the post-pandemic and regional conflicts worldwide.

The reason why multiple linear regression, one of the statistical and machine learning methods, was used in this study is that since there is more than one feature that affects the electricity price and it is known that there is a linear relationship between electricity prices and natural gas prices, the multiple linear regression method was used.

The literature review reveals the use of various statistical and machine learning models to analyze the impact of commodities, such as natural gas, on electricity prices. The evaluation of studies on different countries' energy markets reveals a lack of full investigation into the effect of natural gas prices on electricity prices. This section includes many studies focusing on the importance of electricity price forecasting. Price prediction studies have included many applications. When the energy sector and academic studies are examined, the studies used for price prediction in Türkiye and other country markets are shared below.For electricity price predictions, researchers generally used methods such as the impact of multiple regression and supervised learning algorithms, time-series forecasting techniques, artificial neural networks and hybrid models, deep learning and machine learning models, and other statistical and machine learning approaches.

The article notes that most electricity price predictions focus on the short term rather than the medium or long term, likely due to the electricity market's volatility[1].

In the Spanish market, electricity prices are analyzed using a stochastic approach, considering the impact of COVID-19. The estimation includes natural gas prices, comparing the actual prices in the first half of 2020 with those expected at the end of 2019. This period aligns with the peak impact of COVID-19 on Spanish society. A stochastic model with deterministic and stochastic components determined the expected prices and their future distributions. The latter incorporates mean-reversion and jumps [2].

Electricity prices exhibit high volatility. To address this, a new hybrid machine learning method was proposed that incorporates Automatic Relevance Determination with linear regression and Extra Tree Regression models using partition packing. The model's performance was evaluated through error metrics like MAPE. Robust to the unique characteristics of electricity prices, such as high volatility, rapid spikes, and seasonality, robust machine learning prediction tools are essential. This hybrid approach aims to effectively forecast day-ahead electricity prices[3].

A hybrid model was employed to forecast electricity prices in the Spanish electricity market. This model combines various methods like logistic regression for rare events, decision trees, and a market equilibrium model[4].

The Italian electricity market employs ensemble-based learning (ARMA) for electricity price prediction, highlighting the challenges of forecasting due to the specific characteristics of electricity price series. The study evaluates the performance of an ensemble-based technique for forecasting short-term electricity spot prices in the Italian electricity market[5].

A hybrid model combining the results of multiple linear regression, autoregressive integrated moving average, and Holt-Winters models was used to forecast electricity prices in the Iberian electricity market[6].

In the Turkish electricity market, various methods predict electricity prices, with model performance evaluated through metrics like MAPE. Advanced AI models such as LSTM and CNN-LSTM leverage their analytical capabilities to uncover hidden patterns in time series data. They forecast hourly prices seven days ahead using data from EXIST, where natural gas is a significant exogenous variable in Türkiye's energy generation[7].

In the Turkish electricity market, advanced regression techniques have been utilized to estimate electricity prices. Electricity demand is predicted hourly for the next day, and two simple forecast approaches, also used by grid operators, are compared against these advanced techniques[8].

In the Greek electricity market, forecasts for electricity prices have been developed using the least square error method to estimate model parameters. Similarly, electricity price predictions were also made for Yunnan, China. In this study, a novel grey prediction model, known as the interval GM(0, N) model, was proposed for medium-term electricity market clearing price forecasting. This model improves on the traditional GM(0, N) model by identifying lower and upper bounds for interval estimation, establishing a new whitenization method to determine the forecasting value from the interval, and using an improved particle swarm optimization instead of the least square method to identify parameters. Our thesis employed the LSE technique for determining model parameters[9].

An autoregressive model was utilized to estimate electricity prices in the Italian market. This approach includes a functional forecasting method to accurately predict electricity prices, with a functional autoregressive model of a specific order used for short-term price forecasting [10].

A methodology utilizing artificial neural networks (ANN) was developed for forecasting electricity prices. The study emphasized the importance of selecting and preparing fundamental data that significantly influences electricity prices, as the performance of an ANN model is heavily reliant on appropriate input parameters[11].

In the expanding electricity market, making accurate price predictions is increasingly challenging. With trade liberalization and market harmonization in European markets, accurate price forecasting is difficult for market participants because it necessitates considering features from an expanding array of coupled markets. People recognize LSTM models for their strong performance in managing nonlinear and complex problems when processing time-series data. However, due to our problem's nonlinear nature, LSTM was not used in our specific solution [12].

Four efficient models are proposed for estimating electricity prices in Victoria, Canada: decision tree regression, random forest regression, gradient boosting regression, and linear regression. These productive methods are designed to forecast highly accurate results[13].

This is an article prepared during the thesis study on the subject studied in this thesis. In the thesis, different and improved models were tested, and the results were shared[14].

The Turkish electricity market has been examined, and the inclusion of past natural gas prices in the prediction model was found to lower errors. In our study, we utilized the average of the last two weeks of natural gas prices from the natural gas market as a predictive estimator[15].

It was observed that incorporating the average of the last two weeks' natural gas prices into the forecasting model reduces the error from 15.85% to 14.31%. This suggests that current natural gas prices could impact the electricity market two weeks later, which influenced the approach of our work[16].

Firstly, For other DAMs in Europe and the USA, there are lots of studies in the literature that introduces parameters that have an impact on the electricity price formation; however, there is a limited study on the Turkish DAM price formation, so the first step in this study is to identify the parameters[17]. This article was used to determine the parameters, which is the first step regarding Turkish electricity prices.

Accurate electricity price forecasting has become crucial due to numerous influencing variables. Efficient modeling is vital for market participants to develop bidding strategies and make investment decisions. However, accurate forecasting is challenging because electricity prices are characterized by high volatility, seasonal patterns, calendar effects, and nonlinearity. A functional autoregressive model of a specific order is studied for short-term electricity price forecasting in the Italian market[18].

In the Turkish electricity market, a hybrid model that combines SARIMA, TBATS, and neural network models was developed for predicting hourly electricity prices. Given that time series data can include both linear and nonlinear patterns, this approach enhances forecasting accuracy[19].

Five Western European countries (France, Germany, Italy, Spain, and the United Kingdom) were selected for estimating electricity prices using ARIMA and GARCH models. The study analyzed six datasets of hourly price series exhibiting strong daily seasonality, applying prominent statistical models for time series analysis, such as ARIMA and different GARCH model versions[20].

The Turkish electricity market employs multiple linear regressions to forecast electricity prices, and this study has identified key features crucial for accurate predictions. The lagged electricity prices, including those from the previous day, the prior week, and lagged moving average prices, play a significant role in estimating electricity prices[21].

Electricity price forecasts across 14 European countries were analyzed to identify the most effective time-series forecasting methods[22].

A new time-series model has been developed to estimate electricity prices in the

Turkish market. This model considers the contributions of different resources to the aggregate supply curve, as well as the influence of external factors on supply prices and amounts. It includes a unified Monte Carlo methodology for latent variable tracking, forecasting, and hyperparameter estimation. Specifically, a sequential Markov chain Monte Carlo algorithm is used for tracking latent variables, enabling day-ahead supply curve forecasting[23].

In the Turkish electricity market, past natural gas prices were used as an input for forecasting electricity prices, with predictions made using the CatBoost method. A two-stage forecasting model was proposed, utilizing historical natural gas prices, electricity load forecasts, and past electricity price values. The empirical mode decomposition algorithm was used for decomposition, while the categorical boosting algorithm was employed to accurately predict day-ahead electricity prices. The model's performance was evaluated using error metrics such as MAPE[24].

In the UK market, LSTM and hybrid models were applied for price prediction. The study compares the performance of LSTM-based methods for single-stage and two-stage forecasting of electricity demands and prices. Initially, a reference single-stage LSTM model was created to predict electricity demands or prices based on their historical time series data. Then, the original demand and price series were processed to produce two new time series[25].

Researchers often develop hybrid models that combine statistical methods and artificial intelligence to predict electricity prices. Studies on electricity price prediction using artificial intelligence, particularly deep learning techniques, are gaining traction in the literature. This study employed multiple linear regression, a statistical and machine learning method, as multiple factors influence electricity prices and a linear relationship is known to exist between electricity and natural gas prices. This thesis aims to predict accurate results and reduce errors by comparing regression performances. Electricity price forecasting was analyzed using various metrics.

# CHAPTER III

# DEVELOPMENT OF ELECTRICITY MARKETS IN TURKIYE

MCP and DRP are subject to fluctuations influenced by various factors, including the accessibility and price of fuel, the amount of demand for the commodity, environmental regulations, and geopolitical events. The policies put in place by the nations involved in natural gas exports, the state of supply and demand, prevailing market circumstances, and other determining factors are just a few of the variables that can affect DRP. Oil prices substantially influence natural gas prices due to the frequent indexing of natural gas prices to oil prices. Nevertheless, it is worth noting that the MCP exhibits higher volatility and experiences more frequent fluctuations than the DRP. In contrast, the DRP demonstrates excellent stability and is less responsive to short-term shifts in supply and demand.

Moreover, there is a complex interconnection between MCP and DRP, whereby alterations in the natural gas price can impact the electricity price. As previously stated, natural gas assumes a significant part in the production of electricity across numerous nations, with the cost of electricity production exhibiting a strong correlation with the cost of natural gas. Consequently, an escalation in the price of natural gas results in a corresponding rise in electricity production costs, thereby causing an increase in electricity prices for consumers.

In recent years, substantial growth and advancement have been observed in the electricity and natural gas markets within the Turkish context. The Turkish Electricity Transmission Corporation (TEİAŞ) and the Turkish Natural Gas Pipeline Corporation (BOTAŞ) have been responsible for this development. In Türkiye, the transmission of electricity is under the jurisdiction of TEİAŞ, while the responsibility for the importation, storage, and transmission of natural gas lies with BOTAŞ.

Türkiye's electricity and natural gas markets have also undergone a liberalization process by the government. The government established the Energy Market Regulatory Authority (EMRA) in 2001 to oversee these requirements. EMRA oversees the pricing mechanisms for electricity and natural gas while upholding the imperative of maintaining a dependable supply.

The emergence and growth of Türkiye's electricity and natural gas markets have significantly influenced the economy. The fulfillment of needs has enhanced operational effectiveness within the energy sector, reducing end-user costs. Nevertheless, it is essential to acknowledge that the markets are currently in their nascent phases of development, thereby necessitating the resolution of specific challenges. One illustrative instance pertains to the existing concentration within markets, which indicates a discernible requirement for heightened competition.

The section dedicated to the Development of Electricity and Natural Gas Markets in Türkiye within this thesis will present a comprehensive analysis of these markets' historical progression, regulatory framework, and present condition. This section will additionally examine the challenges encountered by the needs and provide recommendations for effectively addressing these challenges.

## *3.1 Formation Of Energy Markets*

Energy markets are economic systems encompassing various activities associated with producing, distributing, and consuming energy resources. These markets serve as platforms for trading energy commodities among multiple stakeholders, including producers, suppliers, consumers, and intermediaries.

The typical composition of energy markets includes a variety of participants, each playing a distinct role. Suppliers are corporate entities that provide energy resources, representing the market's supply side. They are crucial in ensuring the availability of energy commodities. Producers, on the other hand, are tasked with generating electricity, natural gas, or other energy products. This is done through the processing of various energy resources, making them a key component in the energy supply chain.

Consumers in the energy market include both institutional entities and individuals. Their role is to procure electricity, natural gas, or other energy services, thereby fulfilling their energy-related needs. These consumers are the driving force behind the demand side of the market. Intermediaries are institutional entities that play an essential role in the energy market. They are responsible for facilitating transactions related to the buying and selling of energy. Acting as a bridge, these intermediaries connect various market participants, thereby enabling and streamlining interactions within the energy market. This collaborative ecosystem of suppliers, producers, consumers, and intermediaries forms the backbone of the energy markets, ensuring its efficient functioning and stability.

The fundamental factors influencing energy markets are the dynamic interaction between the availability of energy resources and the level of consumer demand. Supply is controlled by a multitude of factors that have an impact on both the quantity and price of energy resources. Several factors contribute to the determination of supply, such as energy generation, availability of natural resources, production techniques, and political factors. These determinants have a substantial impact on the supply. Demand can be defined as the inherent necessity for energy products and the quantity of consumption. The demand is subject to many factors, including but not limited to economic growth, population growth, climatic conditions, and energy efficiency policies.

The determination of prices in energy markets is contingent upon establishing an equilibrium between the quantity of energy supplied and the quantity demanded. Determining MCP is commonly influenced by the marginal cost, which pertains to the cost incurred in the production or importation of a single unit, and the regional clearing price. Market equilibrium is fundamental to addressing discrepancies between the quantity supplied and the quantity demanded within energy markets. Within this specific procedure exists a dynamic exchange of energy among multiple participants within the market, with the ultimate goal of attaining a state of equilibrium to the price at which the market is ultimately cleared.

## *3.2 Development Of Electricity Markets In Türkiye*

Electricity markets play an essential role in the energy sector and are critical to countries' economic development and energy needs. Türkiye has focused on developing electricity markets through regulations and privatizations in the energy sector. This section examines the historical development of electricity markets in Türkiye and its key milestones.

### 3.2.1 Period Before Unregulated and Divided Market Structures

The development of electricity markets in Türkiye dates back to the period before the privatization and regulation processes began. The old structure had an unregulated and fragmented market structure. Public institutions and private companies managed energy production and distribution. During this period, the government set electricity prices, and there was no competitive environment at the consumer level.

The previous period of the Turkish electricity market, generally covering the 1980s and 1990s, had an unregulated and fragmented market structure for electricity production and distribution. The state's various institutions and private sector companies managed energy production and distribution activities. Public institutions such as the Ministry of Energy and Natural Resources and The Turkish Electricity Authority (TEK) produce and distr ibute energy. TEK combined all of Türkiye's electricity production and distribution activities under a single roof. Therefore, the management of the electricity market had a centralized and unregulated structure.

The Turkish Electricity Authority is shown in figure 1.

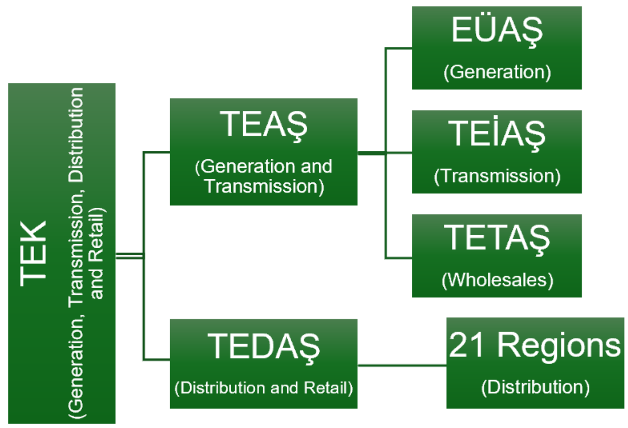


**Figure 1:** The Turkish Electricity Authority

The government set electricity prices according to the energy policies and strategies of the period. This prevented the formation of a competitive market structure and did not allow for price changes at the consumer level. There needed to be more incentives for the efficient and diversified use of energy resources and technologies. In addition, different institutions in different geographical regions carried out energy production and distribution. This led to a fragmented market structure and difficulties balancing energy supply and demand. The fragmented structure prevented the effective and coordinated supply of energy.

This unregulated and fragmented market structure caused Türkiye to experience significant problems regarding efficiency, competitiveness, and sustainability in the energy sector. Therefore, a new approach and policies must be adopted to transform and re-regulate the Turkish electricity market. This need triggered the transformation process that began with the Electricity Market Law 2001.

### 3.2.2 2001 Electricity Market Law and Transformation Process

The fundamental transformation of electricity markets in Türkiye began with the Electricity Market Law 2001. This law included regulations to liberalize the electricity market, promote competition, and initiate privatization processes. In particular, after the adoption of this law, production, transmission, distribution, and retail sales activities were separated, and privatization processes began.

The Electricity Market Law 2001 is essential in transforming the Turkish electricity market. This law was adopted on July 20, 2001, and entered into force on July 30, 2001. The law was prepared to liberalize and open the electricity market to competition. The main objectives of the law encompass creating a free and competitive electricity market, ending public monopolies by opening electricity production and distribution activities to private sector entities and safeguarding energy consumers' rights through regulations ensuring transparent pricing and access to electricity services.

The transformation process began with entering the Electricity Market Law 2001 into force. Various steps have been taken in this process to regulate and privatize the electricity market. The public enterprise TEK has transferred its electricity production and distribution activities to the private sector, and competition in the energy sector has increased. The process of developing electricity markets in Türkiye has been characterized by its lengthy and intricate nature. The nation has made notable advancements in liberalization and opening up the electricity market to foster competition. Nevertheless, specific challenges necessitate attention, including enhancing energy sector efficiency and mitigating the nation's dependence on imported energy resources.

### 3.2.3 The Pursuit of Liberalization and Privatization Initiatives

The implementation of liberalization and privatization measures expedited the process of transforming the electricity market. The establishment of the EMRA occurred to serve as a regulatory entity responsible for overseeing the liberalization process and facilitating market competition. The privatization of electricity generation plants and distribution companies facilitated their transfer to the private sector, resulting in the establishment of a competitive market structure. This stage holds significant importance in transforming the electricity market in Türkiye. The endeavors mentioned above were undertaken to privatize the energy sector and establish a competitive market structure. Below, a more comprehensive overview of Liberalization and Privatization Efforts is provided.

The liberalization process refers to the gradual opening up of economic, political, and social systems, often characterized by the reduction of Liberalization; it refers to the systematic opening up of the energy market to enable the participation of multiple energy companies and private sector entities in a competitive manner. This process occurred after a period where the state solely controlled the energy market. Liberalizing the Turkish energy market was initiated by enacting the Electricity Market Law in 2001. The process involved the Liberalization of electricity generation, transmission, distribution, and trading, allowing for the participation of private sector entities. This led to an escalation in competition within the energy sector.

Privatization endeavors involve the transfer of energy assets, which are currently under the ownership of public enterprises, to the private sector. The Privatization Administration Presidency is responsible for overseeing the privatization process of the electricity market in Türkiye. During this procedure, electricity generation and distribution facilities owned by public entities, such as the Turkish Electricity Authority (TEK), were divested and transferred to private sector investors, resulting in their privatization. This measure led to heightened competition within the energy industry and facilitated the improved operational efficacy of the market.

Licensing is a pivotal aspect of the liberalization process, requiring entities involved in energy production, trade, and distribution to obtain necessary permits. Oversight of this process falls under the jurisdiction of the EMRA, which evaluates companies based on predefined market operating criteria. Through this process, the EMRA ensures adherence to standards, fostering a dependable and consistent provision of services within the energy market.

The liberalization process aims to establish energy prices within a competitive market framework. The escalation in competition has transparently determined energy prices and provided services to consumers at more economically viable rates. In this procedure, the ability of energy companies to determine prices has been restricted, resulting in prices being determined based on prevailing market conditions. Exploration of Potential Investment Prospects: The implementation of liberalization and privatization initiatives has presented the energy sector with fresh prospects for investment. The availability of investing in privatized energy assets and licensed activities has significantly facilitated the growth of energy companies in the market and facilitated the entry of new participants into the sector. Furthermore, this has played a significant role in advancing the energy infrastructure and the subsequent market expansion.

Implementing Liberalization and Privatization Efforts has played a crucial role in transforming the electricity market in Türkiye. These endeavors have resulted in heightened competition within the energy industry, enhanced price visibility, and the establishment of a framework under prevailing market circumstances. Furthermore, implementing privatization initiatives has facilitated the emergence of fresh investment prospects within the energy industry, thereby bolstering the expansion of the market. Effective management of energy policies and maintaining a competitive market structure are crucial considerations for the future.

### 3.2.4 Competitive Landscape and Investment Prospects in the Privatized Electricity Market

The electricity market, which has undergone privatization, has implemented a competitive framework involving private entities' active involvement. The private sector has witnessed a notable surge in energy production and distribution investments, accompanied by the widespread adoption of novel technologies. The presence of a competitive market structure has facilitated a greater level of transparency and competition in the determination of energy prices.

The privatization process of the energy sector has resulted in the establishment of a market structure characterized by competition. The involvement of private sector investors in the privatization of electricity generation and distribution facilities has led to an augmentation of market competition. Private sector enterprises have commenced their operations in the market by acquiring licenses for energy production and trade. This phenomenon has fostered a competitive landscape for companies operating within the energy sector.

The competition facilitates transparent price determination, favoring consumers and enabling more affordable electricity services. The rise in competition within the electricity production and distribution sector catalyzes companies to enhance their operational efficiency, thereby leading to improved productivity. Simultaneously, consumers can reap the benefits of a more comprehensive array of service alternatives. Privatizing the electricity market has resulted in the emergence of fresh investment prospects within the energy sector. The allocation of investments towards privatized electricity assets and licensed activities has facilitated the expansion of energy companies and the entrance of new stakeholders into the sector.

Investment opportunities play a pivotal role in fostering energy infrastructure development, concurrently facilitating the expansion of energy production capacity.

New investments can redirect the energy sector towards sustainable and environmentally friendly technologies. The significance of investment opportunities in the privatized electricity market also extends to the realm of energy security. Implementing projects by multiple investors across various regions plays a significant role in enhancing the diversification of a nation's energy supply and bolstering the reliability of its energy provision.

In summary, the presence of competition and investment prospects within the privatized electricity market has played a significant role in fostering the expansion and advancement of the energy sector. The presence of a competitive market structure incentivizes companies to operate with greater efficiency, leading to the provision of services at more affordable prices for consumers. Simultaneously, investment opportunities are pivotal in fortifying the energy infrastructure and enhancing energy security. In the forthcoming period, the perpetuation of competitive dynamics and the emergence of novel investment prospects are poised to foster the progressive advancement of the electricity market in Türkiye.

### 3.2.5 Day-Ahead Market

The liberalization process in the electricity market began with the Electricity Market Law, aiming to establish a market founded on the principles of transparency, integrity, and competition, while also seeking integration with electricity markets of other countries. This strategic initiative was designed to cultivate a robust and dynamic electricity market landscape, encouraging active participation from market entities and ensuring effective management and monitoring of market operations. As part of this vision, a series of planned actions were executed, significantly advancing the development of a more dynamic and engaged electricity market.

A significant shift from a monopolistic model, which featured a single buyer and single seller, to a liberal and competitive framework marked the transformation. This transition commenced with the introduction of a monthly, three-period financial settlement system on July 1, 2006, followed by the implementation of the Day-Ahead Planning system on December 1, 2009. These stages were pivotal for the evolution of the market, providing stakeholders with valuable experiences and insights, thus laying the foundation for future market models.

A notable milestone was reached on December 1, 2011, with the establishment of the current Day-Ahead Market system, representing the most considerable leap towards the envisioned market structure. Initially, the day-ahead market was established and managed by TEİAŞ. In 2015, EXIST took over the management of the Day-Ahead Market, continuing to operate the market based on the principles of transparency and competitiveness. This development injected new energy and perspective into the Turkish Electricity Market, promoting values of competition.

The Day-Ahead Market introduced several innovations, including the adjustment of consumption based on price levels by the demand side, fostering active market participation, and offering a mechanism for hedging against price volatility. Additionally, it facilitated portfolio management for participants, promoting a more balanced market structure and mitigating imbalances within generation/consumption units. The non-mandatory nature of participation, daily financial settlements, and the establishment of a collateral mechanism to ensure the market's financial stability were also significant advancements.

Central to the day-ahead market is the determination of the MCP, which plays a pivotal role in the day-ahead market. In order to facilitate transactions between electricity suppliers and consumers and meet future electricity needs, EXIST oversees a time-sensitive trading platform that determines this price. The equilibrium of supply and demand determines the MCP, promoting competition and effective pricing. EXIST's supervision and regulation of this market ensure fair and competitive trading, standardizing the procedures for price calculation and market clearing price determination. This system dynamically adjusts pricing in response to supply and demand fluctuations, adhering to the principles and regulations set by EXIST.

As a result of significant regulatory and operational advancements, the liberalization process for the electricity market represents a transformation from a monopolistic to a competitive paradigm. The Day-Ahead Market, with its introduction of the MCP and innovative features, has been instrumental in this transformation. It has fostered a more dynamic, participant-engaged, and financially robust electricity market ecosystem in Türkiye, marking the beginning of a new era in electricity markets with a focus on developing a solid foundation and executing crucial steps for its advancement.

## *3.3 Development Of Natural Gas Markets In Türkiye*

This section looks at the historical growth and significant turning points in Türkiye's natural gas markets. Türkiye has transformed its natural gas industry significantly over the years to satisfy evolving energy needs and provide energy security. Türkiye has developed a robust natural gas market framework as a result of legislative reforms, the creation of regulatory organizations like the EMRA, and calculated infrastructure investments. Significant turning points have included the Natural Gas Market Law's implementation, the market's opening up, and the initiatives to diversify the sources of natural gas supply. These changes have significantly influenced the natural gas markets in Türkiye and improved the resilience and sustainability of the nation's energy supply.

### 3.3.1 The Significance of Natural Gas in the Energy Market of Türkiye

Türkiye relies heavily on natural gas to fulfill a substantial proportion of its energy requirements. Natural gas is a vital energy source across various sectors in Türkiye, encompassing electricity generation, industrial applications, residential and commercial heating, and transportation. This section analyzes the role and historical evolution of natural gas within the energy markets of Türkiye.

The significance of natural gas within Türkiye's energy markets has progressively escalated with the nation's economic expansion and population growth. Türkiye has experienced substantial growth in energy demand due to its rapid urbanization and industrialization since the 1960s. During this time frame, most energy demands were satisfied using fossil fuels, with a particular emphasis on coal and petroleum derivatives. The role of natural gas in Türkiye's energy markets has undergone a substantial transformation since the 1980s. In the 1980s, Türkiye implemented measures aimed at diversifying natural gas utilization and procurement origins. During this particular time frame, Türkiye initiated the procurement of natural gas from Russia and Iran through the execution of long-term agreements to supply said resource.

The Electricity Market Law, enacted in 2001, has been instrumental in shaping and overseeing the energy markets in Türkiye. Through the implementation of this legislation, Türkiye has effectively liberalized its energy markets, fostering an environment that promotes healthy competition. The significance of natural gas in the energy market has become increasingly evident during this time. In recent years, there has been a notable increase in the proportion of natural gas utilized for electricity generation. The proportion of natural gas utilized in electricity production in Türkiye has experienced a significant and rapid growth trajectory since the commencement of the 21st century. The proliferation of natural gas combined cycle power plants and advancements in technology have facilitated the increased utilization of natural gas in electricity generation.

Energy consumption in Türkiye exhibits a persistent upward trend, driven by the rapid pace of economic expansion and industrial development. With the growing demand for energy, the significance and role of natural gas in the energy markets have experienced a notable escalation. Energy policies aim to achieve energy supply security and decrease reliance on energy imports by diversifying the utilization of natural gas alongside domestic and renewable energy sources.

In summary, the significance of natural gas in Türkiye's energy markets is steadily increasing. A strategic objective in energy policies accompanies the growing utilization of natural gas in electricity generation and industrial sectors to promote the diversification of natural gas sources and bolster domestic energy resources. The advancement of energy markets in Türkiye and the significance of natural gas are pivotal factors in guaranteeing the nation's energy supply security and establishing sustainable energy policies.

### 3.3.2 The Discovery and Initial Utilization of Natural Gas in Türkiye

Identifying and utilizing natural gas in Türkiye represent a pivotal milestone within the nation's energy industry. The discovery of natural gas in Türkiye has brought about a substantial shift and alteration in the country's energy provision. This section will provide a comprehensive account of the historical background and procedural aspects of the exploration and initial utilization of natural gas in Türkiye.

Natural gas discovery in Türkiye occurred in the late 1950s and early 1960s. The initial identification of natural gas reserves occurred in 1961 within the Northern Marmara Region, specifically in the vicinity of Tuzla, situated in Aşağıtepe. The discovery mentioned above constituted a crucial milestone in uncovering the potential of Türkiye's subterranean resources.

Commencing natural gas production in İstanbul, specifically at the Ruzgarlıbahce well, marked a significant milestone in 1962. Subsequently, the exploration and extraction of natural gas experienced a notable acceleration. In 1967, the General Directorate of Mineral Research and Exploration (MTA) made a significant discovery of previously unknown natural gas reserves in Karapınar, Zonguldak. This discovery has demonstrated that Türkiye possesses a more extensive natural gas potential.

Nevertheless, it needed to be more economically feasible to utilize and assess the commercial viability of natural gas during this particular period. Consequently, efforts were initiated to explore the cost-effective utilization of natural gas after its initial discoveries.

The initial utilization of natural gas for economic purposes and its subsequent commercial assessment in Türkiye took place in 1987. The utilization of natural gas in the energy industry commenced with the establishment of the İzmit Natural Gas Combined Cycle Power Plant, which is under the ownership of Osmangazi Electricity Production A.Ş. Since this particular date, there has been a significant surge in the utilization of natural gas for energy generation in Türkiye, resulting in its emergence as a crucial fuel source for electricity production.

During the 1990s, there was a significant surge in Türkiye's demand for natural gas and a subsequent increase in its energy dependence. During this particular period, Türkiye established enduring supply agreements with Russia and Iran, utilizing natural gas pipelines to enhance its natural gas supply diversification. Türkiye is actively engaged in many initiatives aimed at augmenting the prominence of natural gas within the energy sector and safeguarding the security of the energy supply. One of the primary objectives of energy policies is to assess Türkiye's domestic natural gas reserves and promote the diversification of the energy supply by utilizing renewable energy sources.

In summary, the identification and initial utilization of natural gas in Türkiye have constituted a momentous milestone within the nation's energy industry. Using natural gas in energy generation has proven indispensable in facilitating economic expansion and has assumed a pivotal position in shaping Türkiye's energy strategies. The strategic pursuit of diversifying Türkiye's natural gas supply and utilizing domestic energy resources are crucial measures in guaranteeing long-term energy supply security, thereby contributing to a sustainable future.

### 3.3.3 The Socioeconomic Consequences of Natural Gas in Türkiye

The utilization of natural gas in Türkiye has a profound impact on both the country's economic landscape and its social fabric, with these effects intensifying as the use of natural gas becomes more prevalent and diversified.

Economically, natural gas plays a vital role in Türkiye's energy sector, with its use spanning electricity generation, industrial processes, residential heating, and commercial applications. This resource significantly contributes to meeting energy demands and reducing production costs, which in turn fosters economic growth. Additionally, the energy efficiency and environmental benefits of natural gas positively influence the economy.

In the business realm, natural gas usage affects various aspects of operations. The industrial sector, known for its high energy consumption, relies heavily on natural gas as a key resource due to its cost-effectiveness and lower environmental impact. Abundant and affordable natural gas supplies lead to reduced production costs for businesses, enhancing their competitiveness. This, in turn, improves energy efficiency and profitability for companies.

For consumers, natural gas is crucial in heating and energy expenses. It provides efficient heating and hot water services to residential and commercial properties, often being a more cost-effective option compared to electricity, especially in winter. However, fluctuations in natural gas prices can have a direct impact on consumer budgets. Socially, natural gas plays a significant role in creating comfortable living environments, essential for good health and well-being in residential and commercial settings. Its status as a cleaner energy source also positively impacts societal well-being, especially in terms of environmental implications and air quality.

Overall, the economic and social impacts of natural gas in Türkiye are substantial and varied, and the expected increase in its usage and diversification will likely enhance these effects even further.

### 3.3.4 The Regulation and Privatization of the Natural Gas Market

Türkiye relies on natural gas imports due to the energy sector's substantial reserve demand. Hence, the regulation and privatization of natural gas play a crucial role in ensuring energy security and fostering a competitive market landscape. The regulatory and privatization procedures of the natural gas market in Türkiye significantly impact the formulation of energy policies and the configuration of the market.

The legal framework governing the natural gas market in Türkiye was established by enacting the Electricity Market Law in 2001 and the subsequent adoption of the Natural Gas Market Law in 2002. These laws aimed to promote the liberalization of energy markets and establish a framework that fosters competition within them. The objective of regulating the natural gas market is to establish a clear boundary between various operational aspects, including energy generation, transmission, distribution, and commercial transactions. In the present scenario, the private sector was granted access to produce and import natural gas, thereby fostering a competitive milieu. EMRA was established to oversee and supervise the operations of the natural gas market.

Furthermore, the regulation of the privatization of the natural gas market in Türkiye was also implemented. The Privatization Law, enacted in 2001, also initiated the privatization procedure within the energy sector. In the context of this study, privatization auctions were utilized to privatize natural gas transmission and distribution companies, power plants, and storage facilities. Implementing regulatory measures and privatization initiatives in the natural gas sector in Türkiye has facilitated the development of a competitive and transparent energy market structure. Privatization policies have fostered an environment conducive to private sector investments, resulting in enhanced operational efficiency and the establishment of a reliable and cost-effective supply of natural gas. Simultaneously, there has been a decline in consumer energy costs due to the decrease in natural gas costs reflected in the prices consumers pay.

The regulatory and privatization measures implemented in Türkiye's natural gas market have facilitated the development of a competitive and efficient structure within the energy sector. The processes mentioned above have benefitted natural gas's economic and social implications, thereby significantly enhancing energy security. Nevertheless, adhering to meticulous protocols regarding the volatilities observed in the international energy markets is imperative. Hence, the ongoing evaluation and revision of energy policies are essential for ensuring Türkiye's energy market's long-term viability and competitive advantage.

### 3.3.5 Imports and Sources of Natural Gas Supply

Türkiye depends on imported natural gas to meet its domestic energy requirements, procuring it from a range of nations such as Russia, Iran, Azerbaijan, Algeria, and Nigeria. Pipelines are the main means of importing natural gas into Türkiye, although there has also been a substantial rise in the importation of liquefied natural gas (LNG).

Russia is the primary provider of natural gas to Türkiye, supplying a significant proportion of the country's imports. Türkiye has an enduring bilateral deal with Russia for the provision of natural gas, which is set to last until 2024. In addition to Russia, Iran is also a significant contributor to the total imports, supplying a considerable portion. The natural gas supply deal between Türkiye and Iran is anticipated to continue in effect until 2026. Azerbaijan is the third-largest supplier of gas to Türkiye, and its contribution plays a significant role in Türkiye's total gas imports. This is made possible through a bilateral agreement that will remain in effect until 2042.

Algeria and Nigeria are also noteworthy providers, with each making up a lesser but still considerable share of the overall gas imports. Türkiye has established bilateral natural gas supply agreements with Algeria and Nigeria, which are expected to remain in effect until 2030 and 2028, respectively. Türkiye's varied range of suppliers ensures a consistent and reliable supply of natural gas, which is essential for meeting its energy needs.

The issue of Türkiye's natural gas imports poses a notable challenge to the nation's energy security. The nation exhibits a significant dependence on a limited number of suppliers, thereby rendering it susceptible to potential disruptions in the supply chain. Türkiye is working to diversify its natural gas sources and decrease its dependence on Russia.

Türkiye possesses a constrained capacity for natural gas storage. The nation has three primary subterranean facilities that store natural gas: Tuz Kuyu, Salt Lake, and Değirmenköy. The aggregate storage capacity of these facilities is approximately 1.5 billion cubic meters. Türkiye actively works to enhance its liquefied natural gas (LNG) storage capacity. The nation possesses a pair of LNG terminals, namely Marmara Ereğlisi and Aliağa. The terminals possess a collective regasification capacity of approximately 10 billion cubic meters annually for LNG. Establishing natural gas storage facilities is crucial in ensuring Türkiye's energy security. This measure aids in safeguarding the nation against potential interruptions in supply and enables it to fulfill its natural gas requirements during periods of high demand.

### 3.3.6 Mechanisms for Pricing Natural Gas

In the intricate landscape of Türkiye's energy market, the pricing mechanisms for natural gas stand as a critical component, influencing not only the market dynamics but also the economic viability of the energy sector. The journey towards liberalization and privatization of the energy market has brought about profound changes, reshaping how natural gas is priced, sold, and distributed. This evolution necessitates a thorough understanding of both market-based mechanisms and the regulatory framework that oversees these processes.

Market-based pricing mechanisms play a foundational role in the natural gas sector. Among these, spot market prices are particularly noteworthy. They represent the immediate, real-time prices at which natural gas is bought and sold, offering a clear reflection of the current supply and demand situation. The spot market serves as a critical platform for the day-to-day trading of natural gas, allowing prices to adjust rapidly in response to fluctuations in supply and demand. Conversely, long-term contracts offer a counterbalance to the volatility of spot market prices. These agreements between suppliers and buyers lock in the purchase of specific quantities of natural gas over extended periods of time, often at fixed prices. Such contracts provide a measure of stability and predictability for both parties, buffering them against the short-term price swings prevalent in the spot market.

Regulatory oversight is paramount in ensuring that the natural gas market operates smoothly and equitably. In Türkiye, this responsibility falls to EMRA. EMRA's role encompasses the careful setting of regulated prices, which are designed to reflect the myriad costs associated with the production, transportation, and distribution of natural gas. Through its regulatory interventions, EMRA aims to ensure that natural gas prices are fair, reasonable, and aligned with the broader economic and operational realities of the energy market.

While EXIST is a central figure in Türkiye's electricity market, its involvement in the natural gas pricing mechanism is more peripheral. EXIST's expertise and operational focus lie primarily in the electricity sector, where it plays a crucial role in setting electricity rates, managing market operations, and ensuring the transparency and efficiency of electricity trading. Despite its limited direct involvement in natural gas pricing, EXIST's management of the electricity market indirectly influences the broader energy sector, including aspects of natural gas trading.

The determination of natural gas prices in Türkiye is a multifaceted process, influenced by a wide array of factors. These include the costs of production, transportation, and distribution, the dynamic interplay of supply and demand, global market trends, exchange rate fluctuations, and the overarching rate of economic growth. EMRA's vigilant oversight of this process is designed to adaptively respond to these factors, adjusting pricing mechanisms as necessary to maintain market balance and competitiveness.

In summary, the pricing of natural gas in Türkiye involves a complex interplay between market-based mechanisms and regulatory oversight. While spot markets and long-term contracts provide the immediate framework for trading, EMRA's regulatory guidance ensures that prices remain fair, equitable, and in line with the economic realities of the energy sector. As Türkiye continues to navigate its path towards an increasingly liberalized and privatized energy market, the roles of EXIST and EMRA will remain pivotal in shaping the market's future, ensuring that it remains competitive, efficient, and responsive to the needs of both suppliers and consumers.

### 3.3.7 Relationship Between Natural Gas and Electricity Markets

Understanding the relationship between MCP and the merit-order method is crucial to grasping the dynamics of the day-ahead electricity market in Türkiye. These concepts are directly influential in matching supply and demand within the day-ahead market and determining electricity prices.

The merit-order method ranks electricity generation units based on their variable costs, from the least to the most expensive. In the day-ahead market, electricity demand is fulfilled according to this ranking to ensure the market consistently benefits from the lowest-cost generation sources, minimizing total production costs. The variable cost of the last unit of electricity required to meet demand in this ranking determines the MCP. This price becomes the price for all electricity transactions that day.

The merit-order method ranks electricity generation units based on their variable costs, from the least to the most expensive. In the day-ahead market, electricity demand is fulfilled according to this ranking to ensure the market consistently benefits from the lowest-cost generation sources, minimizing total production costs. The MCP is determined by the variable cost of the last unit of electricity needed to meet demand in this ranking. This price becomes the price for all electricity transactions that day.

The day-ahead market facilitates electricity production and consumption planning for the next day. In this market, electricity producers and consumers submit their offers to buy or sell electricity for the following day, which are matched based on the merit order, setting the MCP. This process guarantees prices are determined transparently and competitively while prioritizing the lowest-cost generation sources to reduce production costs. The schematic diagram of the MCP and its relationship with the merit-order method is shown in the figure 2.

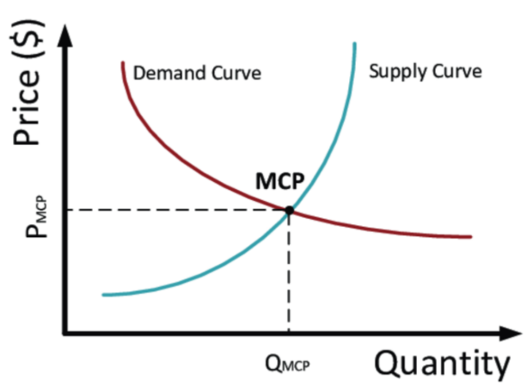


**Figure 2:** The Schematic Diagram of Market Clearing Price

[26]

In Türkiye's electricity market, the MCP is a critical determinant of electricity pricing, with natural gas prices playing a significant role. Given the substantial share of natural gas in electricity generation, fluctuations in natural gas prices directly affect electricity production costs and, consequently, the MCP.

In the merit-order method, natural gas-fired power plants are usually the most expensive generation units due to their relatively high operational costs. These plants often operate as marginal producers, frequently running almost every hour to meet electricity demand. As marginal producers, they set the electricity price since their bids are the last ones accepted to satisfy the demand.

When natural gas prices increase, the bids submitted by these plants also rise, which directly impacts the MCP. During the COVID-19 pandemic and the Russia-Ukraine conflict, global natural gas prices saw significant increases, which translated into substantially higher electricity prices in Türkiye as the natural gas-fired power plants largely determined the MCP.

# CHAPTER IV

# APPROACH FOR COMPREHENDING THE MODEL

A thorough explanation of the study's methodology is given in this section. There will be a discussion on the choice of analytical techniques, data analysis, and conclusion drafting. First, a thorough explanation of the commonly used least-square estimation Method in data analysis will be provided. This technique reduces data set variability, making it a vital tool for evaluating model fit. Subsequently, an in-depth comprehension of the correlations in the analysis will be achieved by examining the Linear and Multiple Regression approach. Finding and estimating the link between dependent and independent variables is a common application of this technique. On the other hand, error metrics will be employed to evaluate how accurate the outcomes are. These metrics offer a range of ways to assess how well the model predicts the future. The associations between the variables in the analysis will be ascertained through correlation. Lastly, a detailed explanation of the method for this situation which was especially created and used in this study will be provided. According to the needs of the analysis and the available data set, this method provides a customized methodology.

## *4.1 Least Square Estimation Method*

The Least Squares Estimation Method(LSE) is a statistical method for estimating the parameters of a mathematical model by minimizing the sum of the squares of the difference between data points. In addition, Multiple Linear Regression (MLR) has gained importance in modeling the relationship between a dependent variable and more than one independent variable. This research focuses on integrating LSE with MLR to create an accurate and efficient forecasting model. The mathematical model of the LSE is shown in equation 1.

$$LSE(\beta_0, \beta_1) = \sum_{i=1}^{n}(y_i - (\beta_0 + \beta_1 x_{1i}))^2 \quad (1)$$

In this study, there is a dataset with one dependent variable ("y") and more than one independent variable ("x"). Our goal is to create a model representing the relationship between the dependent and independent variables. The least-squares method minimizes the sum of squares of the error (the difference between the actual value and the predicted value) as the best way to choose a line (or curve) representing this relationship.

The least-square estimation is used with the Multiple Linear Regression Model to measure the model's performance and evaluate the model's fit. In order to create a model suitable for the data at hand, the parameters (coefficients) of the model are calculated with the least square estimation method, and predictions are made thanks to these parameters. Then, the model's performance is evaluated by calculating the sum of the squared errors between the actual and predicted values.

Using this method, we have chosen the model that best fits the data and measured its performance. Thus, we have statistically documented the relationship between the dependent and independent variables in the thesis topic and evaluated how well the model has predictive power.

## 4.2 Linear And Multiple Regression

Linear regression is a statistical technique employed to examine the association between two variables measured on a continuous scale. The linear regression model postulates a linear association between the dependent variable ($Y$) and a single independent variable ($X$). Its primary aim is to find the optimal parameters of the linear equation that best captures this relationship. Calculating the mean in linear regression involves using the regression equation expressed in equation 2 below:

$$Y = \beta_0 + \beta_1 X_1 \tag{2}$$

Multiple linear regression is an expanded form of regression analysis incorporating multiple predictor variables to predict the response variable. Therefore, this methodology entails predicting a numerical outcome by leveraging multiple variables. Equation 3 shows the multiple linear regression equation.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n \tag{3}$$

In the given context, $Y$ represents the dependent variable, while $x_1, x_2, \ldots, x_n$ represent the independent variables. The independent variables' coefficients are represented as $\beta 0, \beta 1, \beta 2, \ldots, \beta n$ and the error term is represented as $\epsilon$

The study utilized a multiple linear regression model. The Baseline Model, in the context of short-term forecasting, includes calendar factors, demand predictions, and the natural gas reference price. The variables used for the Baseline Model in this investigation, as stated, are provided in equation 4.

$$p_t = \beta_0 + \beta_1 H_t + \beta_2 D_t + \beta_3 W_t + \beta_4 \hat{W}_t + \beta_5 H_t D_t + f(L_t) \tag{4}$$

The location of the price estimate for a given time (hour)$t$ is not specified. The coefficients $\beta i$ for the variables Ht, Wt, and Wtt are estimated using the least squares method. Ht represents the hour of the day, Wt represents the day of the week, and Wtt is a binary indicator that distinguishes between working days (zero) and holidays (one) to determine the hour of the day. The load forecast at time $t$ plays a crucial role in the long-term price forecasting model. It is determined by the current hour's load forecast $L_t$, as represented by equation 5.

$$f(L_t) = \beta_6 L_t + \beta_7 L_t^2 + \beta_8 L_t^3 + \beta_9 H_t L_t + \beta_{10} H_t L_t^2 + \beta_{11} H_t L_t^3 \\ + \beta_{12} H_t L_t + \beta_{13} H_t L_t^2 + \beta_{14} H_t L_t^3 + \beta_{15} W_t L_t + \beta_{16} W_t L_t^2 + \beta_{17} W_t L_t^3 \quad (5)$$

The short-term forecasting model incorporates historical prices, denoted as equation 6.

$$p_t = \beta_0 + \beta_1 H_t + \beta_2 W_t + \beta_3 \hat{W}_t + \beta_4 H_t W + f(\tilde{D}_t) + h(\hat{p}) \quad (6)$$

$$h(\hat{p}) = \beta_{19} \frac{1}{d} \sum_{\tau=1}^{d} \hat{p}_{t-24\tau} + \beta_{20} \frac{1}{w} \sum_{\tau=1}^{w} \hat{p}_{t-168\tau} \quad (7)$$

The variables $d$ and $w$ represent the indicators for the day and week, respectively. In the sixth instance, the utilization of $d = 1$ designates the initial term as the previous day's price at the same hour. Similarly, the utilization of $w = 2$ designates the second term as the average price of the same-day, same-hour from the previous two weeks. The objective of this study is to evaluate the impact of past natural gas prices as a new indication for forecasting electricity prices in the near future. In order to assess the influence of DRP on MCP, a Baseline Model was developed. Therefore, the Baseline Model includes two more predictors, as shown in equation 8.

$$p_t = \beta_0 + \beta_1 H_t + \beta_2 W_t + \beta_3 \hat{W}_t \beta_4 H_t W + f(\tilde{D}_t) + h(\hat{p}) + g(\lambda) \quad (8)$$

$$g(\lambda) = \beta_{21} \frac{1}{24d} \sum_{\tau=t-24(d+1)}^{t-24} \lambda_\tau + \beta_{22} \frac{1}{168w} \sum_{\tau=t-24-168w}^{t-24} \lambda_\tau \quad (9)$$

In this context, the variables $d$ and $w$ represent the day and week indicators, respectively. Additionally, the symbol $\lambda$ denotes the hourly natural gas price. In equation (8), the utilization of $d = 1$ designates the preceding day's natural gas price as the first term. In contrast, using $w = 2$ designates the average daily prices from the past two weeks as the second term. During the period spanning from April 2020 to December 2022, the received data underwent an exploratory data analysis process, resulting in the acquisition of comprehensive information about the dataset. Subsequently, during the data preparation stage, a thorough analysis was carried out to detect any missing values and outliers. The researchers utilized a multivariate linear regression model to evaluate the influence of these variables on the hourly pricing.

## *4.3 Error Metrics*

This section outlines the critical error metrics for evaluating forecast models' performance in the electricity market. Error metrics play a crucial role in assessing the alignment between predictions and actual values, providing valuable insights into the accuracy and effectiveness of the prediction model. The error metrics are defined in equations 10 - 15.

Mean Absolute Percentage Error (MAPE): MAPE is a statistical metric utilized to assess the precision of a forecasting model. It quantifies the average absolute percentage discrepancies between observed and projected values. MAPE measures the degree of imprecision in the forecasts as a percentage. A smaller MAPE number signifies more accuracy, as it implies that the estimated values are closely aligned with the real values. The equation of MAPE is shown in equation 10.

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{A_i - F_i}{A_i}\right| \tag{10}$$

Mean Absolute Error (MAE): MAE is a commonly employed metric for assessing the efficacy of regression models. The metric quantifies the mean absolute deviation between the predicted and actual values within a given dataset. MAE is a reliable metric for evaluating the accuracy of a model, as it focuses solely on the magnitude of errors without considering their direction. The equation of MAE is shown in equation

11.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|A_i - F_i| \tag{11}$$

Mean Squared Error (MSE): MSE is a statistical measure that quantifies the average of the squared differences between expected and actual values. The process entails computing the mean of the squared deviations between the observed and expected values. The Mean Squared Error is frequently used to evaluate the performance of prediction models. A lower MSE indicates a stronger agreement between the predicted values and the actual values. The equation of MSE is shown in equation 12.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(A_i - F_i)^2 \tag{12}$$

Mean Percentage Error (MPE): MPE is a frequently employed metric for assessing the precision of a model's predictions. The metric quantifies the mean percentage deviation between the projected and observed values within a given dataset. MPE offers a comparative assessment of the model's efficacy, enabling us to comprehend the extent of the discrepancies in the actual values. The equation of MPE is shown in equation 13.

$$\text{MPE} = \frac{100\%}{n}\sum_{i=1}^{n}\frac{A_i - F_i}{A_i} \tag{13}$$

R2 Score (R-Squared): R2 Score is a statistical measure that represents the described variance rate of a regression model. The R2 score shows how well the model explains the dependent variable by independent variables. It has a value ranging from 0 to 1, and the closer to 1, the better the model fits. The equation of R2 Square is shown in equation 14.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{14}$$

Root Mean Squared Error (RMSE): RMSE is the square of the average of the squares of the differences between the estimated values and the real values of a regression model. The RMSE shows how many errors there are in the model's estimates, and the smaller the value, the higher the accuracy of the model. The formulation of RMSE is shown in equation 15.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{15}$$

The interpretation of error metrics involves analyzing and understanding various measures used to quantify a model or system's accuracy and performance. These metrics offer valuable insights into the extent of errors or discrepancies between predicted and actual values, facilitating the evaluation process.

In this section, the evaluation of the prediction models utilized in this thesis will be elucidated by employing the error above metrics. Each error metric may possess distinct advantages and limitations. A more comprehensive evaluation of the prediction model's accuracy can be achieved by employing multiple error metrics simultaneously. This approach thoroughly assesses the forecasting model's performance in the electricity market context. In the study, the results will be shared over the MAPE values.

## 4.4 Correlation

Correlation analysis is a statistical method employed to investigate the association between two or more variables. The process involves computing the correlation coefficient, which quantifies the intensity and direction of the linear association between the variables. One can ascertain the relationship between variables by employing suitable correlation approaches, such as the pearson correlation coefficient or spearman rank correlation. The investigation yielded pearson correlation coefficients.

## 4.5 Method for Baseline Model

This study began with a thorough review of existing methodologies, literature, and data analysis requirements. Building on this foundational knowledge, a step-by-step approach was developed and implemented to analyze the collected datasets. The process began with entering the relevant datasets into the system and manually merging them to ensure data consistency. The data then underwent comprehensive statistical analysis to identify patterns and extract meaningful insights. Afterward, the data was carefully cleaned to eliminate anomalies and inconsistencies, ensuring reliable results.

Once the data was prepared, a matrix was constructed using relevant attributes and features identified during the statistical analysis. The least squares estimation method was then applied to this matrix to uncover critical relationships between variables and generate predictive models. Correlation analysis and insights from a previous study further refined the feature selection process. With the data ready for modeling, a multiple linear regression approach was selected based on literature reviews and previous research. Both dependent and independent variables were identified to establish the baseline model. Two enhanced models were also developed by incorporating additional variables and advanced techniques to improve performance. Each model was assessed individually using various error metrics to evaluate predictive accuracy.

The baseline model's performance was measured with a single mean absolute percentage error value, while the two enhanced models were evaluated using 90 MAPE values each. This ensured a robust performance evaluation. Price estimation was then conducted for the cycle that demonstrated the best model performance, providing valuable insights into market trends and energy pricing. Comparative graphs were also generated to visualize the difference between the model-based price estimates and actual observed values, offering a clear representation of predictive accuracy and model efficiency.

This rigorous methodology allowed for accurate market trend analysis and reliable energy price predictions based on historical data. The methodology used in this study is shown in figure 3.

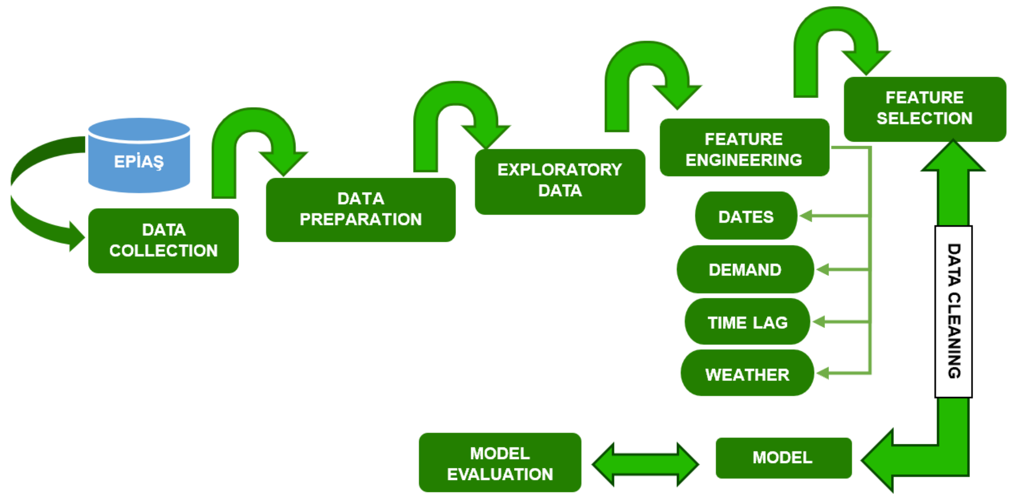


**Figure 3:** Flow Chart for Baseline Model

# CHAPTER V

# CASE STUDY AND RESULTS

A case study on how to anticipate the price of power in the near future in Türkiye's day-ahead market is presented in this section. For those involved in the energy business, predicting the price of power is a strategically significant issue, and precise projections are essential to helping market participants make decisions. This case study thoroughly examines the process of creating a model to forecast power prices in the day-ahead market, as well as the outcomes.

In Türkiye's day-ahead market, it has become necessary to estimate electricity prices in the near future in order to lower market uncertainty and empower energy suppliers to make more informed decisions. This thesis provides a detailed description of a model designed to comprehend the dynamics of Türkiye's energy market and forecast future price changes. This case study outlines the procedures used to forecast energy prices in the day-ahead market, step-by-step, from data collection to data analysis, model building, and model evaluation. This will establish a foundation for delving deeper into the intricate workings of Türkiye's energy market and improving the accuracy of pricing forecasts going forward.

## 5.1 Data Collection and Preprocessing

The dataset sources, as well as information about them, are critical to ensuring the study's reliability and accuracy. This study's data were taken from the EXIST website transparency platform. The data are reliable, accurate, and usable. First, the downloaded files were not used directly while collecting the data, and some numerical adjustments were made. Dollar-based prices were extracted, while TL and Euro-based prices were omitted. The temporal scope of the dataset spans from January 2020 to

December 2022. Data for natural gas prices and electricity prices were downloaded separately, and the.csv files were manually processed and merged. Figure 4 illustrates the historical coverage of the data.

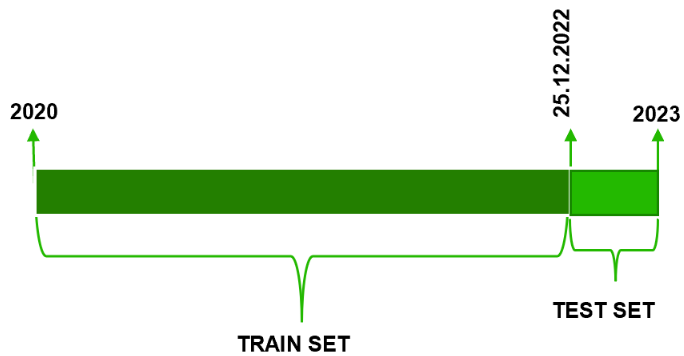


**Figure 4:** The Historical Coverage of The Data

The collection and compilation of this dataset play a crucial role in the research methodology for this study. Data were collected using a quantitative approach, specifically numerical data for analysis. The variables and properties in the dataset include their presentation and definitions.

This study has chosen MCP as the dependent variable. Market Clearing Price: The price is calculated by taking the average of electricity prices for the relevant month on a 24-hour daily basis. The day-ahead market determines the balance between supply and demand. EXIST's transparency website publishes this MCP data. Official hourly electricity prices and load plan data, published on the transparency platform between January 2020 and December 2022, were utilized, along with natural gas reference prices from the same period. MCP data was extracted from the transparency platform's Markets section, specifically the day-ahead market clearing price. Table 1 displays the captured data's titles.

**Table 1:** Market Clearing Price Data Headers

| $Date$ | $Hour$ | $MCP(TL/MWh)$ | $MCP(USD/MWh)$ | $MCP(EUR/MWh)$ |
|---|---|---|---|---|

In this study, the "Daily Reference Price (DRP)," which represents the spot natural gas price in the Turkish natural gas market, will serve as an independent parameter. The price is computed on a daily basis for a volume of one thousand cubic meters of gas in the spot market for natural gas in Türkiye. These DRP data are also published on EPİAŞ's transparency website. Load plan information data has been downloaded from electricity consumption, forecast, and load forecast plan data. The data is downloaded as a date, time, and load forecast plan (MWh). Manual conversions were made in the load forecast plan data (/10000) (Demand(MWh)(L)). For the price of natural gas, manual data sets were downloaded under the headings of Natural Gas Market, Spot Markets, and Price. From January 2020 to the end of December 2022, data was extracted from the existing transparency platform, like DRP. The captured data titles are as follows: the.csv files have been extracted. The DRP (USD/MMBtu) price is used. The titles of the captured data are listed in table 2.

**Table 2:** Daily Reference Price Data Headers

| Gas Day | DRP (TL/1000Sm3) | DRP (USD/1000Sm3) | DRP (EUR/MWh) | DRP (USD/MMBtu) |
|---|---|---|---|---|

## *5.2 Exploratory Data Analysis*

Exploratory Data Analysis (EDA) is a set of techniques and methods used to understand the main characteristics of a data set. This analysis is carried out to identify the basic distribution of data, its trends, relationships between variables, and possible anomalies. This study examined the distribution of critical variables like MCP (market clearing price), DRP (daily reference price), and demand, applied log conversions and standardizations, and analyzed the data's status prior to modeling. The EDA process identifies the structure of the data and potential problems, contributing to more sound and accurate results at the modeling stage.

### 5.2.1 Distribution of Data

The model provides predictions for the last two weeks of December 2022. MCP, DRP, and demand data taken between April 2020 and December 2022 were subjected to exploratory data analysis to obtain comprehensive information about the data. First, comprehensive data on variables received from the EXIST transparency platform will be examined with the.describe() method, examining the distribution of data points to understand the ranges, central tendencies (mean, median), and distributions (variance, standard deviation) of the variables. Calculating and presenting statistical measures, including mean, standard deviation, minimum, maximum, median, and quartiles, is crucial for any variable of interest. The statistics provided make it easy to understand the variables' comprehensive distribution and inequalities. Tables 3, 4, and 5, respectively, provide a brief report of the MCP, DRP, and demand data used in this study.

**Table 3:** A Summary Report of the MCP

| | $Count$ | $Mean$ | $STD$ | $Min.$ | 25% | 50% | 75% | $Max.$ |
|---|---|---|---|---|---|---|---|---|
| $MCP$ | 26.304 | 81 | 60 | 0 | 41 | 54 | 112 | 264 |

**Table 4:** A Summary Report of the DRP

| | *Count* | *Mean* | *STD* | *Min.* | 25% | 50% | 75% | *Max.* |
|---|---|---|---|---|---|---|---|---|
| *DRP* | 26.304 | 465 | 365 | 163 | 199 | 248 | 738 | 1445 |

**Table 5:** A Summary Report of the Demand

| | *Count* | *Mean* | *STD* | *Min.* | 25% | 50% | 75% | *Max.* |
|---|---|---|---|---|---|---|---|---|
| *Demand* | 26.304 | $3,56$ | $0,58$ | $1,70$ | $3,10$ | $3,60$ | $4,00$ | $5,50$ |

### 5.2.2 Visual Representations

Graphs such as histograms, box plots, scatter plots, and bar charts are used to visualize the relationships within the data and the distributions of variables. First of all, the structure of the data set and its variable distributions were looked at to provide an overview of the MCP. Figure 5 shows one-week MCP data for the period from April 17 to April 24, 2022.

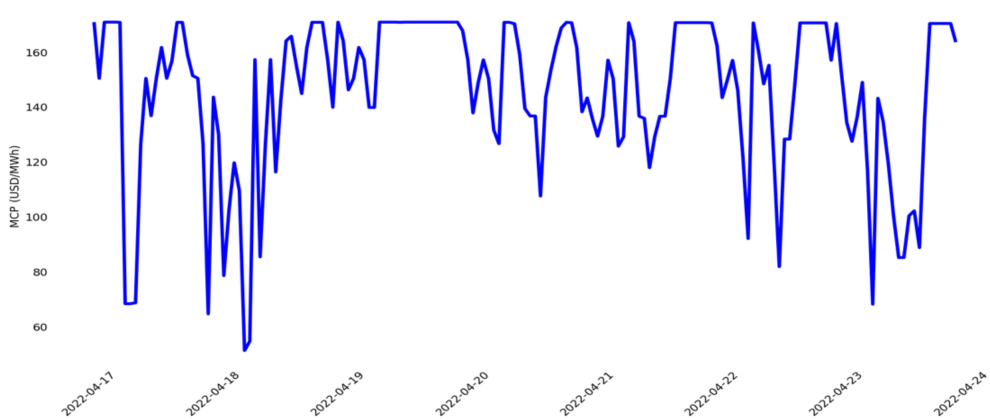


**Figure 5:** Data Period for MCP of April 17th, 2022 to April 24th, 2022

Understanding the structure of data is essential when working in data science. It is beneficial to analyze the outliers in a particular dataset to determine the concentration of the majority of information. One of the methods used to analyze the data set is to draw a box plot. In descriptive statistics, a box plot is a method for demonstrating graphically the locality, spread, and skewness groups of numerical data through their quartiles. The boxplot examines the distribution of data within each sample. As the box's length increases, the data becomes more dispersed. As the size of the box decreases, the information becomes more concentrated. An outlier is a data point that falls outside the range represented by the whiskers of a box plot. Outliers are easily identifiable when looking at the box-whisker plot. To gain a more comprehensive understanding of the data, we divided the box plots based on years and then presented each one separately. Box-plot graphs were employed to visually depict the statistical analysis of MCP. Box-plot charts in figures 6, 7, and 8 display the hourly market clearing prices for the years 2020, 2021, and 2022, respectively.

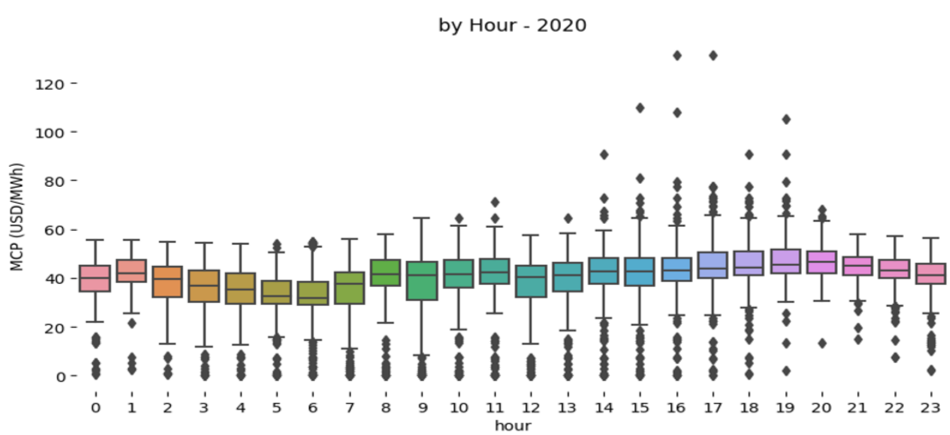


**Figure 6:** Hourly Distribution of MCP Prices by 2020

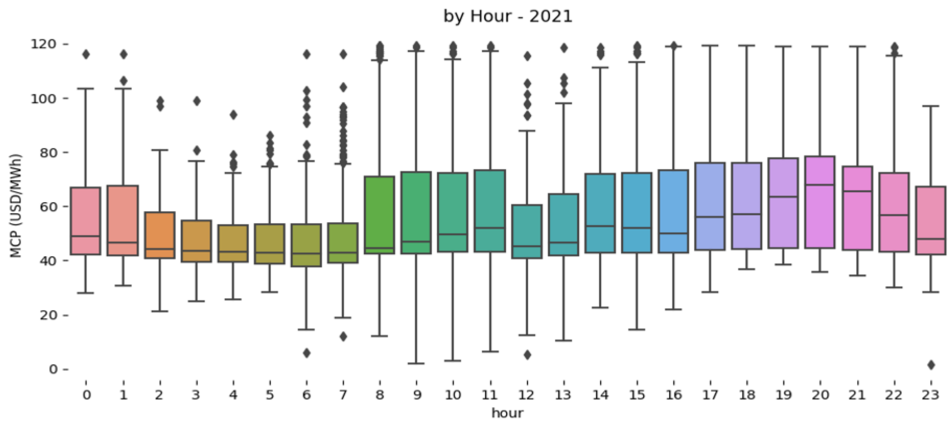


**Figure 7:** Hourly Distribution of MCP Prices by 2021

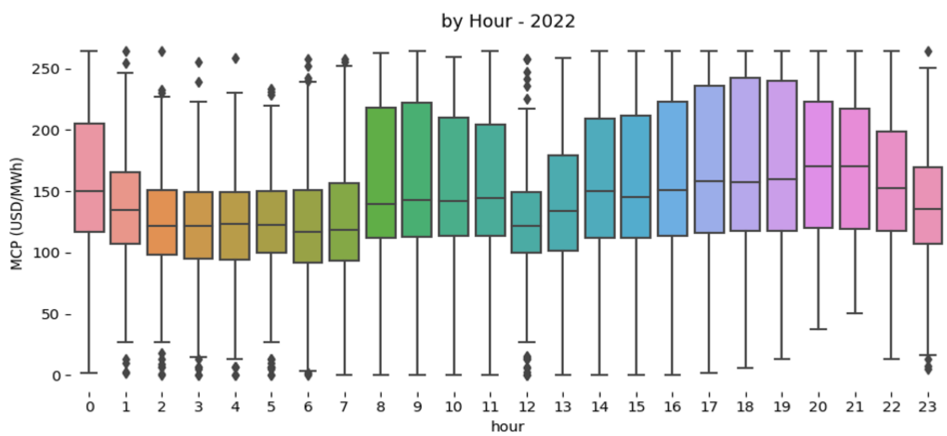


**Figure 8:** Hourly Distribution of MCP Prices by 2022

The provided statistics, presented as histogram graphs displaying monthly market clearing price data distributions for the years 2020, 2021, and 2022, simplify the understanding of the comprehensive distribution and inequalities of the variables. Figures 9, 10, and 11 present box-plot charts displaying monthly market clearing prices, respectively. This box-plot graph allows you to easily see the distribution of

MCP prices by month in the data, their density, and outlier observations.

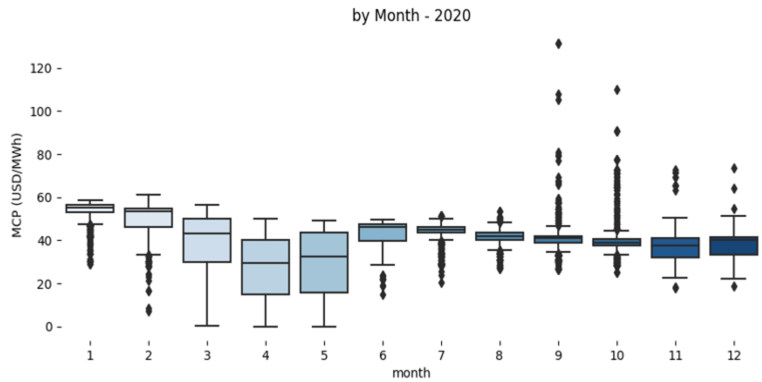


**Figure 9:** Monthly Distribution of MCP Prices by 2020

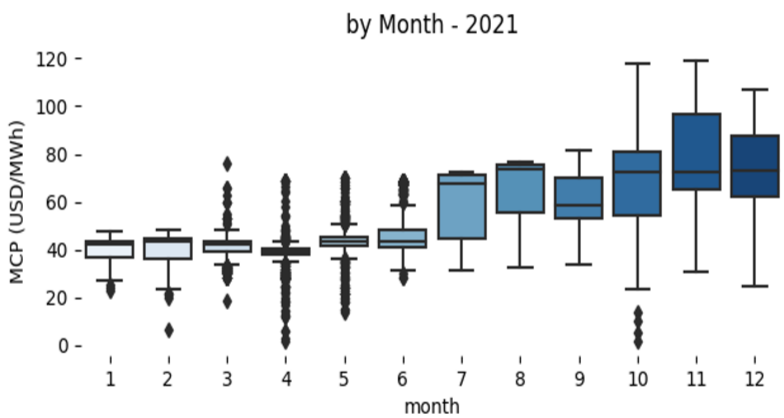


**Figure 10:** Monthly Distribution of MCP Prices by 2021

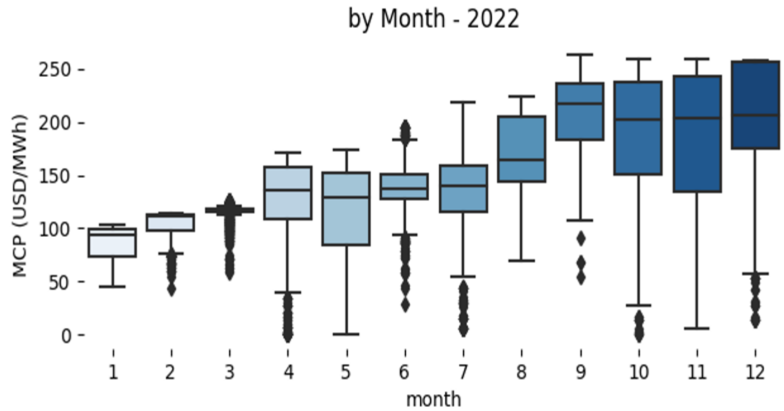


**Figure 11:** Monthly Distribution of MCP Prices by 2022

Histogram graphs were created to examine the distribution of each variable. Histogram plots provide information about the distribution of variables and help understand whether log transformation or standardization is required. Figure 12 displays the histogram graph to demonstrate the normal distribution of MCP across the entire dataset. This chart shows the distribution of MCP prices. MCP prices have a highly right-skewed distribution. The concentration of the highest frequency occurs at lower price values, while the frequency decreases with increasing price.

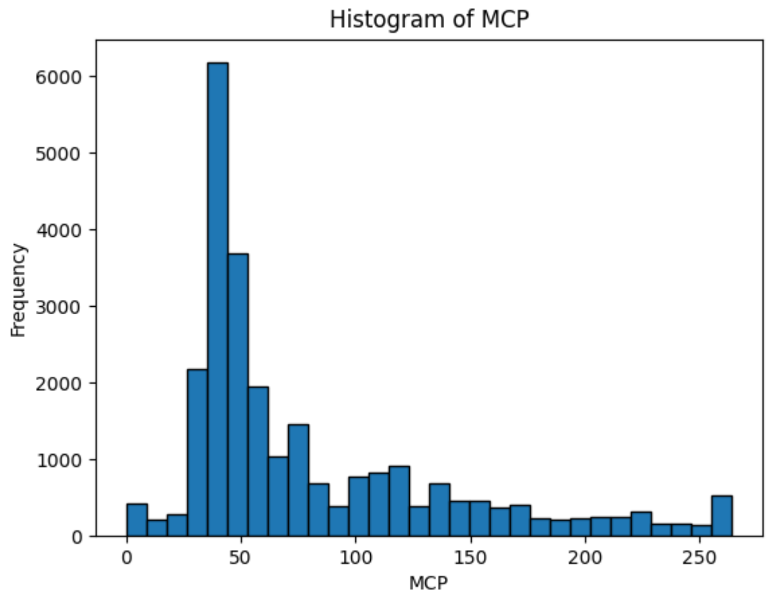


**Figure 12:** Histogram of MCP

The following information pertains to a different variable: The structure of the data set and its variable distributions were looked at to provide an overview of the DRP. Figure 13 displays one-month DRP data for the period April 1 to April 30, 2022.

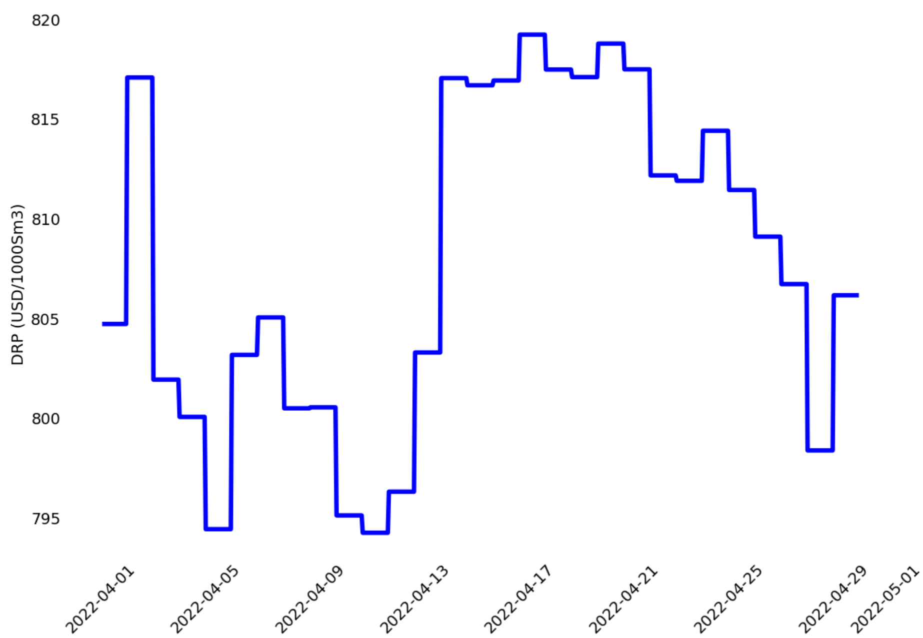


**Figure 13:** Data Period for DRP of April 1st, 2022 to April 30th, 2022

Statistical measurements including mean, standard deviation, minimum, maximum, median, and quartiles were excluded from the DRP price and shared statistically in table 4 in the previous section. Figure 14 depicts the histogram chart for the DRP variable. This histogram shows that the DRP variable has a very right-skewed distribution. It has been found that most of the DRP prices are concentrated at low values. The DRP variable has a highly right-skewed distribution, which can negatively impact the performance of the model.

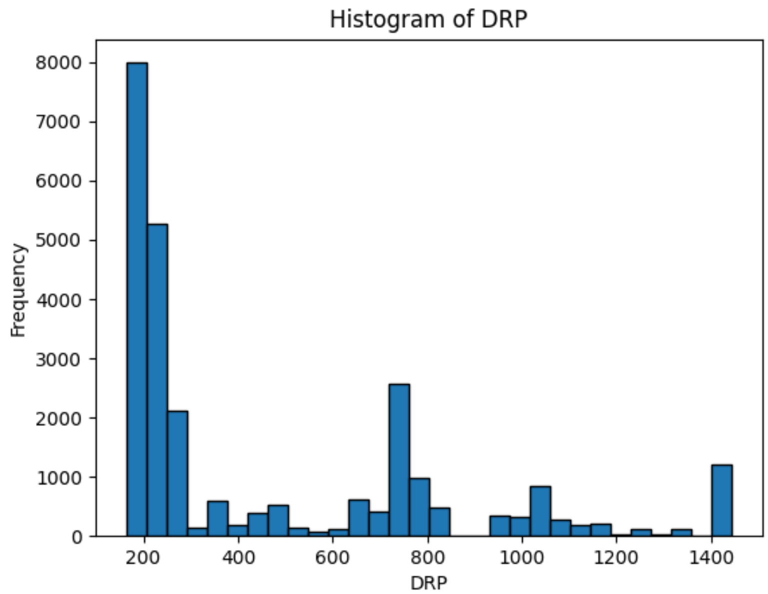


**Figure 14:** Histogram of DRP

Statistical data were examined for another variable, demand. Figure 15 displays one-month demand data for the period April 1 to April 30, 2022. Figure 16 shares the drawn histogram graph for visual analysis. This histogram chart shows the demand variable's distribution. The distribution's general shape is close to normal and symmetrical. The distribution of the demand variable is close to normal, suggesting that it may not require transformation during the modeling process.

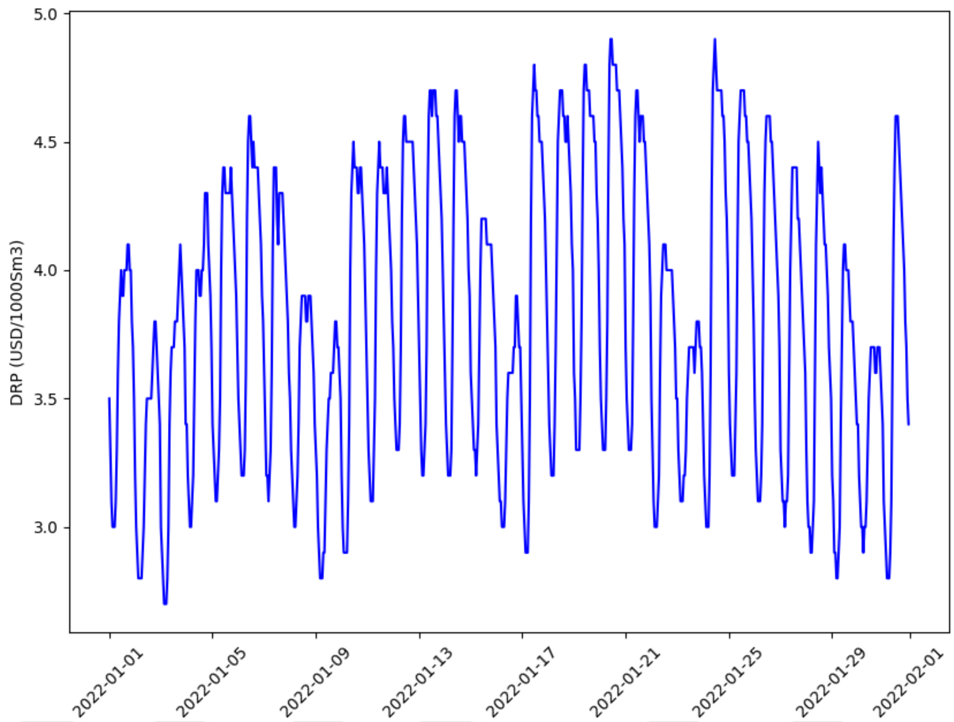


**Figure 15:** Data Period for Demand of April 1st, 2022 to April 30th, 2022

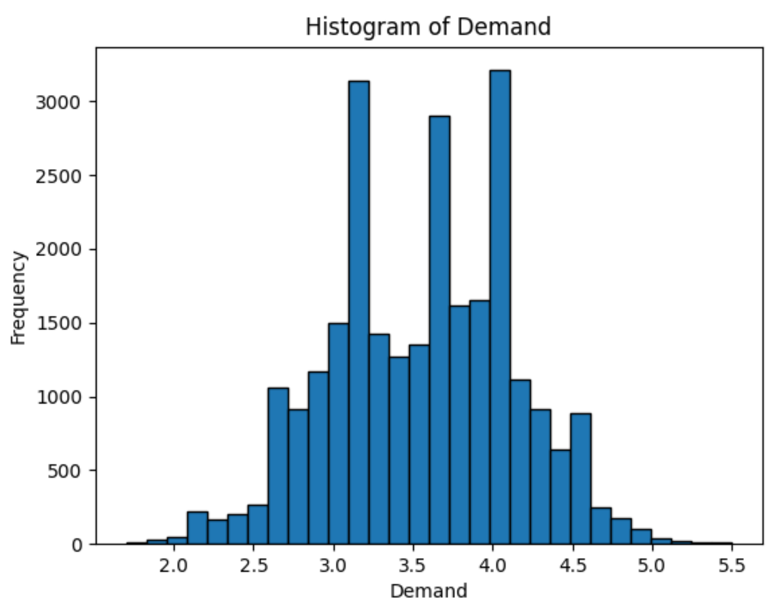


**Figure 16:** Histogram of Demand

Examining variables separately usually gives the best results.However, due to the size of the data set and the manual nature of the analysis, the MCP, DRP, and demand data were thoroughly examined. Upon the examination of the histogram graphics, the following actions were decided to be implemented. Looking at the histogram graphics, the decision was made to apply a log transformation. If a variable has a right-skewed distribution, a log transformation can help correct for this skewness. Since MCP and DRP variables are skewed to the right, a log transformation was applied. Log transformation will reduce this skewness, making the distribution more symmetrical and normal. If a variable had a left-skewed distribution, a different transformation (for example, a square root transformation) might be more appropriate. This was not implemented. When a variable is already close to a normal distribution, there is no need to log-transform. Since the demand variable shows a normal distribution, the log transformation was not applied. Next, log-transformed data should be standardized. This will improve the performance of the model. Since the demand variable is already close to normal distribution, there is no need for log transformation. However, standardizing this variable is also necessary. The histogram graphics made it clear that the variables' scales differed, leading to the decision to standardize. Figures 17 and 18, respectively, display the graphs of the MCP and DRP variables after log transformation.

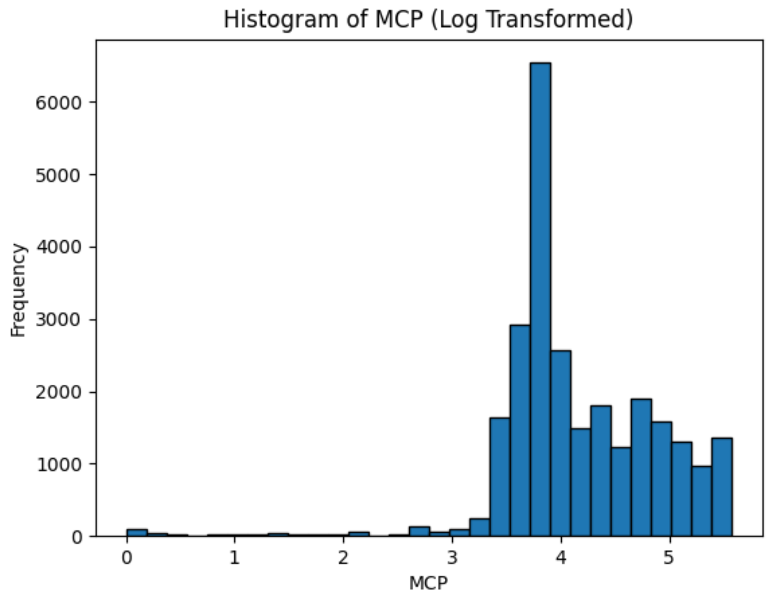


**Figure 17:** Histogram of MCP(Log Transformed)

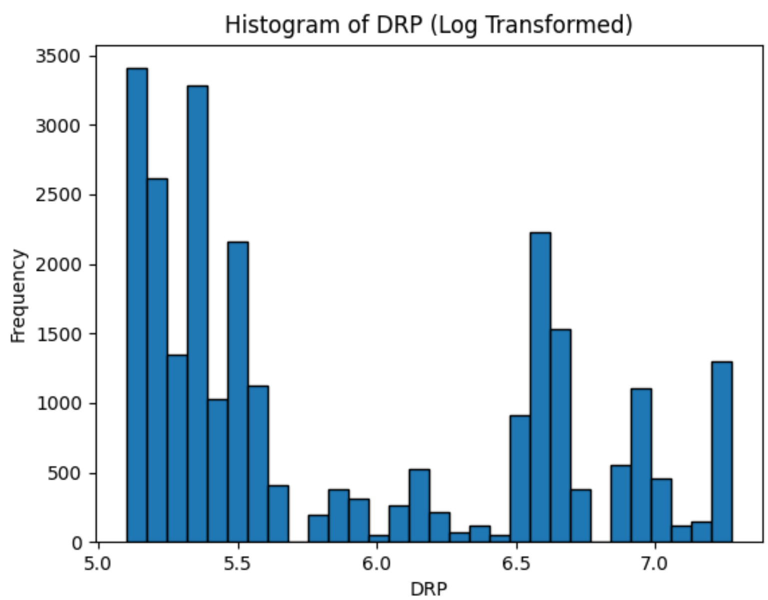


**Figure 18:** Histogram of DRP(Log Transformed)

Figure 17 shows the MCP variable distribution after applying the log transformation. The log transformation rectified the right-skewed distribution, resulting in a more symmetrical distribution.  nsformation made the distribution of the MCP variable closer to the normal distribution. Figure 18 shows the distribution of the DRP variable after applying the log transformation. The log transformation rectified the right-skewed distribution, resulting in a more symmetrical distribution. The log transformation brought the DRP variable's distribution closer to normal.

It is observed that after a log transformation is applied to both variables, their distributions become more symmetrical and closer to the normal distribution. Log transformations for these two variables can help the model perform better. This transformation was implemented specifically because it can improve the performance of linear models such as linear regression.

Standardizing the variables after log conversion allows the model to learn more steadily and quickly. If all variables are on a similar scale and the model's performance is adequate, standardization may not be necessary. Some variables were standardized, others were not. Whether or not they were constantly changing was decided. MCPs (USD/MWh) are directly measurable values, just like demand (MWh). Also, the numbers and measurements that come from collisions or derivatives, like L2 (the square of L) and H*L (the collision of Hour and Demand), are always changing because they are derived from collisions or derivatives. Standardization Applicable variables are: MCP, Demand, L2, L3, H*L, H*L2, H*L3, D*L, D*L2, D*L3, W*L, W*L2, W*L3. These are the variables that are not subject to standardization. The variables include Hour (H), Day (D), Week (W), Holiday (Whead), Hour, and Day. Data normalization, or standardization, is the process of improving a model's performance by bringing data on different scales to a similar scale. The standardization (Z-Score Normalization) method has been chosen for the data normalization process. This method scales the data to an average of 0 and a standard deviation of 1. As a result, standardization applies to all variables, except for the date variables. Figure 19, figure 20, and figure 21 respectively, display the scaled MCP, DRP and Demand graphs.

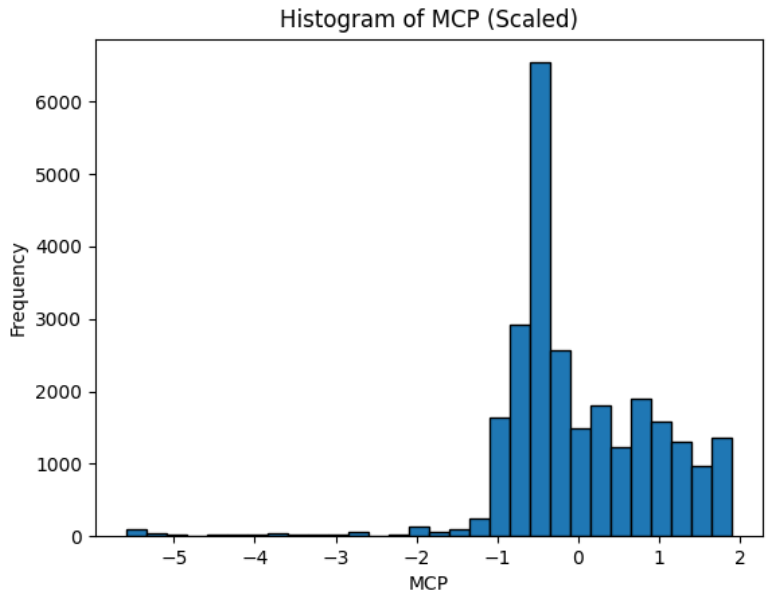


**Figure 19:** Histogram of MCP(Standard Scaled)

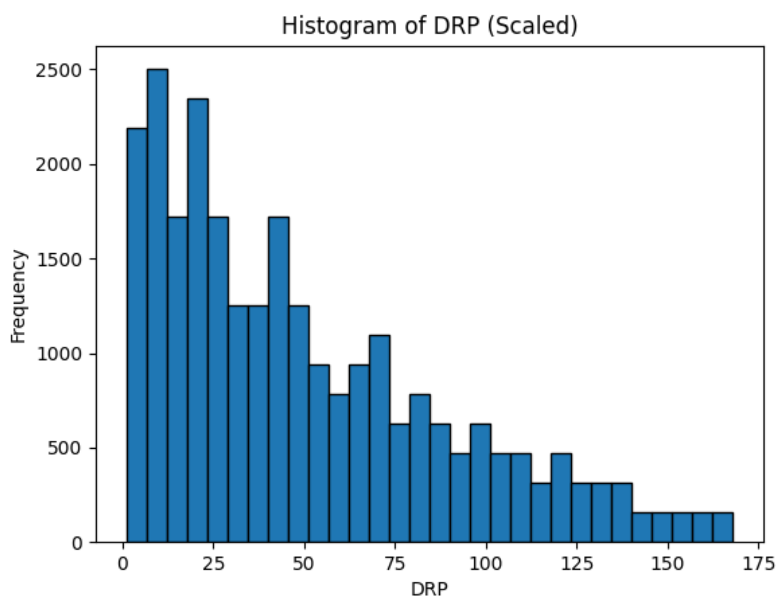


**Figure 20:** Histogram of DRP(Standard Scaled)

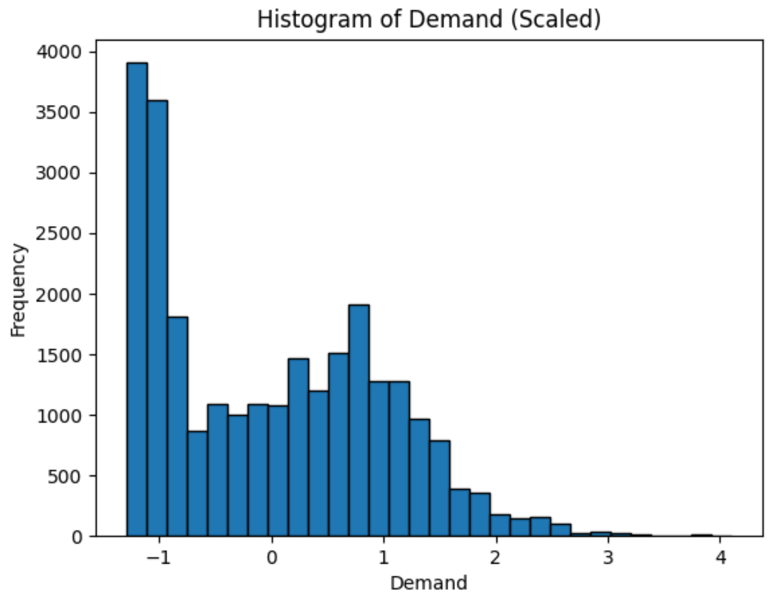


**Figure 21:** Histogram of Demand(Standard Scaled)

Additionally, it was concluded that the data set is heterogeneous. In other words, we determined appropriate transformation techniques by examining each variable separately, given their different distribution and scale properties. These steps enhance the model's performance by making the data more fitting. After preparing your data with log transformation and standardization, the model training and prediction process will begin. Both variables' distributions become more symmetrical and closer to a normal distribution after applying the log transformation. Log transformations for these two variables can help the model perform better. After the modeling process, this step was not done because log transformation negatively affected the good performance of the model.

### 5.2.3 Outlier Analysis

A more detailed outlier analysis can be performed for the extreme values (outlier values) seen in the histogram. This can be useful to improve the performance of the model. Outlier analysis is an important step to improve model performance and identify extreme outliers in the dataset. Outliers are data points that are extreme relative to the overall data and can negatively impact the model's prediction accuracy. Various methods can be used for outlier analysis. In this study, the IQR (interquartile range) method was chosen. IQR identifies outliers based on the data's interquartile range. IQR is the difference between Q1 (1st quartile) and Q3 (3rd quartile).

Outlier analysis can be performed both at the beginning and after the modeling process. Which stage to take depends on the structure of the data set, modeling objectives, and preferred approach. The advantages and disadvantages of both approaches are compared. In the first place, performing outlier analysis before the model has the following advantages: Identifying and removing outliers at the beginning can make the data cleaner and more reliable; anomalies can prevent overfitting or mislearning during the model's training; and cleaning the data can increase the model's performance rapidly. The disadvantages are as follows: Initially, removing outliers may cause you to lose important information in your data set. If outliers represent a significant trend or structure in your data set, removing them may lead to inaccurate models.

Outlier analysis is performed. The model offers the following advantages: Identifying outliers by analyzing the model's residuals allows you to make more detailed and informed decisions. The model's residuals determine outliers, allowing for a more accurate evaluation of its predictive power. Outliers can be specifically removed or fixed to improve the model's performance.The disadvantages are as follows: Performing outlier analysis after training the model and making predictions requires an extra step and time. Outliers can have a negative impact on the initial model's training, causing it to misperform.

Both approaches can be useful in different situations. However, the most common practice is to perform outlier analysis on residuals after training the model and making predictions. It was decided to share the outlier analysis after training the model, as this allows a more accurate evaluation of the model's prediction performance and a better understanding of the impact of outliers.

### 5.2.4 Handling Missing Values

This section describes identifying missing data points by assigning or removing them, depending on their impact and the amount of missing data, and deciding how to handle them. While doing this project, error metrics will be used after the model is trained. We assigned very small values, like "1e-10," to the missing data in the error metrics discussed in the previous section. This was because the missing data caused the MAPE values to increase, resulting in an infinite and meaningless formula. Other small data points and outliers are not included. This study includes both small and large data points. They also wanted to detect whether there was a trend or a special situation. The feature engineering section commenced after the EDA processes concluded.

## 5.3 Feature Engineering

Feature engineering, a preprocessing step in supervised machine learning and statistical modeling, transforms raw data into a more effective set of inputs. Each input comprises several attributes, known as features. This study produced new features from data such as MCP, DRP, and date.

Firstly, the data set used is to find the effect of natural gas prices on electricity prices. The data set is simple and includes the following parameters: MCP, DRP, time, and date information. The MCP data includes hourly MCP (USD/MWh) data spanning from January 2020 to December 2022. Daily DRP (USD/Sm3) data for DRP from January 2020 to December 2022. In order to combine the two parameters in an Excel file, the price of the DRP was written 24 hours a day by repeating it because the DRP is traded at the same price 24 hours a day. Primary data have MCP, DRP, demand, date, and time parameters. New features were produced from the date and time data. Initially, we extracted the following data from the date-time information to incorporate it into the model: In the holiday part, Saturday and Sunday are added as 1 and other days as 0. Table 6 displays the date-time headers.

**Table 6:** Date-Time Headers

| | | | | |
|---|---|---|---|---|
| $Hour(H)$ | $Day(D)$ | $Week(W)$ | $Holiday(Whead)$ | $HourandDay$ |

Date and time parameters led to the development of new features. Prices starting one day before and up to 15 days ago were produced from the MCP price used. These features were added from MCP-Lag1 to MCP-Lag15. The previous value of each day is added separately, starting from the day before and up to 15 days before the MCP. The linear relationship between the MCP value and the n-day ago value was examined. From these data, the first week's MCP data was averaged, the 1st week average was taken, the second week's average was taken, and the 2nd week average was also added as a new feature. The linearity relationship between the mean of the data from the previous week and two previous weeks to the MCP value and the MCP was also examined, and the independent variables W1-Average and W2-Average were formed. Figure 22 shows the features used in the baseline model.

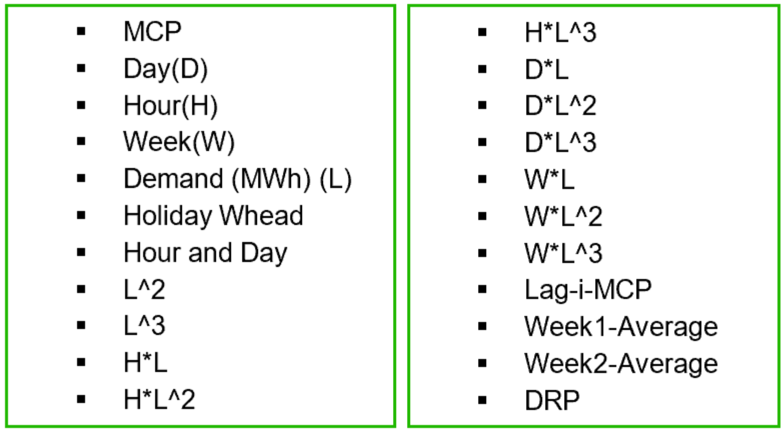


**Figure 22:** Features for Baseline Model

The baseline model also included the DRP's own value as a parameter. Two different enhanced models were developed by removing the DRP value from the baseline model and adding the previous values of DRP to the model. Then, in the enhanced model developed from the baseline model, natural gas data from DRP_Lag1 to DRP_Lag90, 3 months ago, were added one by one, according to the DRP data. The developed model adds and calculates the characteristics of natural gas prices from the first cycle separately, as shown in figure 23. Figure 24 shows the addition of the second loop-by-loop model without its removal from the model.

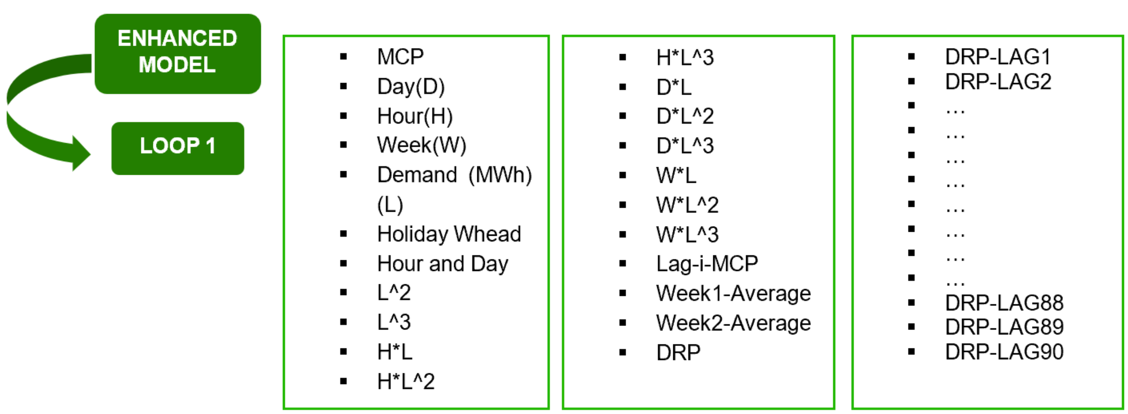


**Figure 23:** Features of the Enhanced Model in its First Loop

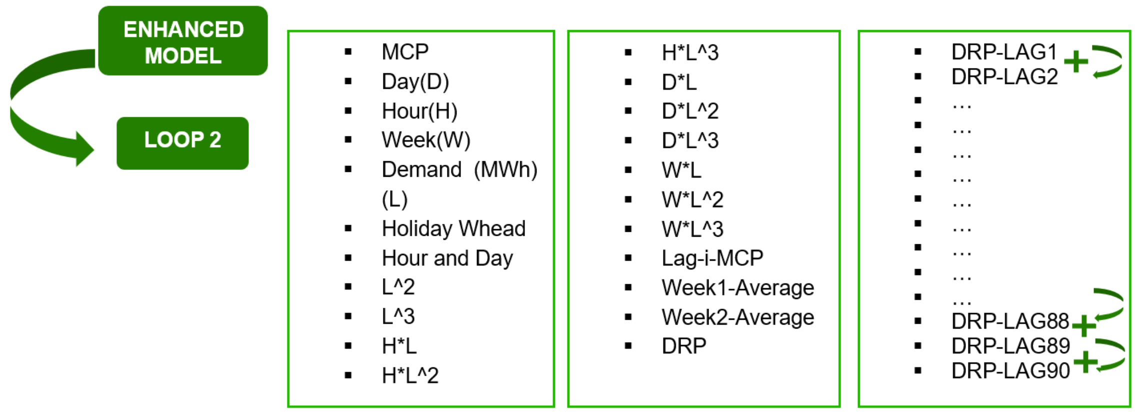


**Figure 24:** Features of the Enhanced Model in its Second Loop

## *5.4 Correlation Between MCP And DRP*

Correlation analysis is performed to check the relationships between variables and understand how they may affect each other. This includes calculating correlation coefficients and creating correlation matrices. This section explores the results of the correlation analysis carried out between MCP and the DRP. This analysis seeks to quantify the strength and direction of the relationship between MCP and DRP, thus enhancing our understanding of how these variables interact within broader market dynamics.

Employing correlation analysis between MCP and DRP enables us to assess the operational and economic efficiency of the electricity market. Understanding how changes in demand and supply influence market prices is crucial, as these dynamics significantly shape market behavior. The outcomes of this correlation analysis provide valuable insights for market participants and policymakers involved in decision-making processes, offering a clear perspective on market efficiency. Table 7 displays the year-based representation of the correlation values between them.

**Table 7:** The Correlation Between MCP and DRP

| The Correlation | | |
|---|---|---|
| **2020** | **2021** | **2022** |
| 0.23 | 0.70 | 0.59 |

The results of this correlation analysis are visually represented in a heat map, displayed in figure 25. This visual aid helps to clearly communicate the relationship between MCP and DRP, highlighting areas of strong or weak correlation and thus aiding in a more intuitive understanding of market dynamics.

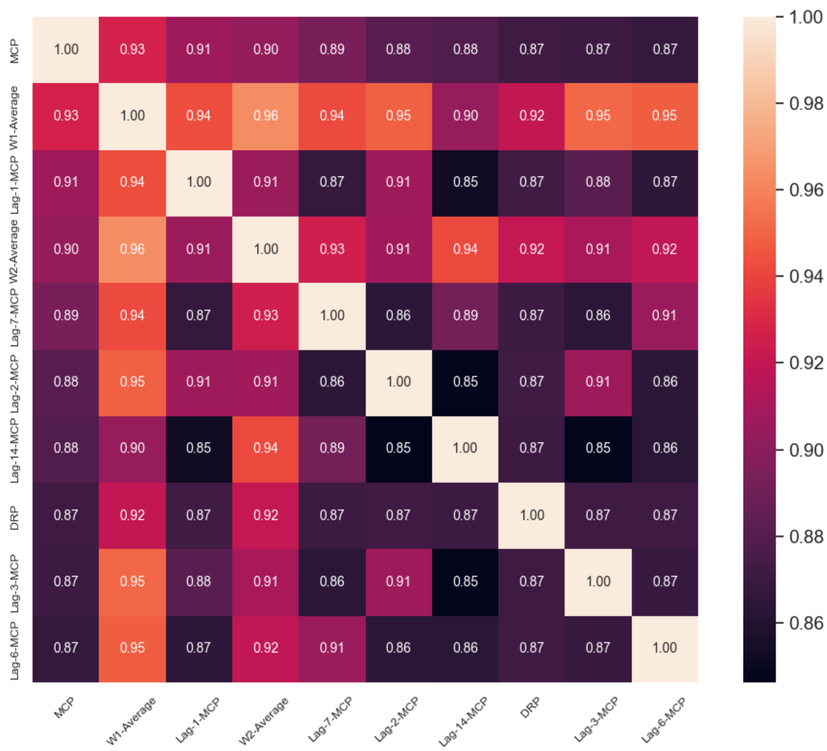


**Figure 25:** Heat Map for Correlation Analysis

## 5.5 Data Cleaning

Data cleaning is a crucial process that identifies and rectifies errors within a dataset to ensure its accuracy and integrity. This process is foundational for maintaining data quality and involves several routine activities. One common issue encountered in datasets is missing data, which can significantly impact the analysis results. It's essential to detect missing data and delve into the reasons behind its absence. The resulting data sets were verified using the 'df.sum()' method to ensure that there were no empty data fields.

Another crucial step in data cleaning is analyzing abnormalities or anomalies in the data, which are essentially data points that starkly deviate from the rest. These could be due to data entry errors or outliers. A box plot can illustrate the nature and cause of these anomalies, crucial for determining whether to correct or remove them. The previous section involved the drawing of the box-plot graphs. Beyond identifying and addressing issues, data correction and standardization are critical. Standardization ensures that the dataset has consistent formats and measurement units, making the data uniform and easier to analyze.

This thesis did not follow all the data cleaning steps. Box-plot graphics were examined to detect anomalies and check for any empty elements in the datasets. Some steps were intentionally omitted due to the nature of the real-time data used. These routine data cleaning activities highlight the importance of this process in research methodology by assessing the reliability and accuracy of the data. The resulting data sets were verified using the 'df.sum()' method to ensure the absence of any empty data fields.

## 5.6 Data Modeling and Evaluation

The specified date ranges divide the dataset into train and test sets, covering the period from the beginning of 2020 to the end of 2022. A multiple linear regression model was developed using the training data set to analyze the relationships between the independent variables and the dependent variable according to the short-term prediction model formulation. Independent and dependent variables (X and y) in the training and test sets were determined. The model was then applied to the independent variables in the test set (x_test) to generate predicted values (y_pred). These predictions were compared with the actual outcomes (y_test) within the test set, enabling the evaluation of the model's accuracy and the calculation of the error value.

MAE, R2, RMSE, and MAPE values were calculated. These values will also be shared. The MAPE was used as the primary metric to assess the model's performance. MAPE measures the size of the error in percentage terms, providing a clear indication of the accuracy of predictions in relation to the actual observed values. Figure 26 emphasizes the difference between the training data and the validation period. It demonstrates the baseline model's usage from December 25th to December 31st, 2022. This period was specifically chosen to validate the model's effectiveness in predicting near-term electricity prices based on historical data up to the end of the previous year.

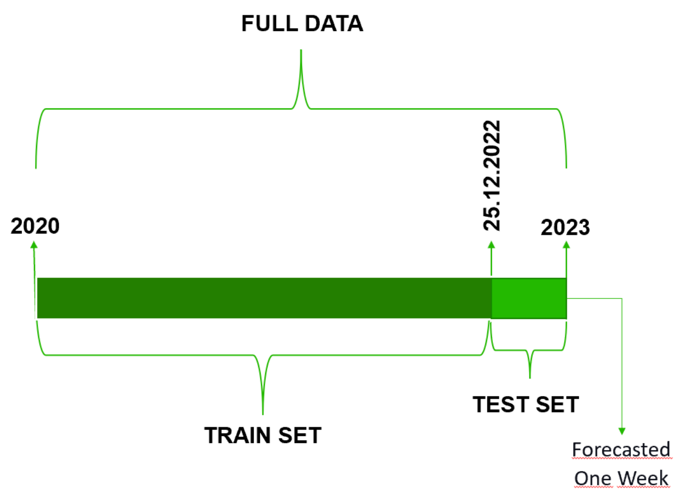


**Figure 26:** The Difference Between the Training Data and the Validation Period

In this study, a baseline model was created based on the short-time forecasting model. And in this model, different DRP variables were added to the formulation to measure the effect of the DRP variable. These models are called Enhanced Model Loop 1 and Enhanced Model Loop 2. The tabs below will share the model steps and results for each model. Additionally, residual analyses that are important in multiple linear regressions will be shared. Outlier analysis, which is not shown in the exploratory data analysis section, will be shared, and its comparison with the residuals in the model will be shared.

### 5.6.1 Baseline Model

Baseline Model, designed for short-term electricity price forecasting, uses a multiple linear regression approach, which effectively addresses the linearity between the dependent and independent variables. This model is crafted using parameters defined by the least square estimation method, aimed at minimizing prediction errors by fitting a linear equation to the observed data. Such an approach confirms the model's simplicity and clarity in interpreting the relationships among variables.

The core of the Baseline Model is represented by equation 8, where each variable's contribution to predicting electricity prices is quantified through regression coefficients. These coefficients, derived from data spanning April 1st, 2020, to December 25th, 2022, reflect the strength and direction of the influence each predictor has on the market price of electricity. This linear formulation allows for straightforward interpretation and application in predictive scenarios. The effectiveness of Baseline Model is quantified using the MAPE, which measures the accuracy of the predictions against actual market prices. MAPE Values are shown in table 8. Apart from MAPE, R2, MAE and RMSE error metrics were also calculated. Their values are shown in table 9. With a MAPE of 16.67% during the validation period from December 25th to December 31st, 2022, the model demonstrates a reliable performance in capturing and forecasting price trends. Comparing these results with other models or benchmarks from the literature helps validate the model's precision and capability. B coeffients and incercept values obtained in the models are given in the Appendix section.

**Table 8:** The MAPE for Baseline Model

| **The MAPE (%)** |
| --- |
| 16.67 |

**Table 9:** Error Metrics for Baseline Model

| $MAE$ | $R2$ | $RMSE$ |
| --- | --- | --- |
| 0.33 | 0.47 | 0.43 |

These error metrics can be interpreted like this: MAE represents the average absolute difference between estimated values and actual values. The equation is calculated on the basis of 11. In this case, the model's estimates make an average error of about 0.333 units. A low MAE indicates that the model usually makes correct predictions. However, the scale of the data set should determine how to assess this value. Since the MCP values are not very large, this error may be quite small and acceptable. The R2 score shows how well the model explains data variance. This is calculated by equation 14. A value close to 1 indicates that the model is performing very well, whereas a value near 0 does not adequately explain the model's data variance. Here, the R2 score is 0.477, which indicates that the model explains approximately 47.7% of the data variance. This performance can be considered intermediate. The model can explain a significant portion of the data variance, but there is also a variance that is still unexplained. The RMSE represents the square of the average square error between the estimated values and the actual values. The RMSE is calculated using equation 15. Here, the model's estimates make an average error of about 0.432 units. RMSE is used to determine whether errors are large or not. MAPE shows the average absolute percentage error of estimated values compared to real values. Equation 10 calculates this error metric. Here, the model's estimates make an average error of 16.67% compared to real values. The MAPE value of 16.67% indicates that the model generally performs well, but that the estimates have some deviations from the actual values. Generally, MAPE values below 10% are considered very good, while 10–20% can be considered good, 20–50% medium, and 50% above as poor performance.

The present study employs a short-term electricity price forecasting baseline model, represented by equation 8. Equation 8 coefficients were estimated using a dataset spanning three years of historical data, specifically from April 1st, 2020, to December 25th, 2022. The hourly cost of the baseline model was calculated over a period of one week using a dataset acquired from the power market. The estimates in figure 27 cover the period from December 25 to December 31, 2022, based on the training data for the basic model and the validation period.

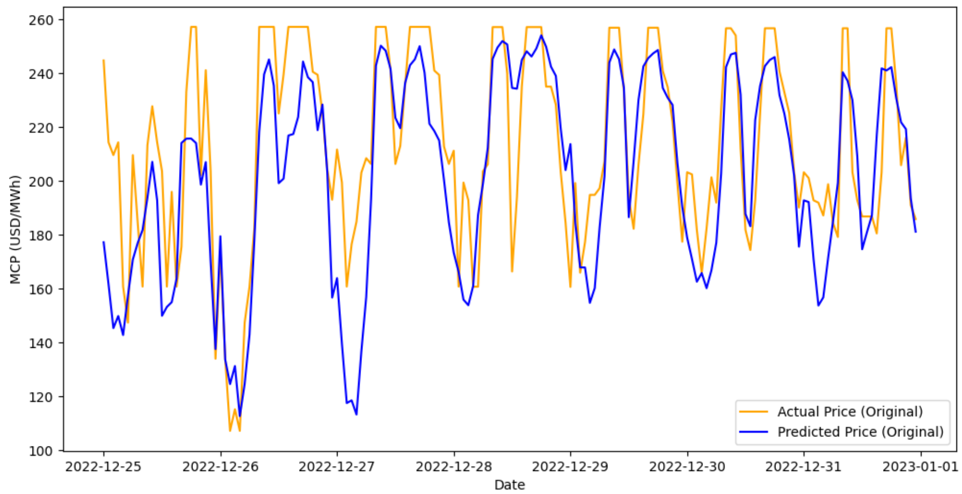


**Figure 27:** Forecasted Price Using the Baseline Model and Actual Price Over the Validation Period from December 25th to December 31st, 2022.

After estimates and the calculation of error metrics, residuals are calculated and analyzed. Residuals are calculated using the multiple linear regression formula in the eighth equation. The normal distribution of residuals increases the accuracy and validity of the model. That's why residual analysis is important. As a result, the following two graphs for residuals have been obtained: figures 28 and 29 show the model's prediction performance and the distribution of errors.

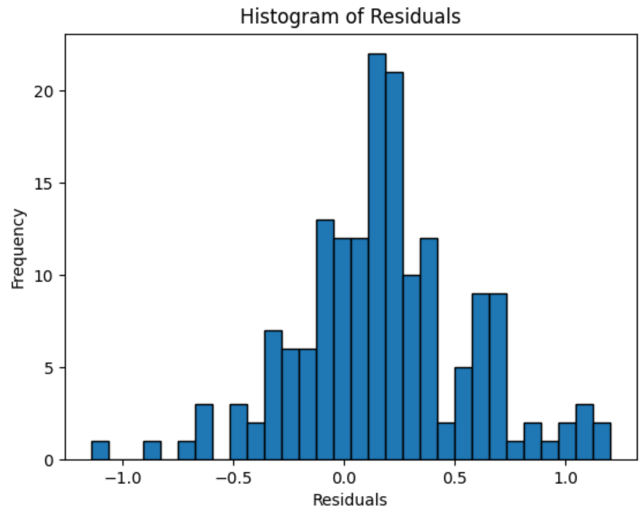


**Figure 28:** Histogram of Residuals

The histogram of the residuals chart shared in figure 28 shows the distribution of the differences between the predicted values and the actual values. The histogram's center concentrates around 0. This means that the predictions are usually correct. The distribution is not symmetrical, and there are slightly more values in the left tail. The concentration of the histogram in the center suggests that the model produces generally accurate predictions. However, the fact that the distribution is not symmetrical indicates that some errors (especially negative) are more frequent in the model than others. The residuals distribution should ideally resemble a normal distribution, signifying the random distribution of model errors and the absence of systematic errors in the model.

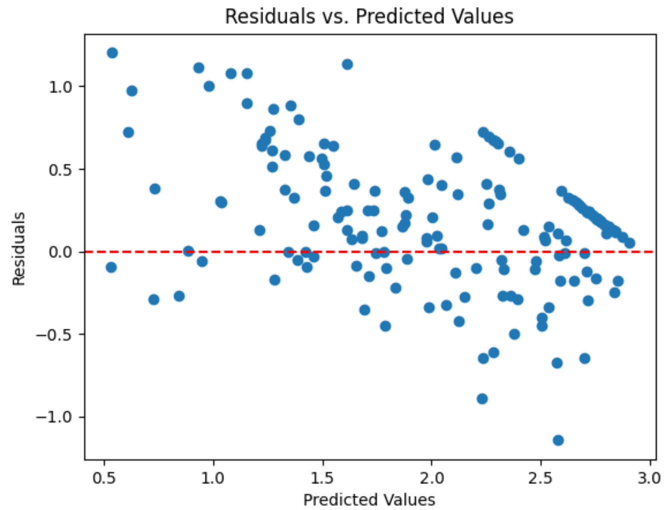


**Figure 29:** Residuals - Predicted Values

Residuals vs. shared in figure 29: predicted values the plot shows the relationship between predicted values and residuals (errors). Estimated values are on the horizontal axis, and residuals are on the vertical axis. The red dashed line (y = 0) on the graph represents the point where the error is zero. Data points are expected to be distributed around this line. The trend observed in this graph shows that residuals (errors) decrease as predicted values increase. The asymmetric distribution of data points around the red dashed line suggests the presence of systematic errors in the model. The fact that the residuals seem to decrease as the predicted values increase indicates that the model makes more errors at low values and fewer errors at high values. Ideally, residuals should have no correlation with predicted values. This indicates that the model exhibits a comparable error rate for every predicted value, and the errors follow a random distribution.

The residuals histogram shows the relationship between the residuals and predicted values. Plot charts are useful tools for examining the model's prediction errors. The distribution of the residuals and their relationship with the predicted values indicate areas where the model needs improvement. Using these graphs, the performance of the model can be evaluated and the necessary improvements can be made. Residuals analysis is an important step towards increasing the accuracy and validity of the model. Before moving to the enhanced model steps created by adding a new feature, a detailed outlier analysis was performed for the extreme values seen in the histogram. This may help to improve the model's performance.

Outlier analysis is an important step in improving model performance and identifying extreme end values in the data set. Outliers are data points that are generally at the extreme edge of the data and may have an adverse effect on the model's prediction accuracy. Various methods can be used for outlier analysis. The previous chapter indicated that he employed IQR analysis for outlier examination. The IQR method identifies and visualizes outliers. Using the boxplot and scatter plot, the graphs obtained in this step are shown below in figures 30 and 31, respectively.

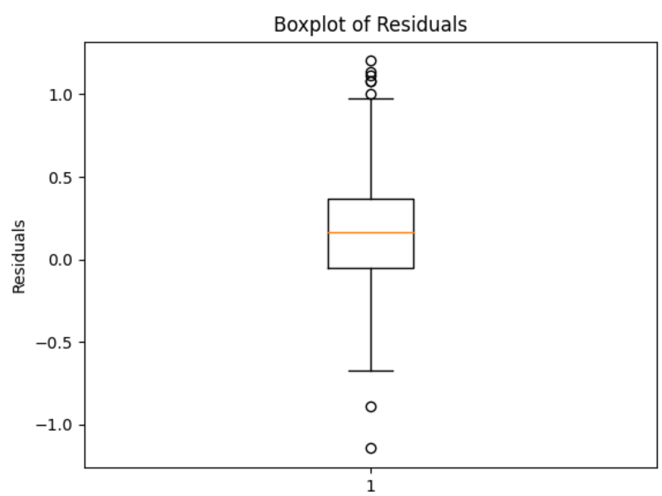


**Figure 30:** Box Plot of Residuals

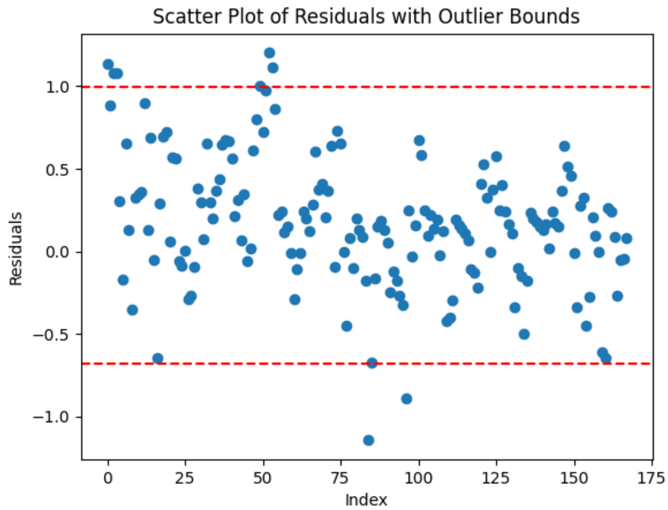


**Figure 31:** Scatter Plot of Residuals with Outliers

After the outliers were removed, no significant improvement was observed in MAE and R2 values. Therefore, no operation has been conducted, nor has an outlier operation been carried out to identify in detail the effects of the pandemic and regional conflicts. The model generally performs well when it comes to error metrics, but there are some areas that require improvement. These metrics serve as crucial indicators for assessing the model's performance and offer valuable suggestions for its enhancement. This study used MAPE to interpret the error metrics and write the results.

### 5.6.2 Enhanced Model

In order to investigate the impact of DRP, we extracted the DRP value from the features identified in the baseline model. The training dataset consist of values collected between April 1, 2020, and December 25, 2022. The first cycle of the enhanced model incorporated DRP values cyclically, starting one day in advance and extending up to 90 days. Each added value was subtracted in the next loop. Equation 8 represents this process. In the second cycle of the enhanced model, we cyclically incorporated DRP values into the model, starting one day in advance and extending up to 90 days.

Each added value remained in the next loop, calculated by adding the added value. Equation 8 represents this process.

Predictive values of one-week MCP data in both cycles after December 25 were obtained. Heat maps were created using MAPE values from the short-term forecast model prices during the validation period. The results, MAPE, R2, MAE, RMSE error metrics are presented separately in the relevant section tables. B coeffients and intercept values obtained in the models are given in the Appendix section. The MAPE value for the first model, Lag-85, is 16.17%. The MAPE values were obtained by applying each model separately. The plotted predicted and actual values obtained from the model with the lowest MAPE in a pool of 90 models are shown in figure 34. The MAPE value for the second model is shown in figure 41, Lag-3, is 17.05%. The MAPE values were obtained by applying each model separately.

### 5.6.3 First Loop For Enhanced Model

Visual and tabular presentations illustrate the process and results of the first loop for the enhanced model, showcasing the model's analytical approach and effectiveness. Figure 32 illustrates the flow chart for this first loop, detailing the sequential steps and data flow utilized in the model's computation. This visual representation offers a clear and structured view of how data inputs and processing steps culminate in the final predictions.

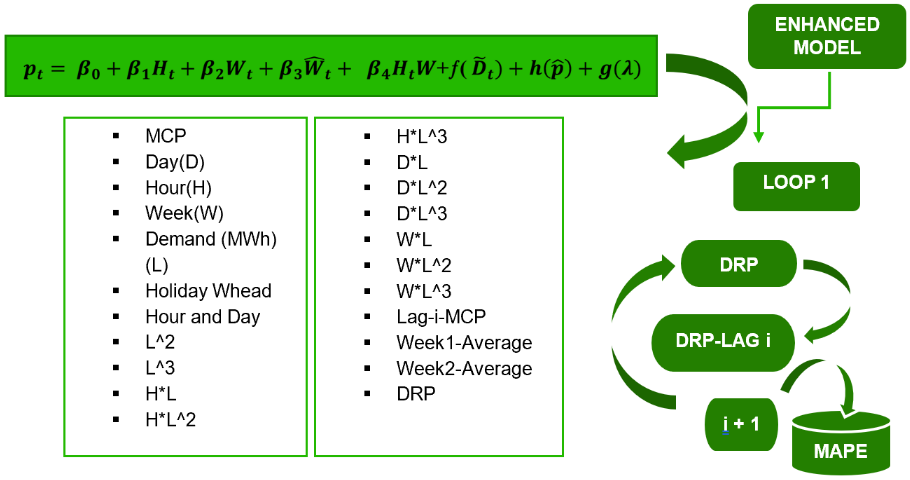


**Figure 32:** Flow Chart for First Enhanced Model

For this model, the MCP values for the week following December 25 were forecasted using the short-term forecasting model developed. After generating these predictions, a heat map was created to visualize the MAPE for each day within the validation period. Reflecting the model's prediction accuracy, this heat map reveals the distribution of error rates across different lags, with the minimum MAPE value recorded at 16,17% observed at Lag-85. This indicates that the model reached its peak accuracy when combining DRP values from about three months ago.

The detailed results from this analysis are presented in figure 33, which documents the MAPE values achieved through the individual execution of each model variant within the pool of 90 models. This table provides a quantitative insight into the model's performance, supporting the selection of the best model based on its predictive accuracy.

| MAPE (%) | | | | | |
|---|---|---|---|---|---|
| LAG1: | 17,06 | LAG31: | 16,95 | LAG61: | 16,49 |
| LAG2: | 17,03 | LAG32: | 16,96 | LAG62: | 16,42 |
| LAG3: | 16,99 | LAG33: | 17,19 | LAG63: | 16,43 |
| LAG4: | 16,99 | LAG34: | 17,02 | LAG64: | 16,39 |
| LAG5: | 16,98 | LAG35: | 16,89 | LAG65: | 16,37 |
| LAG6: | 16,98 | LAG36: | 16,92 | LAG66: | 16,40 |
| LAG7: | 16,98 | LAG37: | 16,92 | LAG67: | 16,41 |
| LAG8: | 16,98 | LAG38: | 16,91 | LAG68: | 16,40 |
| LAG9: | 16,98 | LAG39: | 16,89 | LAG69: | 16,40 |
| LAG10: | 16,98 | LAG40: | 16,86 | LAG70: | 16,39 |
| LAG11: | 16,98 | LAG41: | 16,85 | LAG71: | 16,42 |
| LAG12: | 16,99 | LAG42: | 16,84 | LAG72: | 16,40 |
| LAG13: | 16,99 | LAG43: | 16,84 | LAG73: | 16,30 |
| LAG14: | 16,99 | LAG44: | 16,84 | LAG74: | 16,26 |
| LAG15: | 16,98 | LAG45: | 16,84 | LAG75: | 16,24 |
| LAG16: | 16,98 | LAG46: | 16,83 | LAG76: | 16,24 |
| LAG17: | 16,98 | LAG47: | 16,82 | LAG77: | 16,25 |
| LAG18: | 16,97 | LAG48: | 16,83 | LAG78: | 16,23 |
| LAG19: | 16,96 | LAG49: | 16,81 | LAG79: | 16,23 |
| LAG20: | 16,96 | LAG50: | 16,78 | LAG80: | 16,25 |
| LAG21: | 16,96 | LAG51: | 16,75 | LAG81: | 16,25 |
| LAG22: | 16,93 | LAG52: | 16,71 | LAG82: | 16,25 |
| LAG23: | 16,94 | LAG53: | 16,68 | LAG83: | 16,22 |
| LAG24: | 16,96 | LAG54: | 16,63 | LAG84: | 16,19 |
| LAG25: | 16,94 | LAG55: | 16,61 | LAG85: | 16,17 |
| LAG26: | 16,92 | LAG56: | 16,55 | LAG86: | 17,00 |
| LAG27: | 16,92 | LAG57: | 16,54 | LAG87: | 17,00 |
| LAG28: | 17,08 | LAG58: | 16,55 | LAG88: | 16,86 |
| LAG29: | 16,97 | LAG59: | 16,56 | LAG89: | 16,95 |
| LAG30: | 17,13 | LAG60: | 16,50 | LAG90: | 16,96 |

**Figure 33:** Heat Map Values for First Enhanced Model

Calculations were made using MAPE error metrics to assess percentage error rates. When evaluating a model, one or more error metrics can be used to measure its accuracy and reliability. This study also shares R2, RMSE, and MAE values. Error metric values obtained in the models are given in the Appendix section.The appendix contains the R2 values calculated to evaluate the general descriptive power of the model, RMSE error values, which are more sensitive to the size of errors and give more weight to large errors, and MAE values, which represent the average absolute difference between predicted values and actual values.

These error metrics can be interpreted like this: MAE represents the average absolute difference between estimated values and actual values. The equation is calculated on the basis of 11. In this case, the model's estimates make an average error of about 0.333 units. A low MAE indicates that the model usually makes correct predictions. However, the scale of the data set should dictate how to assess this value. Since the MCP values are not very large, this error may be quite small and acceptable. The R2 score shows how well the model explains data variance. This is calculated by equation 14. A value close to 1 indicates that the model is performing very well, whereas a value near 0 does not adequately explain the model's data variance. Here, the R2 score is 0.477, which indicates that the model explains approximately 47.7% of the data variance. This performance can be considered intermediate. The model can explain a significant portion of the data variance, but there is also a variance that is still unexplained. The RMSE represents the square of the average square error between the estimated values and the actual values. The RMSE is calculated using equation 15. Here, the model's estimates make an average error of about 0.432 units. RMSE is used to determine whether errors are large or not. MAPE shows the average absolute percentage error of estimated values compared to real values. Equation 10 calculates the error metric. Here, the model's estimates make an average error of 16.67% compared to real values. The MAPE value of 16.67% indicates that the model generally performs well, but that the estimates have some deviations from the actual values. Generally, MAPE values below 10% are considered very good, while 10–20% can be considered good, 20–50% medium, and 50% above as poor performance.

The model generally performs well when it comes to error metrics, but there are some areas that require improvement. These metrics serve as crucial indicators for assessing the model's performance and offer valuable suggestions for its enhancement. This study used MAPE to interpret the error metrics and write the results. Furthermore, the actual and forecasted price graph, utilizing the parameters that resulted in the best MAPE value from this loop, is displayed in figure 34. This graph offers a direct visual comparison between the model's predictions and the actual prices during the validation period, underscoring the practical utility and accuracy of Enhanced Model in real-world applications. This graphical representation not only validates the model's effectiveness but also provides a clear and intuitive understanding of its performance in the context of electricity market forecasting.

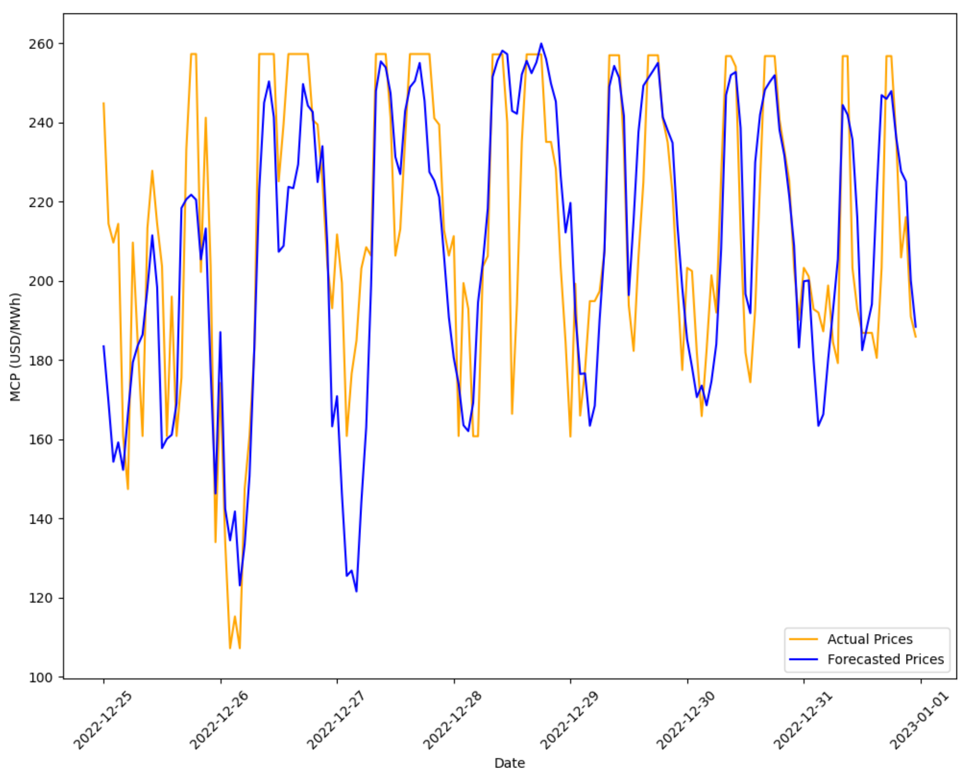


**Figure 34:** Comparison of Predicted Price and Actual Price for the First Enhanced Model

The surpluses calculated and analyzed in the Baseline Model are made under this model. As a result, the following two graphs of advantages have been obtained. Figure 35, the histogram graph shows the differences (residuals) between estimated values and actual values. In the graph, residuals usually range from -80 to 60. The histogram shows that residuals have an approximately normal distribution. This indicates that the model's errors are random and do not have a specific layout. Most residual values are concentrated around zero, which indicates that the model generally makes correct estimates. The chart contains end values such as -80 and 60. This indicates that the model made major mistakes in some estimates. The Histogram of Residuals graph shows that the overall performance of the model is good, but that some end values increase the model's errors. The normal distribution of residuals indicates that the errors in the model are random and that the model usually makes correct predictions.

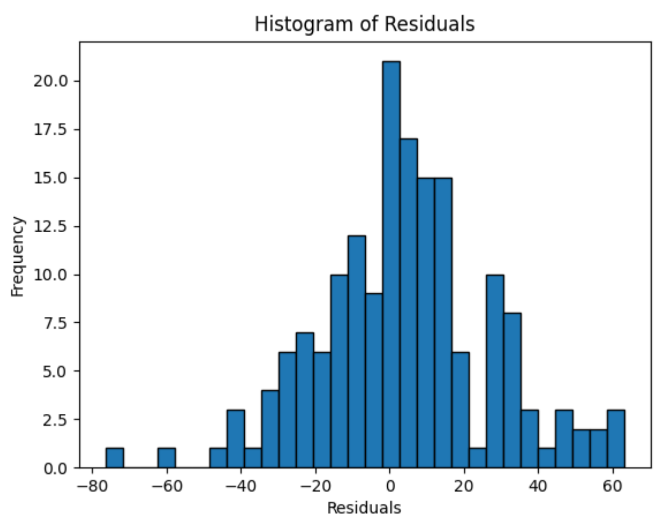


**Figure 35:** Histogram of Residuals for First Enhanced Model

Figure 36, Residuals vs. Predicted Values In the Plot Graph, the estimated values (x axis) and residuals (y axis), are shown. Estimated values range from approximately 120 to 260. Residuals are spread up and down from the estimated values. It seems that this spread does not have a specific arrangement. The red cut line (y=0) represents the average of the residuals. Residuals are scattered symmetrically around this line.

However, a small amount of negative bias is observed in residuals. This may indicate that the model has systematically low estimates while predicting higher values. The Residuals vs. Predicted Values graph shows that the model's predicted errors are randomly distributed against the estimated values, but have a slightly negative trend. This indicates that the overall performance of the model is good, but that the model can systematically make errors at some high estimates. The lack of a specific layout of the distribution in the graph indicates that the model's errors are largely random, and that it does not make excessive or inadequate predictions.

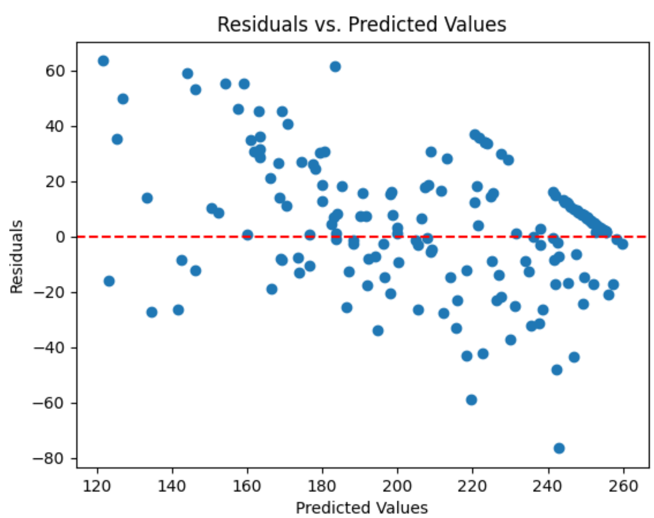


**Figure 36:** Residuals-Predicted Values for First Enhanced Model

Overall, the performance of the model looks good when you look at these two graphs. The Histogram of Residuals graph shows that residuals have a normal distribution and that the model generally makes correct estimates. Residuals vs. The Predicted Values graph shows that the model's errors are randomly distributed, but there is a negative trend at some higher values. This suggests that there are some areas in the model that need to be improved, especially for higher values, where predictions should be made more carefully.

The two charts, presented below in order, aid in evaluating the predictive errors of the model and identifying potential contradictory values. The boxplot of residuals graph, shown in figure 37, shows the central trend and spread of residuals. The median (middle line) residual value is around 0, which indicates that the average of the residuals is 0. The box shows the interval between the 25th and 75th percentiles of residuals (IQR). The bottom and top lines (whiskers) show the lowest and highest values of residuals (excluding outliers). Points outside the box have the potential to have contradictory values. There are contradictory values at both the bottom and upper ends. There are several contradictory values at the upper edge, specifically around the 60 level. At the bottom, there are opposite values around -80. The Residuals Boxplot graph reveals that the residuals typically exhibit a central trend of 0, with a significant distribution between the 25th and 75th percentiles. However, there are some contradictory values, both positive and negative. These contradictory values indicate that the model makes major errors in some cases, and that the causes of these errors can improve the model's research performance.

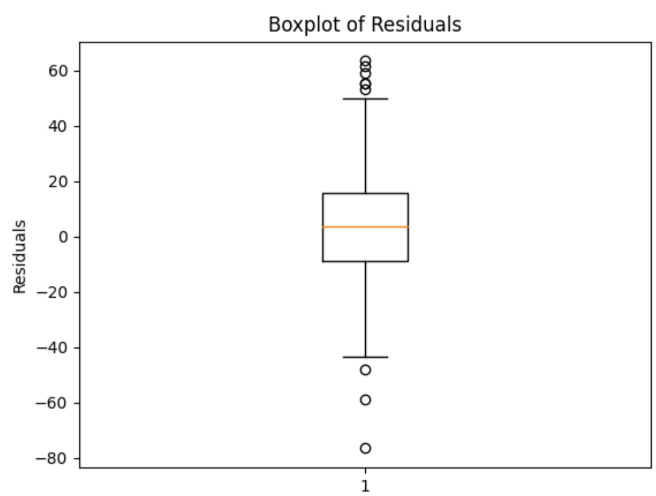


**Figure 37:** Box Plot of Residuals

The scatter plot of residuals with outlier bounds, shown in figure 38, shows residuals (y axis) and indexes (x axis). Residuals range from about -80 to 60. Without any specific arrangement, the residuals spread up and down. This spread indicates that the model's errors are random. Red cut lines indicate the outlier boundaries (bottom and upper boundaries). Points outside these boundaries have the potential to be contradictions. In the graph, there are several opposite values in both positive and negative directions. The scatter plot of residuals with outlier bounds reveals a largely random distribution of the model's predictive errors, albeit with the presence of some extreme values. Correct values indicate that the model makes major mistakes in some cases. Investigating the root causes of these contradictions can improve the model's performance.

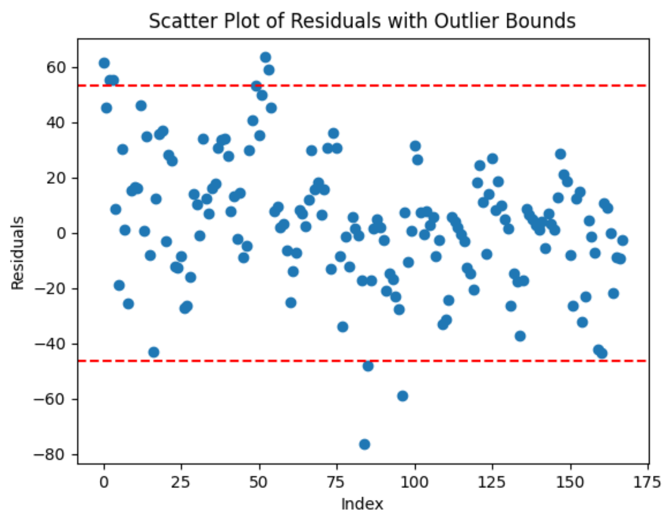


**Figure 38:** Scatter Plot of Residuals with Outliers

The scatter plot of residuals with outlier bounds and the boxplot of residuals graphs demonstrate a generally random distribution of the model's predictive errors, albeit with some extreme values. Examining these contradictory values could be a significant step towards improving the model's performance. Understanding the reasons for the differential values can be used to increase the accuracy of the model.

### 5.6.4 Second Loop For Enhanced Model

In the second loop of Enhanced Model, the approach to analyzing the impact of lagged values on the MCP deviates from that of the first loop. Instead of subtracting Lag-n values in each iteration, this loop involves adding lag values progressively to examine their cumulative effect on forecasting accuracy. The flow chart for the second loop of Enhanced Model is illustrated in figure 39.

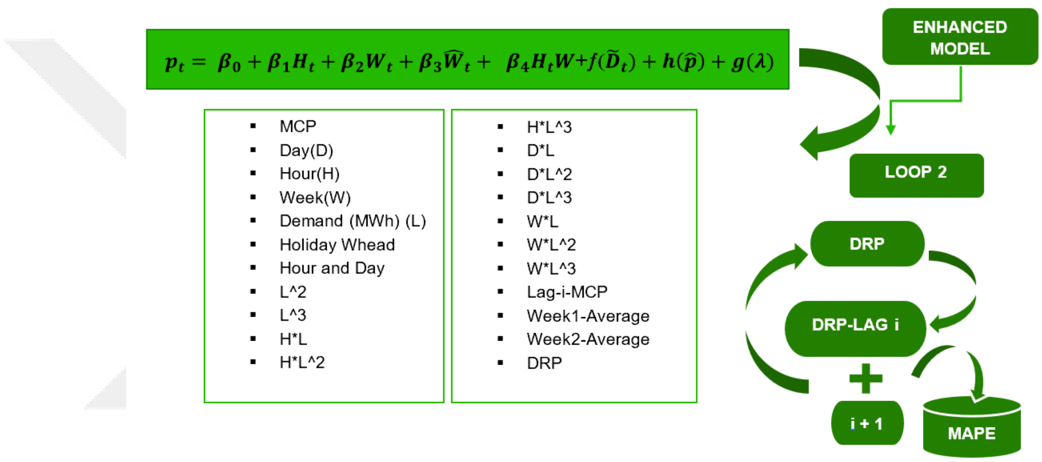


**Figure 39:** Flow Chart for Second Enhanced Model

A detailed heat map was generated to visualize the MAPE across various lags, showcasing how each additional lag influences the model's predictive performance. This visual representation helps in pinpointing the specific lag that optimizes forecasting accuracy. The results of this analysis are systematically presented in figure 40, which documents the MAPE values for each model run independently within the loop.

| MAPE (%) | | | | | |
|---|---|---|---|---|---|
| LAG1: | 17,06 | LAG31: | 17,86 | LAG61: | 18,54 |
| LAG2: | 17,05 | LAG32: | 18,54 | LAG62: | 18,51 |
| LAG3: | 17,05 | LAG33: | 18,57 | LAG63: | 18,57 |
| LAG4: | 17,08 | LAG34: | 18,59 | LAG64: | 18,40 |
| LAG5: | 17,10 | LAG35: | 18,76 | LAG65: | 18,40 |
| LAG6: | 17,10 | LAG36: | 18,86 | LAG66: | 18,79 |
| LAG7: | 17,10 | LAG37: | 18,86 | LAG67: | 18,81 |
| LAG8: | 17,10 | LAG38: | 18,88 | LAG68: | 18,77 |
| LAG9: | 17,10 | LAG39: | 18,85 | LAG69: | 18,80 |
| LAG10: | 17,10 | LAG40: | 18,72 | LAG70: | 18,78 |
| LAG11: | 17,11 | LAG41: | 18,63 | LAG71: | 18,83 |
| LAG12: | 17,11 | LAG42: | 18,57 | LAG72: | 18,82 |
| LAG13: | 17,11 | LAG43: | 18,58 | LAG73: | 18,50 |
| LAG14: | 17,10 | LAG44: | 18,57 | LAG74: | 18,42 |
| LAG15: | 17,17 | LAG45: | 18,70 | LAG75: | 18,24 |
| LAG16: | 17,17 | LAG46: | 18,67 | LAG76: | 18,21 |
| LAG17: | 17,17 | LAG47: | 18,61 | LAG77: | 18,31 |
| LAG18: | 17,16 | LAG48: | 18,70 | LAG78: | 18,33 |
| LAG19: | 17,16 | LAG49: | 18,76 | LAG79: | 18,35 |
| LAG20: | 17,18 | LAG50: | 18,64 | LAG80: | 18,35 |
| LAG21: | 17,18 | LAG51: | 18,70 | LAG81: | 18,41 |
| LAG22: | 17,11 | LAG52: | 18,65 | LAG82: | 18,41 |
| LAG23: | 17,11 | LAG53: | 18,48 | LAG83: | 18,43 |
| LAG24: | 17,15 | LAG54: | 18,22 | LAG84: | 18,35 |
| LAG25: | 17,16 | LAG55: | 18,25 | LAG85: | 18,09 |
| LAG26: | 17,25 | LAG56: | 18,09 | LAG86: | 18,10 |
| LAG27: | 17,35 | LAG57: | 18,09 | LAG87: | 18,12 |
| LAG28: | 18,40 | LAG58: | 18,32 | LAG88: | 18,00 |
| LAG29: | 18,30 | LAG59: | 18,44 | LAG89: | 17,97 |
| LAG30: | 18,38 | LAG60: | 18,56 | LAG90: | 17,95 |

**Figure 40:** Heat Map Values for Second Echanced Model

The lowest MAPE recorded was %17.05, observed at Lag-3, indicating that this particular lag configuration provides the most accurate predictions for the MCP. This figure highlights the precision of the forecasts generated by the model with the lowest MAPE from a pool of 90 models, demonstrating its robustness in accurately predicting short-term electricity prices. This study also shares R2, RMSE, and MAE values. Error metric values obtained in the models are given in the Appendix section.

Furthermore, the actual and forecasted price graph, utilizing the parameters that resulted in the best MAPE value from this loop, is displayed in figure 41. This graph offers a direct visual comparison between the model's predictions and the actual prices during the validation period, underscoring the practical utility and accuracy of Enhanced Model in real-world applications. This graphical representation not only validates the model's effectiveness but also provides a clear and intuitive understanding of its performance in the context of electricity market forecasting.

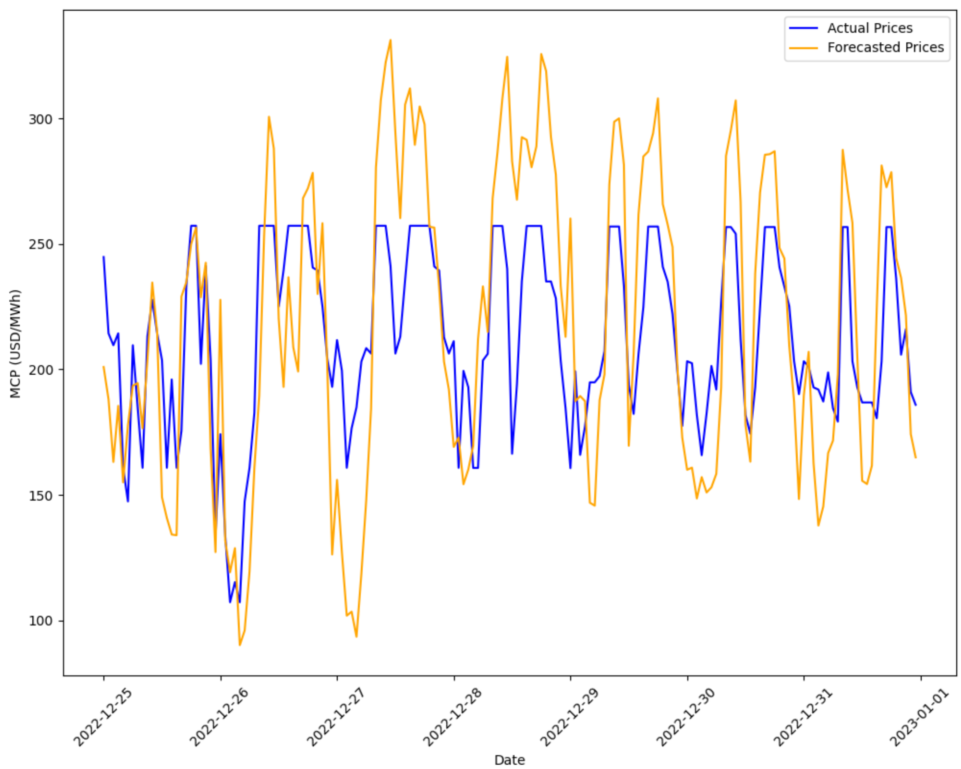


**Figure 41:** Comparison of Predicted Price and Actual Price for the Second Enhanced Model

### 5.6.5 Summary of Results

The baseline model in the thesis employs linear regression techniques to establish the foundational relationship between natural gas and electricity prices in Türkiye. Using daily data from the EXIST transparency platform, it analyzes how these prices interact, leveraging key performance metrics like mean absolute error and mean absolute percentage error to gauge predictive accuracy. Although this model offers a basic level of prediction, it serves as the groundwork for more sophisticated enhancements.

Subsequently, the advanced model was created, the approach was enhanced, and further data was used to enhance the accuracy of predictions. Loop one encompassed more comprehensive data, including hourly prices and other market characteristics. By adding the Lag-DRP parameters little by little, the baseline model got more accurate in terms of MAPE values compared to the original baseline. The LAG-85 had the lowest MAPE value. This data indicated that there was a delay of around 85 days between the daily period and the impact of changes in natural gas prices.

In loop second, the identical parameters that were taken into account in loop one were utilized. The incremental addition of Lag DRP parameters and their retention in subsequent cycles resulted in lower accuracy, as indicated by the MAPE values, when compared to the baseline. Through this iteration, it became evident that the inclusion of Lag-DRP data led to a substantial rise in the prediction errors. Both loops of the enhanced model confirm that natural gas prices impact electricity prices with a time lag, and sophisticated modeling is crucial to capture this relationship. This iterative modeling process provides valuable insights into managing Türkiye's energy market and informs policy decisions.

The research findings point to a significant time-lag effect in the Turkish energy market, specifically regarding the impact of natural gas prices on electricity pricing. This time-lag effect has been determined to occur roughly 85 days after fluctuations in natural gas prices. The accuracy of the model has been evaluated using the error metric mean absolute percentage error.

# CHAPTER VI

# DISCUSSION

In this study, the initial analysis of the dataset revealed that the data did not exhibit a normal distribution, as indicated by the histogram plots. This discovery sparks a discussion about the appropriateness of data transformations, such as logarithmic transformations, which are commonly used to standardize data distributions. However, transformations were not employed in this instance primarily because the regression techniques used—linear and multiple linear regression—do not require the dependent or independent variables to follow a normal distribution. The nature of the regression methods used supported the decision to proceed without transformation because they are robust to non-normality in the explanatory variables.

Preliminary investigations indicating a linear relationship between the key variables of interest—natural gas prices and electricity prices—led to the decision to employ linear and multiple linear regression methods. These methods were selected because they provide a clear framework for understanding the influence of independent variables on a dependent variable and are particularly effective for prediction when relationships between variables are linear. The use of multiple linear regression allowed for the inclusion of several explanatory variables, providing a comprehensive model that accounts for various factors simultaneously affecting electricity prices.

To establish the linear relationships between the parameters, statistical techniques such as correlation analysis and scatter plots were employed. These methods facilitated the visualization and quantitative assessment of relationships between variables. Correlation coefficients quantified the strength and direction of these relationships, providing the necessary statistical basis for employing linear modeling techniques.

This approach played a crucial role in confirming the hypothesis that changes in natural gas prices are linearly associated with shifts in electricity pricing, justifying the use of linear regression models in the study.

A critical question arising from this study is whether future electricity price estimations can be made using the models developed here. The results suggest that these models can indeed forecast future electricity prices, albeit within a specified range and under certain conditions. However, the inherent unpredictability of energy markets and the linear nature of the models themselves constrain their accuracy. Therefore, while these models provide valuable baseline predictions, their limitations should be acknowledged, particularly in cases where extreme market volatility or structural changes occur.

Another important methodological decision concerns the division of the dataset into training and testing subsets. In existing literature, it's common practice to allocate the training and test sets using either an 80-20% or a 75-25% split. This study followed a similar approach, but with modifications tailored to the unique characteristics of the dataset. The training set encompassed the majority of the data to optimize model learning, while the test set comprised a smaller portion to ensure robust validation.

Over a two-year period, the selection of 15 days for the 'y_pred' test set was a deliberate decision based on several factors. First, it aimed to strike a balance between providing sufficient data for meaningful testing and minimizing the risk of overfitting. Second, this period allowed the model to account for short-term fluctuations while still offering practical predictive scope. Despite potential concerns about underselection, the study's objectives and the unique characteristics of the data deemed this selection appropriate. Underselecting the test period could lead to errors, but a careful evaluation showed that the chosen period provided a reliable benchmark for assessing the model's predictive power.

Given that this study builds on prior work and has specifically applied its methodology to the Turkish energy market, the potential for broader applicability in other countries is a key consideration. The framework used here can be adapted to other energy markets with suitable modifications to account for local regulatory environments, market dynamics, and energy sources. Given the inherent flexibility of linear and multiple linear regression models, adjusting them for varying market conditions could yield valuable predictive insights. The model's success in the Turkish market demonstrates that similar relationships between natural gas and electricity prices likely exist in other regions, though further research would be required to validate this hypothesis.

Another significant finding is the observed 85-day lag in the Turkish market between fluctuations in natural gas prices and their impact on electricity prices. This period is crucial for policymakers and market participants to understand, as it provides a window for strategic decision-making in response to price changes. If the lag period were longer than 60 days, this could result in delayed policy responses, potentially exacerbating market volatility. Conversely, a shorter lag would require more rapid adjustments, posing challenges for energy suppliers and consumers alike. The study provides useful insights for anticipating electricity price changes, enabling stakeholders to plan ahead and mitigate potential risks.

Regarding the impact of recently discovered natural gas reserves in Türkiye, it's anticipated that this new supply will influence electricity prices positively in the long term. However, the timeline for these effects to manifest is dependent on the reserves' development pace, integration into the national energy grid, and the extent to which this new supply offsets imports. Initially, the direct impact on electricity prices may be gradual, but as the reserves become a more significant part of the energy mix, they could contribute to stabilizing or reducing electricity prices.

# CHAPTER VII

# CONCLUSIONS

This study examines the influence of natural gas market prices on electricity price forecasting. This study analyzes the delayed effects of natural gas prices on electricity pricing in Türkiye's day-ahead market after the pandemic. EXIST operates transparent marketplaces for determining natural gas and electricity prices in Türkiye. The relationship between natural gas and power pricing has become increasingly important due to Europe's sudden and unpredictable increases in natural gas costs following the pandemic and regional wars, which have disrupted the balance between fuel supply and demand. The current study has employed regression techniques to analyze the timing of the impact of changes in natural gas prices on electricity prices. The empirical analysis was carried out using data acquired from the EXIST transparency platform.

The research findings point to a significant time-lag effect in the Turkish energy market, specifically regarding the impact of natural gas prices on electricity pricing. This time-lag effect has been determined to occur roughly 60 days after fluctuations in natural gas prices. The accuracy of the model has been evaluated using different error metrics, such as MAE, MSE, MAPE, MPE, R2 and RMSE. The paper specifically reports the results of MAPE. The error metrics utilized in this study yield results that indicate a strong level of reliability, thereby providing additional support for the validity of our findings.

The findings of this study hold great importance for policymakers, participants in the energy market, and researchers. They contribute to a deeper comprehension of the interplay between these two crucial energy resources. Policymakers are responsible for making informed decisions regarding policies that promote the efficient operation of the energy market and guarantee the security of the energy supply. This tool enables the identification of strategic decisions and the development of risk management strategies for players, suppliers, and consumers in the energy market. Moreover, this comprehension can offer essential direction for the industry's strategic decision-making and attempts to reduce risks.

Examining the relationship between natural gas and electricity pricing is crucial in energy markets and economic analysis. The findings of this specific study can be used to understand and efficiently manage similar linkages in various energy markets. Diverse energy markets can extend and implement the research findings, which have wide-ranging implications. The study highlights the importance of the time delay and the underlying mechanisms by which changes in natural gas prices affect electricity pricing. This discovery has the potential to help us understand similar patterns in other energy markets.

It is crucial to acknowledge that this study has specific constraints. While the market conditions that have emerged following the epidemic are unique, they may only partially represent the market's long-term trends and dynamics. Furthermore, we must recognize that these findings may have limited applicability to different markets or periods. Nevertheless, it is worth exploring the potential of testing the proposed methodology in alternative market contexts. Subsequent investigations may enhance the scope of this study by encompassing a wider array of energy sources, varying time-lag intervals, or incorporating supplementary market variables that could impact the correlation between natural gas and electricity prices.

# APPENDIX A

# APPENDIX

## *A.1 B Coefficients and Intercepts*

B coefficients and intercepts coefficients for the Baseline Model, First Loop Enhanced Model, and Second Loop Enhanced Model obtained from the regression analysis are shared in detail below.

### A.1.1 B Coefficients for Baseline Model

The intercept coefficient for the Baseline Model obtained from the regression analysis is 2.456. The B coefficients for Baseline Model obtained from the regression analysis are as follows:

| **Variable** | **Coefficient** |
|---|---|
| Hour(H) | -0.391 |
| Day(D) | 0.809 |
| Week(W) | -0.399 |
| Holiday(Whead) | 1.796 |
| Hour and Day | -1.974 |
| Demand | 0.726 |
| $L^2$ | -1.569 |
| $L^3$ | 1.543 |
| H*L | -0.559 |
| $H*L^2$ | 3.954 |
| $H*L^3$ | -3.992 |
| D*L | 1.446 |
| $D*L^2$ | 0.343 |
| $D*L^3$ | 0.071 |
| W*L | 0.039 |
| $W*L^2$ | 0.017 |

| | |
|---|---|
| $W*L^3$ | 0.026 |
| Lag-1-MCP | -0.026 |
| Lag-2-MCP | 0.195 |
| Lag-3-MCP | -0.057 |
| Lag-4-MCP | -0.011 |
| Lag-5-MCP | -0.021 |
| Lag-6-MCP | 0.010 |
| Lag-7-MCP | -0.021 |
| Lag-8-MCP | -0.005 |
| Lag-9-MCP | 0.154 |
| Lag-10-MCP | 0.101 |
| Lag-11-MCP | 0.007 |
| Lag-12-MCP | 0.136 |
| Lag-13-MCP | -0.081 |
| Lag-14-MCP | 0.290 |
| W1-Average | -0.095 |
| W2-Average | -0.017 |

| | |
|---|---|
| DRP | -0.000 |

### A.1.2 B Coefficients for First Loop Enhanced Model

The intercept coefficient in LAG 3, which is the best MAPE value among the 90 models of the first enhanced model obtained as a result of the regression analysis, is -0.003995.The B coefficients for the First Loop Enhanced Model obtained from the regression analysis are as follows:

| Variable | Coefficient |
|---|---:|
| Hour(H) | 0.574 |
| Day(D) | -1.074 |
| Week(W) | 0.541 |
| Holiday(Whead) | -0.166 |
| Hour and Day | 0.308 |
| Demand | -0.154 |
| $L^2$ | 0.181 |
| $L^3$ | -0.450 |
| H*L | 0.242 |
| $H*L^2$ | -0.742 |
| $H*L^3$ | 1.493 |
| D*L | -0.749 |
| $D*L^2$ | 0.339 |
| $D*L^3$ | 0.066 |
| W*L | 0.034 |
| $W*L^2$ | 0.020 |

| | |
|---|---|
| W*L$^3$ | 0.021 |
| Lag-1-MCP | -0.029 |
| Lag-2-MCP | 0.192 |
| Lag-3-MCP | -0.056 |
| Lag-4-MCP | -0.012 |
| Lag-5-MCP | -0.027 |
| Lag-6-MCP | 0.006 |
| Lag-7-MCP | -0.024 |
| Lag-8-MCP | -0.002 |
| Lag-9-MCP | 0.148 |
| Lag-10-MCP | 0.097 |
| Lag-11-MCP | 0.005 |
| Lag-12-MCP | -0.009 |
| Lag-13-MCP | 0.387 |
| Lag-14-MCP | -0.220 |
| W1-Average | -0.088 |
| W2-Average | -0.021 |

| | |
|---|---|
| DRP | 0.065 |
| DRP_LAG85 | 0.075 |

### A.1.3 B Coefficients for Second Loop Enhanced Model

The intercept coefficient in LAG 85, which is the best MAPE value among the 90 models of the second enhanced model obtained as a result of the regression analysis, is -0.0002357.The B coefficients for the Second Loop Enhanced Model obtained from the regression analysis are as follows:

| **Variable** | **Coefficient** |
|---|---:|
| Hour(H) | 0.518 |
| Day(D) | -0.986 |
| Week(W) | 0.509 |
| Holiday(Whead) | -0.170 |
| Hour and Day | 0.343 |
| Demand | -0.192 |
| $L^2$ | 0.178 |
| $L^3$ | -0.449 |
| H*L | 0.243 |
| $H*L^2$ | -0.667 |
| $H*L^3$ | 1.342 |
| D*L | -0.670 |
| $D*L^2$ | 0.346 |
| $D*L^3$ | 0.071 |
| W*L | 0.037 |
| $W*L^2$ | 0.019 |

| | |
|---|---|
| W*L$^{3}$ | 0.023 |
| Lag-1-MCP | -0.027 |
| Lag-2-MCP | 0.196 |
| Lag-3-MCP | -0.052 |
| Lag-4-MCP | -0.011 |
| Lag-5-MCP | -0.025 |
| Lag-6-MCP | 0.009 |
| Lag-7-MCP | -0.021 |
| Lag-8-MCP | -0.001 |
| Lag-9-MCP | 0.154 |
| Lag-10-MCP | 0.100 |
| Lag-11-MCP | 0.008 |
| Lag-12-MCP | 0.377 |
| Lag-13-MCP | -0.244 |
| Lag-14-MCP | -0.047 |
| W1-Average | 0.014 |
| W2-Average | 0.218 |

| | |
|---|---|
| DRP_LAG1 | -0.078 |
| DRP_LAG2 | -0.087 |
| DRP_LAG3 | -0.020 |

## *A.2 Error Metrics*

MAE, R2, RMSE and MAPE error metric values for the Baseline Model and Enhanced Model obtained from the regression analysis are shared in detail below.

### A.2.1 Error Metrics for First Loop Enhanced Model

The error metrics in LAG 3, which is the best MAPE value among 90 models of the first enhanced model obtained from the regression analysis, are as follows:

| DRP_LAG | MAE | $R^2$ | RMSE | MAPE |
|---|---|---|---|---|
| 1 | 0.340 | 0.453 | 0.442 | 17.062 |
| 2 | 0.339 | 0.453 | 0.442 | 17.029 |
| 3 | 0.338 | 0.453 | 0.442 | 16.984 |
| 4 | 0.339 | 0.454 | 0.441 | 16.983 |
| 5 | 0.339 | 0.455 | 0.441 | 16.983 |
| 6 | 0.339 | 0.455 | 0.441 | 16.983 |
| 7 | 0.339 | 0.455 | 0.441 | 16.983 |
| 8 | 0.339 | 0.456 | 0.441 | 16.982 |
| 9 | 0.339 | 0.456 | 0.441 | 16.983 |
| 10 | 0.339 | 0.456 | 0.441 | 16.982 |
| 11 | 0.339 | 0.457 | 0.441 | 16.983 |
| 12 | 0.339 | 0.456 | 0.441 | 16.986 |
| 13 | 0.339 | 0.456 | 0.441 | 16.985 |
| 14 | 0.339 | 0.458 | 0.440 | 16.984 |
| 15 | 0.339 | 0.455 | 0.441 | 16.983 |
| 16 | 0.338 | 0.457 | 0.440 | 16.979 |

| | | | | |
|---|---|---|---|---|
| 17 | 0.338 | 0.459 | 0.440 | 16.975 |
| 18 | 0.338 | 0.458 | 0.440 | 16.971 |
| 19 | 0.338 | 0.459 | 0.440 | 16.963 |
| 20 | 0.338 | 0.459 | 0.440 | 16.960 |
| 21 | 0.338 | 0.459 | 0.440 | 16.955 |
| 22 | 0.337 | 0.463 | 0.438 | 16.930 |
| 23 | 0.337 | 0.462 | 0.439 | 16.941 |
| 24 | 0.338 | 0.459 | 0.440 | 16.955 |
| 25 | 0.337 | 0.460 | 0.439 | 16.945 |
| 26 | 0.336 | 0.464 | 0.437 | 16.919 |
| 27 | 0.337 | 0.463 | 0.438 | 16.923 |
| 28 | 0.341 | 0.448 | 0.444 | 17.076 |
| 29 | 0.341 | 0.454 | 0.442 | 16.974 |
| 30 | 0.341 | 0.453 | 0.442 | 17.129 |
| 31 | 0.337 | 0.468 | 0.436 | 16.946 |
| 32 | 0.339 | 0.469 | 0.436 | 16.964 |
| 33 | 0.343 | 0.457 | 0.441 | 17.187 |

| | | | | |
|---|---|---|---|---|
| 34 | 0.339 | 0.464 | 0.437 | 16.952 |
| 35 | 0.336 | 0.470 | 0.435 | 16.892 |
| 36 | 0.336 | 0.470 | 0.435 | 16.920 |
| 37 | 0.336 | 0.471 | 0.435 | 16.923 |
| 38 | 0.336 | 0.472 | 0.434 | 16.906 |
| 39 | 0.335 | 0.475 | 0.433 | 16.887 |
| 40 | 0.334 | 0.478 | 0.432 | 16.865 |
| 41 | 0.333 | 0.481 | 0.431 | 16.847 |
| 42 | 0.333 | 0.482 | 0.430 | 16.838 |
| 43 | 0.333 | 0.481 | 0.430 | 16.842 |
| 44 | 0.333 | 0.482 | 0.430 | 16.836 |
| 45 | 0.333 | 0.484 | 0.430 | 16.840 |
| 46 | 0.332 | 0.485 | 0.429 | 16.831 |
| 47 | 0.332 | 0.487 | 0.428 | 16.819 |
| 48 | 0.332 | 0.485 | 0.429 | 16.829 |
| 49 | 0.332 | 0.487 | 0.428 | 16.812 |
| 50 | 0.330 | 0.493 | 0.426 | 16.778 |

| | | | | |
|---|---|---|---|---|
| 51 | 0.329 | 0.496 | 0.424 | 16.748 |
| 52 | 0.327 | 0.501 | 0.422 | 16.713 |
| 53 | 0.326 | 0.505 | 0.421 | 16.683 |
| 54 | 0.324 | 0.510 | 0.419 | 16.630 |
| 55 | 0.323 | 0.511 | 0.418 | 16.607 |
| 56 | 0.321 | 0.517 | 0.415 | 16.555 |
| 57 | 0.320 | 0.520 | 0.414 | 16.539 |
| 58 | 0.320 | 0.517 | 0.415 | 16.553 |
| 59 | 0.320 | 0.516 | 0.416 | 16.560 |
| 60 | 0.318 | 0.520 | 0.414 | 16.505 |
| 61 | 0.317 | 0.524 | 0.413 | 16.485 |
| 62 | 0.315 | 0.528 | 0.411 | 16.424 |
| 63 | 0.315 | 0.528 | 0.411 | 16.430 |
| 64 | 0.313 | 0.531 | 0.409 | 16.391 |
| 65 | 0.313 | 0.531 | 0.409 | 16.369 |
| 66 | 0.314 | 0.526 | 0.412 | 16.403 |
| 67 | 0.315 | 0.525 | 0.412 | 16.415 |

| | | | | |
|---|---|---|---|---|
| 68 | 0.314 | 0.527 | 0.411 | 16.400 |
| 69 | 0.314 | 0.527 | 0.411 | 16.399 |
| 70 | 0.314 | 0.527 | 0.411 | 16.395 |
| 71 | 0.315 | 0.523 | 0.413 | 16.424 |
| 72 | 0.314 | 0.526 | 0.411 | 16.396 |
| 73 | 0.310 | 0.538 | 0.406 | 16.304 |
| 74 | 0.307 | 0.545 | 0.403 | 16.262 |
| 75 | 0.305 | 0.549 | 0.401 | 16.240 |
| 76 | 0.305 | 0.550 | 0.401 | 16.238 |
| 77 | 0.306 | 0.548 | 0.402 | 16.245 |
| 78 | 0.305 | 0.550 | 0.401 | 16.227 |
| 79 | 0.305 | 0.549 | 0.402 | 16.233 |
| 80 | 0.306 | 0.546 | 0.403 | 16.253 |
| 81 | 0.306 | 0.545 | 0.403 | 16.248 |
| 82 | 0.306 | 0.546 | 0.403 | 16.245 |
| 83 | 0.305 | 0.547 | 0.402 | 16.217 |
| 84 | 0.303 | 0.552 | 0.400 | 16.188 |

| | | | | |
|---|---|---|---|---|
| 85 | 0.301 | 0.557 | 0.398 | 16.171 |
| 86 | 0.339 | 0.454 | 0.442 | 17.002 |
| 87 | 0.339 | 0.454 | 0.442 | 16.995 |
| 88 | 0.338 | 0.461 | 0.439 | 16.861 |
| 89 | 0.338 | 0.457 | 0.441 | 16.695 |
| 90 | 0.338 | 0.456 | 0.441 | 16.955 |

### A.2.2 Error Metrics for Second Loop Enhanced Model

The error metrics in LAG 85, which is the best MAPE value among 90 models of the second enhanced model obtained from the regression analysis, are as follows:

| DRP_LAG | MAE | $R^2$ | RMSE | MAPE |
|---|---|---|---|---|
| 1 | 0.340 | 0.452 | 0.442 | 17.06 |
| 2 | 0.340 | 0.452 | 0.443 | 17.057 |
| 3 | 0.340 | 0.452 | 0.443 | 17.053 |
| 4 | 0.340 | 0.453 | 0.442 | 17.08 |
| 5 | 0.340 | 0.455 | 0.441 | 17.10 |
| 6 | 0.340 | 0.455 | 0.441 | 17.096 |
| 7 | 0.340 | 0.455 | 0.441 | 17.10 |
| 8 | 0.340 | 0.455 | 0.441 | 17.095 |
| 9 | 0.340 | 0.455 | 0.441 | 17.104 |
| 10 | 0.340 | 0.455 | 0.441 | 17.104 |
| 11 | 0.340 | 0.454 | 0.441 | 17.108 |
| 12 | 0.340 | 0.454 | 0.442 | 17.11 |
| 13 | 0.340 | 0.454 | 0.442 | 17.109 |
| 14 | 0.340 | 0.457 | 0.441 | 17.101 |
| 15 | 0.342 | 0.450 | 0.443 | 17.170 |
| 16 | 0.341 | 0.452 | 0.442 | 17.170 |

| | | | | |
|---|---|---|---|---|
| 17 | 0.341 | 0.454 | 0.442 | 17.17 |
| 18 | 0.341 | 0.454 | 0.442 | 17.164 |
| 19 | 0.341 | 0.455 | 0.441 | 17.159 |
| 20 | 0.341 | 0.453 | 0.442 | 17.18 |
| 21 | 0.341 | 0.453 | 0.442 | 17.175 |
| 22 | 0.339 | 0.459 | 0.440 | 17.110 |
| 23 | 0.339 | 0.459 | 0.440 | 17.113 |
| 24 | 0.340 | 0.456 | 0.441 | 17.15 |
| 25 | 0.341 | 0.455 | 0.441 | 17.16 |
| 26 | 0.345 | 0.438 | 0.448 | 17.248 |
| 27 | 0.349 | 0.425 | 0.453 | 17.35 |
| 28 | 0.372 | 0.339 | 0.486 | 18.36 |
| 29 | 0.372 | 0.343 | 0.485 | 18.276 |
| 30 | 0.372 | 0.344 | 0.484 | 18.342 |
| 31 | 0.358 | 0.389 | 0.467 | 17.83 |
| 32 | 0.374 | 0.368 | 0.475 | 18.520 |
| 33 | 0.375 | 0.367 | 0.476 | 18.550 |

| | | | | |
|---|---|---|---|---|
| 34 | 0.375 | 0.366 | 0.476 | 18.566 |
| 35 | 0.379 | 0.359 | 0.479 | 18.735 |
| 36 | 0.381 | 0.353 | 0.481 | 18.83 |
| 37 | 0.381 | 0.353 | 0.481 | 18.831 |
| 38 | 0.382 | 0.351 | 0.482 | 18.855 |
| 39 | 0.381 | 0.355 | 0.480 | 18.82 |
| 40 | 0.378 | 0.362 | 0.477 | 18.704 |
| 41 | 0.376 | 0.365 | 0.476 | 18.608 |
| 42 | 0.375 | 0.369 | 0.475 | 18.55 |
| 43 | 0.375 | 0.368 | 0.475 | 18.558 |
| 44 | 0.375 | 0.369 | 0.475 | 18.55 |
| 45 | 0.378 | 0.366 | 0.476 | 18.682 |
| 46 | 0.377 | 0.369 | 0.475 | 18.65 |
| 47 | 0.376 | 0.373 | 0.473 | 18.594 |
| 48 | 0.377 | 0.367 | 0.476 | 18.67 |
| 49 | 0.378 | 0.361 | 0.478 | 18.735 |
| 50 | 0.375 | 0.371 | 0.474 | 18.620 |

| | | | | |
|---|---|---|---|---|
| 51 | 0.377 | 0.366 | 0.476 | 18.678 |
| 52 | 0.375 | 0.372 | 0.474 | 18.631 |
| 53 | 0.371 | 0.386 | 0.469 | 18.469 |
| 54 | 0.364 | 0.402 | 0.462 | 18.216 |
| 55 | 0.365 | 0.400 | 0.463 | 18.248 |
| 56 | 0.361 | 0.413 | 0.458 | 18.086 |
| 57 | 0.361 | 0.414 | 0.458 | 18.082 |
| 58 | 0.366 | 0.390 | 0.467 | 18.309 |
| 59 | 0.371 | 0.374 | 0.473 | 18.425 |
| 60 | 0.374 | 0.365 | 0.476 | 18.536 |
| 61 | 0.373 | 0.367 | 0.476 | 18.520 |
| 62 | 0.372 | 0.369 | 0.475 | 18.488 |
| 63 | 0.373 | 0.365 | 0.476 | 18.548 |
| 64 | 0.369 | 0.369 | 0.475 | 18.38 |
| 65 | 0.369 | 0.370 | 0.475 | 18.385 |
| 66 | 0.378 | 0.350 | 0.482 | 18.774 |
| 67 | 0.379 | 0.347 | 0.483 | 18.790 |

| | | | | |
|---|---|---|---|---|
| 68 | 0.378 | 0.350 | 0.482 | 18.75 |
| 69 | 0.378 | 0.351 | 0.482 | 18.780 |
| 70 | 0.378 | 0.351 | 0.481 | 18.763 |
| 71 | 0.379 | 0.349 | 0.482 | 18.81 |
| 72 | 0.379 | 0.349 | 0.482 | 18.796 |
| 73 | 0.371 | 0.373 | 0.473 | 18.484 |
| 74 | 0.368 | 0.386 | 0.468 | 18.391 |
| 75 | 0.364 | 0.400 | 0.463 | 18.211 |
| 76 | 0.363 | 0.403 | 0.462 | 18.186 |
| 77 | 0.365 | 0.397 | 0.464 | 18.269 |
| 78 | 0.365 | 0.396 | 0.465 | 18.277 |
| 79 | 0.366 | 0.394 | 0.465 | 18.296 |
| 80 | 0.366 | 0.394 | 0.466 | 18.297 |
| 81 | 0.368 | 0.390 | 0.467 | 18.361 |
| 82 | 0.368 | 0.390 | 0.467 | 18.360 |
| 83 | 0.368 | 0.388 | 0.468 | 18.39 |
| 84 | 0.366 | 0.395 | 0.465 | 18.326 |

| | | | | |
|---|---|---|---|---|
| 85 | 0.360 | 0.412 | 0.458 | 18.073 |
| 86 | 0.360 | 0.411 | 0.459 | 18.077 |
| 87 | 0.361 | 0.410 | 0.459 | 18.099 |
| 88 | 0.359 | 0.415 | 0.457 | 17.975 |
| 89 | 0.359 | 0.417 | 0.457 | 17.950 |
| 90 | 0.358 | 0.417 | 0.456 | 17.793 |

Performance Metrics for Different DRP_LAG Values

# REFERENCES


[1] I. El Azab, R. Swief, N. El-Amary, and H. Temraz, "Machine and deep learning approaches for forecasting electricity price and energy load assessment on real datasets," *Ain Shams Engineering Journal*, vol. 15, p. 102613, 04 2024.

[2] L. M. Abadie, "Energy market prices in times of covid-19: The case of electricity and natural gas in spain," *Energies*, 2021.

[3] A. N. Alkawaz, A. Abdellatif, J. Kanesan, A. S. M. Khairuddin, and H. M. Gheni, "Day-ahead electricity price forecasting based on hybrid regression model," *IEEE Access*, vol. 10, pp. 108021–108033, 2022.

[4] A. Bello, J. Reneses, and A. Muñoz, "Medium-term probabilistic forecasting of extremely low prices in electricity markets: Application to the spanish case," *Energies*, vol. 9, pp. 1–27, 2016.

[5] N. Bibi, I. Shah, A. Alsubie, S. Ali, and S. A. Lone, "Electricity spot prices forecasting based on ensemble learning," *IEEE Access*, vol. 9, pp. 150984–150992, 2021.

[6] D. Bissing, M. Klein, R. A. Chinnathambi, D. Selvaraj, and P. Ranganathan, "A hybrid regression model for day-ahead energy price forecasting," *IEEE Access*, vol. 7, pp. 36833–36842, 2019.

[7] C. Castillo-Berrio and  Bozlak, "Analyzing the impact of natural gas prices on day-ahead electricity price forecasting in turkey: A comparative study of lstm and cnn- lstm model," 12 2023.

[8] M. Çetinkaya and T. Acarman, "Next-day electricity demand forecasting using regression," in *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, pp. 1549–1554, 2021.

[9] C. tian Cheng, B. Luo, S.-M. Miao, and X. Wu, "Mid-term electricity market clearing price forecasting with sparse data: A case in newly-reformed yunnan electricity market," *Energies*, vol. 9, p. 804, 2016.

[10] F. Jan, I. Shah, and S. Ali, "Short-term electricity prices forecasting using functional time series analysis," *Energies*, 2022.

[11] D. Keles, J. Scelle, F. Paraschiv, and W. Fichtner, "Extended forecast methods for day-ahead electricity spot prices applying artificial neural networks," *Applied Energy*, vol. 162, pp. 218–230, 2016.

[12] W. Li and D. Becker, "Day-ahead electricity price prediction applying hybrid models of lstm-based deep learning methods and feature selection algorithms under consideration of market coupling," *Energy*, vol. 237, p. 121543, 2021.

[13] S. Orenc, E. Acar, and M. S. Özerdem, "The electricity price prediction of victoria city based on various regression algorithms," in *2022 Global Energy Conference (GEC)*, pp. 164–167, 2022.

[14] B. E. Ozmen and G. Poyrazoglu, "The time-lag effect of natural gas prices on electricity prices in turkiye: An analysis of the post-pandemic market," *2023 19th International Conference on the European Energy Market (EEM)*, pp. 7–5, 2023.

[15] O. G. Poyrazoglu and G. Poyrazoglu, "Impact of natural gas price on electricity price forecasting in turkish day-ahead market," in *2019 1st Global Power, Energy and Communication Conference (GPECOM)*, pp. 435–439, 2019.

[16] G. Poyrazoglu, "Impact of gas price on electricity price forecasting via supervised learning and random walk," in *2019 16th International Conference on the European Energy Market (EEM)*, pp. 1–5, 2019.

[17] G. Poyrazoglu, "Price estimation by random walk mean reversion in day-ahead electricity market," *2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, pp. 1–5, 2019.

[18] A. Prajesh, P. Jain, S. Dhingra, Rishabh, A. Malik, and S. Agrawal, "Short term electricity price forecasting by optimized lstm model of deep learning with genetic algorithm," in *2023 7th International Conference on Computer Applications in Electrical Engineering-Recent Advances (CERA)*, pp. 1–6, 2023.

[19] B. Taş, "Electricity price forecasting using hybrid time series models," Master's thesis, Middle East Technical University, 2018.

[20] S. Tehrani, J. Juan, and E. Caro, "Electricity spot price modeling and forecasting in european markets," *Energies*, 2022.

[21] T. Ulgen and G. Poyrazoglu, "Predictor analysis for electricity price forecasting by multiple linear regression," *2020 International Symposium on Power Electronics, Electrical Drives, Automation and Motion (SPEEDAM)*, pp. 618–622, 2020.

[22] T. Ulgen, A. E. Sayed, and G. Poyrazoglu, "Prediction algorithm learner selection for european day-ahead electricity prices," *2020 2nd Global Power, Energy and Communication Conference (GPECOM)*, pp. 285–286, 2020.

[23] S. Yıldırım, M. Khalafi, T. Güzel, H. Satık, and M. Yılmaz, "Supply curves in electricity markets: A framework for dynamic modeling and monte carlo forecasting," *IEEE Transactions on Power Systems*, vol. 38, no. 4, pp. 3056–3069, 2023.

[24] C. Yıldız, "A two stage model for day-ahead electricity price forecasting: Integrating empirical mode decomposition and catboost algorithm," *Konya Journal of Engineering Sciences*, 2023.

[25] S. Zhu, M. Zou, S. Z. Djokic, J. Gunda, and I. Papic, "Comparison of single-stage and two-stage lstm-based methods for forecasting electricity demands and prices," in *2022 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, pp. 1–5, 2022.

[26] M. H. Fouladfar, A. Loni, M. Bagheri Tookanlou, M. Marzband, R. Godina, A. Al-Sumaiti, and E. Pouresmaeil, "The impact of demand response programs on reducing the emissions and cost of a neighborhood home microgrid," *Applied Sciences*, vol. 9, no. 10, 2019.

# VITA

I graduated with a bachelor's degree in electrical electronics engineering from TOBB Economy and Technology University in 2016. Afterward, I worked as an R&D engineer at Boğaziçi Elektrik Dağıtım Şirketi until 2022, leading innovative projects like the "Harvest Phase 2 Project" and the "Feeder Measurement, Monitoring, Analysis, and Protection System Design Project." I then transitioned to Özel Şifa Hastanesi, where I currently work as a data science specialist, utilizing my expertise in Python and machine learning to streamline healthcare data processes. My skills encompass Python, SQL, data analysis, machine learning algorithms, and project management, having successfully led projects in the healthcare sector.