ANALYSIS OF THE IMPACT OF CEILING PRICES

ON TURKISH DAY-AHEAD ELECTRICITY MARKET CLEARING PRICE

MUSTAFA TALHA SAĞMAN

BOĞAZİÇİ UNİVERSITY

2024

ANALYSIS OF THE IMPACT OF CEILING PRICES

ON TURKISH DAY-AHEAD ELECTRICITY MARKET CLEARING PRICE

Thesis submitted to the

Institute for Graduate Studies in Social Sciences

in partial fulfillment of the requirements for the degree of

Master of Arts

in

Management

by

Mustafa Talha Sağman

Boğaziçi University

2024

Analysis of the Impact of Ceiling Prices

on Turkish Day-Ahead Electricity Market Clearing Price

The thesis of Mustafa Talha Sağman

has been approved by:

Assoc. Prof. Ali Coşkun (Thesis Advisor)	_____
Halil Ertürk Esen, Phd (Thesis Co-Advisor)	_____
Prof. Vedat Akgiray	_____
Assoc. Prof. Ulaş Akküçük	_____
Assoc. Prof. Ahmet Göncü (External Member)	_____

January 2024

DECLARATION OF ORIGINALITY

I, Mustafa Talha Sağman, certify that

- I am the sole author of this thesis and that I have fully acknowledged and documented in my thesis all sources of ideas and words, including digital resources, which have been produced or published by another person or institution;
- this thesis contains no material that has been submitted or accepted for a degree or diploma in any other educational institution;
- this is a true copy of the thesis approved by my advisor and thesis committee at Boğaziçi University, including final revisions required by them.

Signature…….....

Date …….....

ABSTRACT

Analysis of The Impact of Ceiling Prices

on Turkish Day-Ahead Electricity Market Clearing Price

After the energy crisis in the post-pandemic period, the ceiling price has become more effective in the Turkish day-ahead electricity market. The limiting impact of ceiling prices is evident for hours when we see ceiling prices touched in the market. On the other hand, the effect of ceiling prices on hours when ceiling prices are not touched is essential to consider. Another critical point of consideration is the impact of ceiling prices on the other determinants of market clearing price. In this study, a comprehensive machine learning model comprising data sources of production, consumption, supply-demand curve points, weather conditions, and many other features are constructed, and ceiling price information of the trading day is integrated into this model. The model's behavior, which focuses on ceiling price and other features, is analyzed using the SHAP package in Python. With this analysis, the impact on hours where the ceiling price is not touched has been analyzed. The effect of ceiling prices on the determinants of ceiling prices is analyzed by building a statistical model that considers the ceiling price as a censor to the linear regression model, namely the Tobit model. A classical linear regression model with the same features as the Tobit model as a control model is constructed. The impact of ceiling prices on all hours of the day and the determinants of the market clearing price were found to be evident as the result of the analysis of this thesis.

ÖZET

Tavan Fiyat Uygulamasının Türkiye Gün Öncesi Elektrik Piyasa Takas Fiyatı Üzerindeki Etkisinin Analizi

Pandemi sonrası enerji krizinin ardından, tavan fiyat uygulaması Türk gün öncesi elektrik piyasasında etkin hale gelmiştir. Piyasada tavan fiyatların görüldüğü saatlerde tavan fiyatlarının kısıtlayıcı etkisi açıktır. Diğer yandan, tavan fiyatların görülmediği saatlerdeki etkisi de değerlendirilmesi gereken önemli bir konudur. Dikkate alınması gereken bir diğer önemli nokta ise, tavan fiyatlarının piyasa takas fiyatını belirleyen faktörler üzerindeki etkisidir. Bu çalışmada, üretim, tüketim, arz-talep eğrisi noktaları, hava durumu ve birçok diğer özellik veri kaynaklarından beslenen kapsamlı bir makine öğrenimi modeli oluşturulmuş ve bu modelin içine işlem gününün tavan fiyat bilgileri dahil edilmiştir. Bu modelin davranışı, tavan fiyat ve diğer özellikler arasındaki ilişkiler Python'daki SHAP paketi kullanılarak analiz edilmiştir. Bu analizle, tavan fiyatının görülmediği saatlerdeki etkisi değerlendirilmiş ve tavan fiyatın bu saatlerde de etkili olduğu görülmüştür. Tavan fiyatın piyasa takas fiyatının belirleyicileri üzerindeki etkisi araştırılmak üzere, tavan fiyatın sansür olarak değerlendirildiği bir lineer regresyon modeli olan Tobit modeli kurulmuştur. Kontrol modeli olarak ise Tobit modeli ile aynı özelliklere sahip klasik bir lineer regresyon modeli oluşturulmuştur. Oluşturulan bu iki modelin davranışları incelenmiş ve bu incelemeler sonucunda tavan fiyat uygulamasının tavan fiyatı belirleyen faktörler üzerinde de etkili olduğu sonucuna varılmıştır.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my thesis advisor, Assistant Professor Ali Coşkun, and my co-advisor, Halil Ertürk Esen, for their support and valuable contributions throughout my research journey. Their guidance, expertise and continuous feedback have been instrumental in refining the quality and direction of this study.

Furthermore, I would like to express my heartfelt gratitude to my supportive wife, Merve Sağman, whose encouragement and understanding have been my pillar of strength throughout this academic endeavor. Her patience and belief in my abilities have been a constant source of motivation. Also my deepest gratitude to my dear son, my best playmate and baby, Yusuf Zahid Sağman, who makes every day brighter.

Finally, I would like to express my deepest gratitude to my beloved mother and father, whose support and sacrifices have made my educational pursuits possible. Their help and belief in my potential has allowed me to focus my whole heart on my studies. I would like to express my gratitude to my beloved father, Naci Sağman, who recently passed away. He was not only a man of faith, but also my mentor who guided me with his support and efforts and showed me the path I should follow in life. I would also like to extend my heartfelt appreciation to my dear brother and sister, who have always be with me throughout all the moments that have shaped my journey.

I would like to thank my beloved family for their invaluable support and love. Their patience and endless guidance make all the challenges achievable throughout my life.

With deepest appreciation,

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION 1

CHAPTER 2: LITERATURE REVIEW 12

2.1 Ceiling prices 12

2.2 Ceiling prices in electricity markets 14

2.3 Electricity markets 15

2.4 Electricity price prediction 20

2.5 Impact of ceiling prices on price prediction 29

CHAPTER 3: METHODOLOGY 30

3.1 Data 33

CHAPTER 4: RESULTS 75

4.1 Preprocessing 75

4.2 Train-test split 75

4.3 Correlation check 76

4.4 Selected model and hyperparameters 77

4.5 Prediction on test period 79

4.6 Feature importance plots 80

4.7 SHAP analysis 83

4.8 Tobit model built and analyze on determinants of market clearing price 89

4.9 Sensitivity analysis on MCP predictions 96

CHAPTER 5: CONCLUSION 99

CHAPTER 6: LIMITATIONS AND FUTURE WORK 102

REFERENCES 104

LIST OF TABLES

Table 1. Top 10 Cities Having Highest Proportion of Energy Consumption for May 2023 56

Table 2. Top 10 Cities of Wind Production Capacities 58

Table 3. Top 10 Cities with Highest Unlicenced Solar Power Generation 60

Table 4. Tobit Regression Model Results 92

Table 5. Standard Linear Regression Model Results 93

LIST OF FIGURES

Figure 1. Trade volumes across electricity markets 3
Figure 2. Daily ceiling prices and daily market clearing price mean, max, min 4
Figure 3. Histograms of market clearing prices for different ceiling prices 7
Figure 4. Histogram of market clearing prices after 2022 ceiling prices are highlighted 8
Figure 5. Proportion of ceiling price effective hours by periods of ceiling prices..... 9
Figure 6. Daily consumption per hour and daily mean market clearing price 34
Figure 7. Daily consumption per hour and daily mean market clearing price 35
Figure 8. Hourly consumption and hourly market clearing price for a specific date 36
Figure 9. Daily production per hour and daily mean market clearing price 38
Figure 10. Daily mean production per hour and daily mean market clearing price.. 39
Figure 11. Hourly production amounts and market clearing price for a specific date 39
Figure 12. Methodology used to calculate YEKDEM-supported facilities 42
Figure 13. Daily mean YEKDEM supported facilities production per hour and daily mean MCP..... 43
Figure 14. Daily mean wind sourced production per hour excluding YEKDEM and daily mean MCP..... 44
Figure 15. Daily mean dammed hydropower production per hour excluding YEKDEM supported and daily mean MCP 45
Figure 16. Daily mean natural gas production amount per hour and daily mean market clearing price..... 47

Figure 17. Daily mean imported coal sourced production amount and mean market clearing price. 49
Figure 18. Daily mean domestic coal-based production per hour and daily mean market clearing price 50
Figure 19. Daily USD/TRY exchange rate and daily mean market clearing price ... 52
Figure 20. Daily mean gas reference price and mean market clearing price 53
Figure 21. Is working day dummy feature and daily mean market clearing price.... 54
Figure 22. Consumption weighted temperature and daily mean market clearing price 56
Figure 23. Daily absolute difference of consumption weighted temperature from the mean and daily market clearing price 57
Figure 24. Wind production weighted wind speed and daily mean market clearing price 59
Figure 25. Solar sourced production weighted sun hour and daily mean market clearing price 61
Figure 26. Daily mean price independent sale offer amount and daily mean market clearing price 62
Figure 27. Daily mean price independent bid order amount per hour and daily mean market clearing price 63
Figure 28. Supply demand curve of 17 July 2023 14.00 64
Figure 29 Minimum and maximum demand amounts & market clearing price for a specific date 65
Figure 30 Minimum and maximum supply amounts & market clearing price for a specific date 66

Figure 31 Daily mean matching amounts per hour and daily mean market clearing price 67
Figure 32. Merit order representation 68
Figure 33 Marginal costs of production types 71
Figure 34. Daily ceiling prices and daily mean market clearing price 73
Figure 35. Number of hours ceiling price is touched 74
Figure 36. Correlation matrix of the model 76
Figure 37. Expected and Predicted values with the model proposed 80
Figure 38. Split-based feature importance plot 81
Figure 39. Gain-based feature importance plot 82
Figure 40. SHAP Summary Plot 84
Figure 41. Shap dependency plot for ceiling price and the target variable 86
Figure 42. First example of SHAP force plot 87
Figure 43. Second example of SHAP force plot 88
Figure 44. Tobit model correlation matrix 90
Figure 45. Intersection points of Tobit and OLS model features 95
Figure 46. Histogram of predictions where ceiling price is 2700 TL/MWh 96
Figure 47. Histogram of predictions where ceiling price is 2900 TL/MWh 97
Figure 48. Histogram of predictions where ceiling price is 3200 TL/MWh 98

CHAPTER 1:

INTRODUCTION

Electricity is one of the vital components of today's modern world. One of the most essential characteristics of electricity is that it cannot be stored. Because of this feature, it is necessary to balance supply and demand constantly to sustain electricity supply. This is making electricity prices unique in itself. Electricity distribution companies are responsible for the electricity supply when it is demanded by anyone who opens any lights. Although it is produced by electricity generation companies from several resources like renewables, coal, and natural gas, it is the distribution companies' responsibility to meet the demand with trade operations with generation and trade companies. The distribution companies with trade operations must source electricity before consumption. Moreover, this trade can happen in several ways: It can be sourced with bilateral agreements between buyer and seller. These types of contracts are unique to the parties of the agreement; the price of the electricity, the time that electricity will be delivered, the load that will be provided, and other types of characteristics are determined by buyers and sellers. Generally, the period is broad, like weekly, monthly, etc. Over-the-counter markets have counterparty risk; even if companies agree on a contract, they may fail to fulfill the needs of the contract. There is a lack of a price discovery function that would help us gain insight into the long-term price of electricity because agreements in over-the-counter markets are not published regularly. To overcome these difficulties, on 01 June 2021, the Power Futures Market in Exist is served in EXIST. This market introduces a central counterparty feature, and long-term physical delivery is assured. With tools in this market, participants can hedge their risks and see price prospects for the future.

Due to the close maturity of transactions, markets where the physical delivery is closer than one day are called spot markets. When we come closer to the consumption time, like a day before the consumption, it can be traded in the day-ahead electricity market. Which is a regulated market conducted by EXIST. In this market, market participants give their bid and sell offers for 24 hours of the next day, and as a result of all of the offers, a market clearing price (MCP) is created for every hour of the next day. Supply and demand curves for each hour are created, and the final price is determined in this market. Approximately %40 of the load for a day is getting traded in this market.

If the electricity needed is within the day, it can be traded in intra-day market. EXIST also operates the intra-day market. Till 1 hour in advance of physical delivery, participants can trade to eliminate their imbalances and balance their portfolios. When time gets closer to physical delivery, if market participants have imbalances, they can hover in the Power Balancing Market, which is governed by Türkiye Elektrik İletim A.Ş (Türkiye Electricity Transmission Company TEİAŞ). Real-time operation and management are carried out in this market. It is mandatory to participate in this market as a balancing unit. However, balancing in this market is much costlier than in other markets; because of this, participants try not to be obliged to suspend in this market. The distribution of daily traded electricity by energy markets can be found in Figure 1.

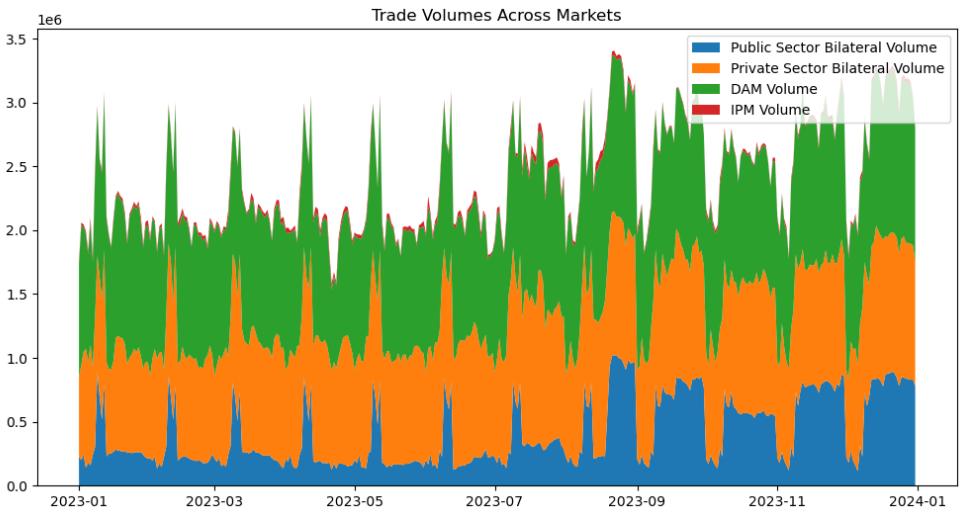

Figure 1. Trade volumes across electricity markets

It is seen that the day ahead electricity market most of the time has the highest proportion compared to the other markets. Because of this, the MCP created in this market is also crucial for the energy sector. MCP acts like a reference price for other markets, meaning that the prices formed in the day-ahead electricity market represent the supply and demand balance of the market very close to the physical electricity supply. (Boru İpek, 2021). Market participants aim to optimize their portfolios by trading in these markets; because MCP is a benchmark price, predicting it before it is announced for trading opportunities is essential. If they can predict MCP more precisely, they can give the most beneficial offers to the market and, therefore, balance their portfolios much more effectively. Participants can also benefit from arbitrage opportunities between these markets.

Ceiling prices in electricity markets refer to the maximum price that can be charged for electricity set by a regulatory body. They are typically applied to protect consumers from excessive prices, often when there is a monopoly or limited competition in the electricity market. This ensures that electricity remains affordable for all consumers while allowing electricity providers to reasonably cover their costs

and profit. Ceiling prices are making the electricity market clearing price data censored with values that could not exceed some amounts. There is a noticeable impact on hourly prices, where it reaches the ceiling price and cannot jump above it, but there is a possible impact on the prices of all hours.

Graph of the day's minimum, maximum, and mean MCP and the ceiling price determined by the regulatory authority for the specific date can be found in Figure 2. It can be seen that the maximum amount of MCP could be, at most, the ceiling price. Also, the daily mean market clearing price can sometimes reach the ceiling price determined. For the days where it reaches the ceiling price, it means that all 24 hours of the day are recorded as the ceiling price. On the other hand, we are seeing that the minimum MCP can reach 0 TL but cannot go below 0 TL because negative pricing is not possible in the Turkish day-ahead electricity market.

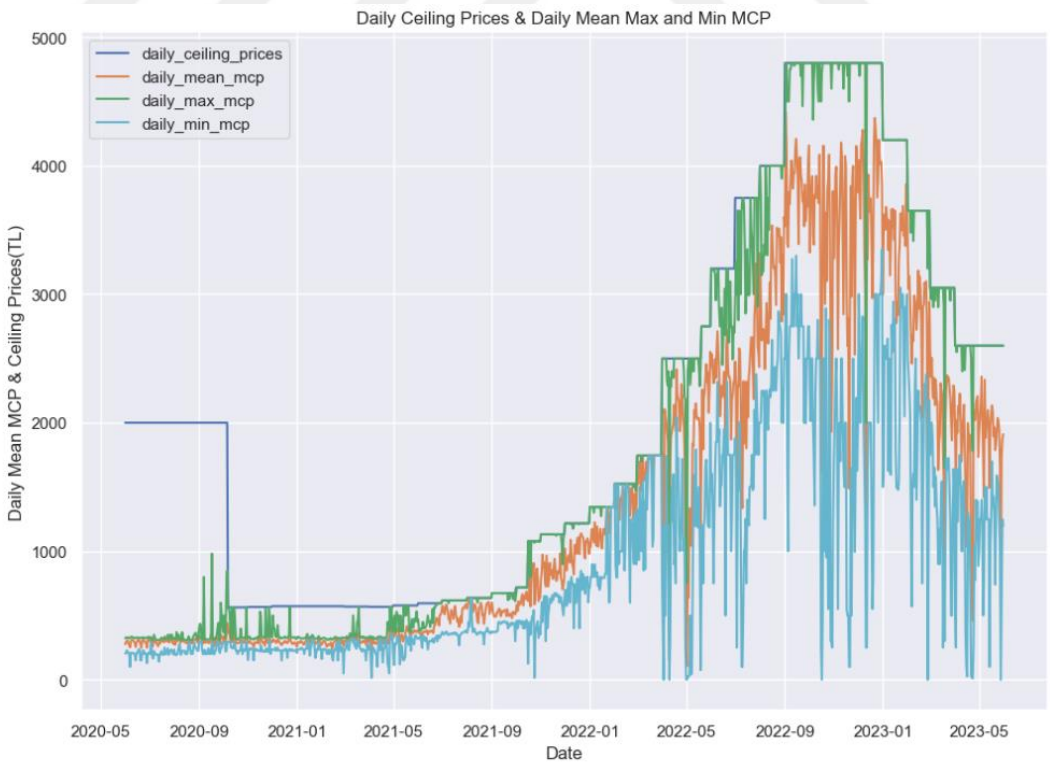

Figure 2. Daily ceiling prices and daily market clearing price mean, max, min

In this data, ceiling prices are effective after 2021-06, the post-pandemic period in which we started seeing spikes in energy prices. The crisis's roots lay in a complex interplay of economic elements. Key among them was the swift resurgence of the global economy after the pandemic, a rebound that surpassed the available energy supplies. This situation further deteriorated into an acute energy shortfall on a worldwide scale, particularly following the military actions of Russia in Ukraine. Consequently, the price of natural gas soared to unprecedented levels, which, in turn, drove up electricity prices in several markets. We started seeing ceiling prices touched in this crisis period, and we are seeing this behavior today, too. When considering the period before the energy crisis, the ceiling prices were ineffective in the market. The number of hours we see ceiling price equals market clearing prices, which are rare in this period. Because of this, the impact of ceiling prices on electricity prices was relatively mild. However, after the crisis, it is very strong.

When we get the histogram of Market Clearing Prices for specific dates where the ceiling price is constant, it can be seen that MCPs are limited by ceiling prices in the distribution.

In the below histograms, the MCP histogram is plotted for ceiling price values that are 2500 (ceiling price from 01/04/2022 to 20/05/2022), 2600 (ceiling price from 01/04/2023 to 04/07/2023), 2700 (ceiling price from 04/07/2023 to current day of this study), 3050 (ceiling price from 01/03/2023 to 01/04/2023), 3200 (ceiling price from 01/06/222 to 01/07/2022), 3750 (ceiling price from 01/07/2022 to 01/08/2022), 4000 (ceiling price from 01/08/2022 to 01/09/2022), 4200 (ceiling price from 01/01/2023 to 01/02/2023), and 4800 (ceiling price from 01/09/2022 to 01/01/2023) can be found. Each ceiling price period analyzed here will also be

tracked in the ceiling price stairs based on the previous comparison between MCP and ceiling prices.

In these histograms, it can be seen that the normal distribution of MCP is usually broken with ceiling prices. This inference is likely much stronger when we see more ceiling prices equal market clearing price cases.

For instance, in Figure 3, the histogram of ceiling price equals 3200, which is closer to the normal distribution. When we check the proportion of ceiling price equal to MCP hours to the total number of hours in this period, it is equal to %0.97, which is much lower than other periods in this sample. On the other hand, when we consider the ceiling price of 4200, it can be seen that the distraction of normal distribution is strongly interfered with by the ceiling price of 4200; when we check the same proportion for this period, it is recorded as %23.52 which equals to almost one-quarter of the hours in this period.

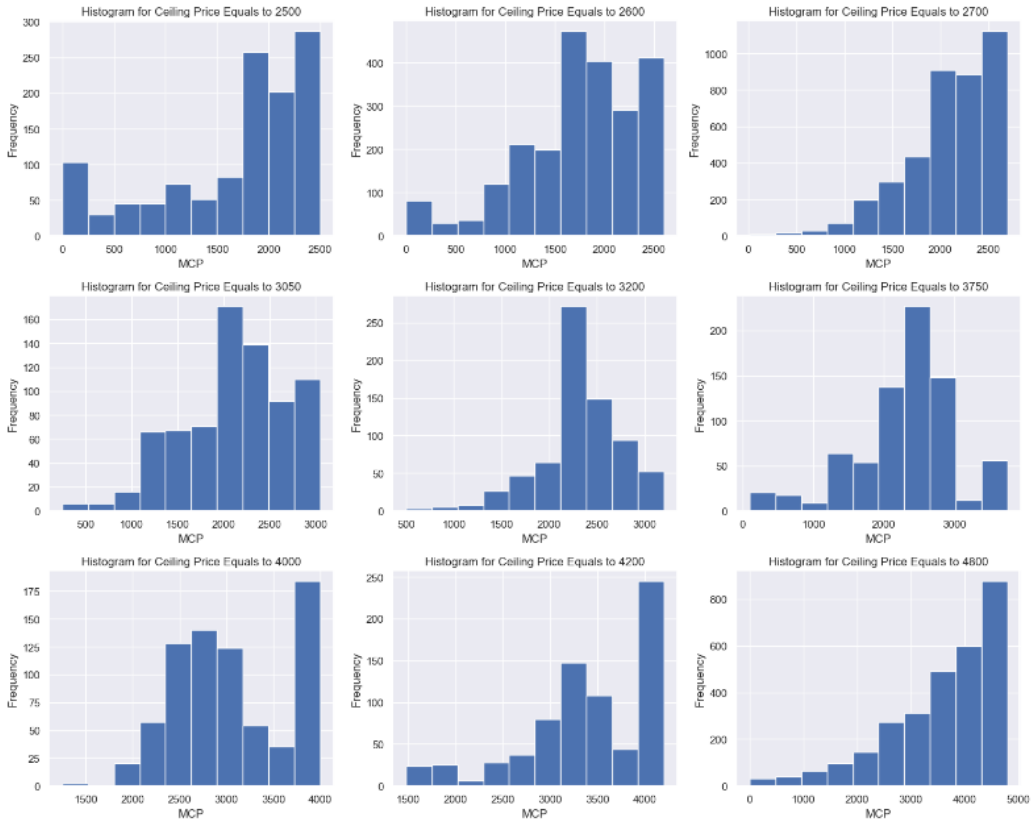

Figure 3. Histograms of market clearing prices for different ceiling prices

In Figure 4, historical hourly market clearing prices are plotted in the histogram to inspect the overall impact. In this histogram, it can be seen that some peaks influence the overall normal distribution of the data. All the ceiling prices set after 2022 are plotted with red-dashed vertical lines to understand the dynamics behind these peaks. It can be seen that those peaks are highly related to the ceiling prices determined in various periods.

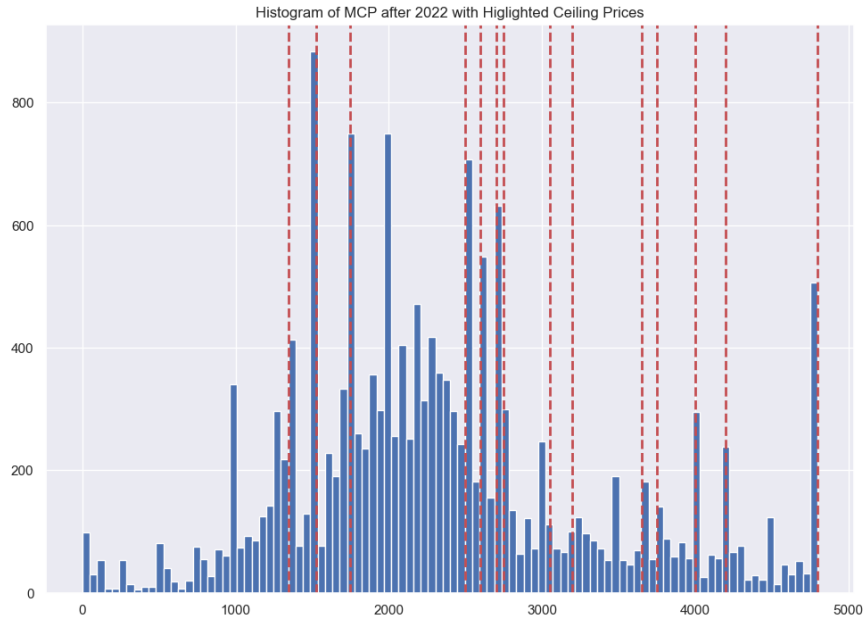

Figure 4. Histogram of market clearing prices after 2022 ceiling prices are highlighted

To further analyze the impact of ceiling prices, the proportion of hours we have ceiling prices equal to MCP to the total number of ceiling prices is plotted in Figure 5.

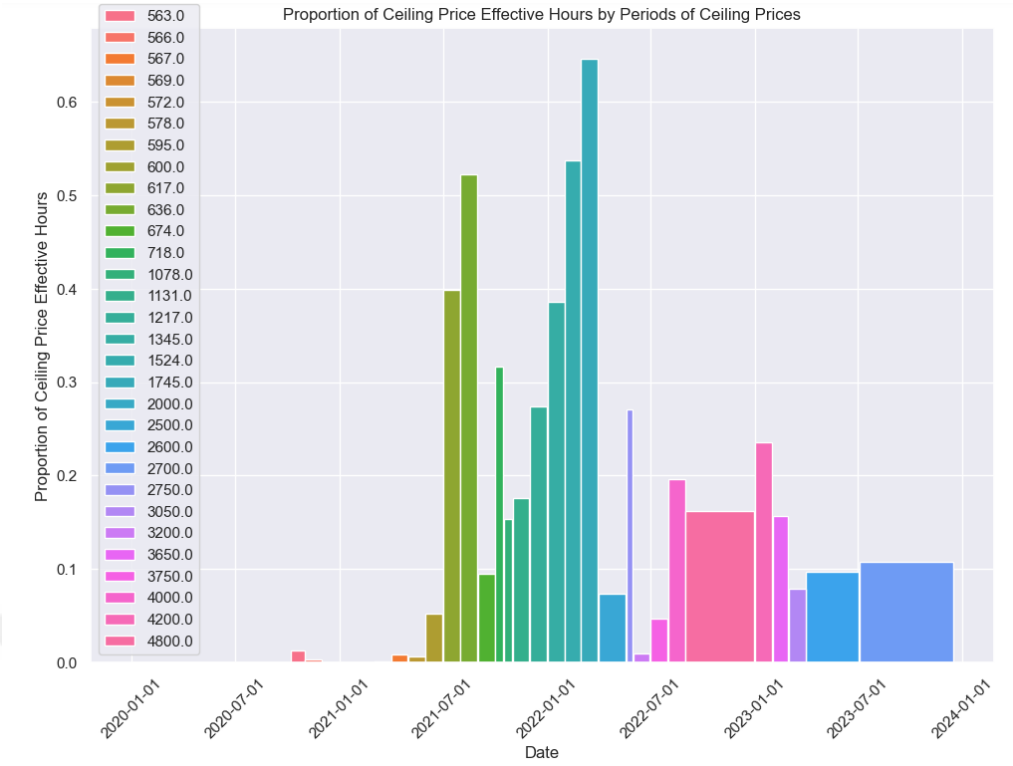

Figure 5. Proportion of ceiling price effective hours by periods of ceiling prices

The height of this plot shows the proportion of hours where the ceiling price is seen. For instance, in 2022-03-01, the proportion of ceiling price seen hours is increased to %65 percent. That means most of the hours' prices in this period are equal to the ceiling price of the period. In the legend, the ceiling price of the period can be tracked, which is 1745 TL. The length of the x-axis of this plot shows the timespan for which the specific ceiling price is set.

Ceiling prices impact the MCP by just analyzing these histograms and data provided. On the other hand, the ceiling price can impact hours where MCP does not touch the ceiling price. The ceiling price mechanism is making the MCP data a censored one. Censored data refers to a subset of data modified or restricted for various reasons, such as privacy concerns, governmental regulations, or organizational policies. This type of data manipulation can significantly affect the

integrity and usefulness of the data set. When specific data points are censored, it can distort the overall analysis and interpretation. This is because the exclusion or alteration of particular information results in a lack of representativeness and completeness in the data set. For instance, in statistical analysis, censored data can lead to biased estimates and incorrect conclusions if not appropriately handled. The impact of data censoring is pervasive across all data points, as the absence or modification of some data alters the underlying relationships and distributions within the data set. Therefore, it is crucial to recognize and account for censored data in any analytical process to ensure the accuracy and reliability of the findings.

This master's thesis aims to analyze further the impact of ceiling prices on electricity market clearing prices. For this purpose, the literature about electricity markets, market clearing price prediction models, and ceiling prices will be carefully reviewed in the next section.

Prediction models are practical tools to understand the relationships between variables and outputs. To understand the impact of ceiling prices on MCP, a prediction model that utilizes the ceiling price feature is constructed and tested for a specific period. This study's methodology section introduces the methods applied throughout the study, and the data used for the prediction models are analyzed.

The specifications of the prediction model and the outputs are analyzed in the results section. The impact of each feature on the market clearing price is analyzed in detail using the SHAP method in this section. Because ceiling prices can affect other determinants of market clearing prices, we need to employ some statistical techniques to analyze this impact. To cater to this purpose, we have selected the important features from the feature importance of the machine learning model and

built a Tobit model, the version of linear regression that considers the impact of censoring in predictions. This model creation aims to get the equation coefficients to determine the market clearing price. Another control linear regression model has been built to analyze the impact of ceiling price censor on the coefficients of determinant features. The only difference between Tobit and the linear regression model is that the linear regression model does not consider ceiling price censor. The coefficients and intercepts of both models are analyzed to understand the impact of ceiling prices on the other factors affecting the market clearing price.

CHAPTER 2:

LITERATURE REVIEW

2.1 Ceiling prices

Ceiling prices refer to a government-imposed limit on the price charged for a particular good or service. This regulatory mechanism is typically used to protect consumers from excessively high prices, especially during periods of scarcity or market instability. By setting a maximum allowable price, a ceiling aims to ensure affordability and prevent suppliers from exploiting high demand or limited supply situations.

However, while ceiling prices are intended to shield consumers, they can also have unintended consequences. For instance, artificially low prices may lead to shortages as suppliers find it unprofitable to produce or sell at the capped price. This can result in reduced product or service availability, potentially harming the consumers the policy seeks to protect. Additionally, price ceilings can discourage investment and innovation in the affected industry, as reduced profit margins make it less attractive for businesses to invest in improving or expanding their offerings.

Governments often resort to ceiling prices in response to crises like natural disasters, wars, or economic upheavals, where essential goods such as food, fuel, or housing might otherwise become unaffordable for a large population segment. However, careful consideration and planning are necessary to balance the intended benefits of price controls with their potential drawbacks.

In economic theory, a price ceiling is a regulation that sets the maximum price a seller is allowed to charge for a product or service (Mankiw, 2018).

Policymakers often justify this intervention in the free market as a means to protect consumers from exploitative pricing, especially in markets characterized by monopolies or limited competition. According to standard microeconomic models, while a price ceiling below the equilibrium price can make goods more affordable in the short term, it often leads to a market disequilibrium, where the quantity demanded exceeds the quantity supplied, resulting in shortages (Katz & Rosen, 1998). Economists argue that these shortages can exacerbate the very problems the price ceilings aim to solve, as they often lead to inefficient resource allocation and can incentivize black markets.

The historical application of price ceilings provides a rich context for understanding their impact. One prominent example is rent control, a form of price ceiling applied to housing rentals. Studies have shown that while rent control can provide short-term relief for existing tenants, it can also lead to negative long-term consequences, such as reduced incentives for landlords to maintain or improve properties, leading to a deterioration in the quality of housing stock (Glaeser & Luttmer, 2003). Furthermore, empirical evidence suggests that rent control policies can exacerbate housing shortages by discouraging new construction and reducing the availability of rental units, as landlords may convert properties to non-rental uses (Arnott, 1995)

Governments often implement price caps in essential markets such as utilities, healthcare, and pharmaceuticals to prevent monopolistic abuses and ensure that essential services remain accessible (Joskow, 2007). In these markets, price caps are used to balance profitability and public interest. The economic rationale is that in markets with high entry barriers and significant economies of scale, unchecked pricing can lead to monopolistic pricing, which is detrimental to consumer welfare

(Abbott, 1995). However, setting these caps requires careful consideration to ensure they are not counterproductive; they must be high enough to encourage firms to invest in maintaining and improving the quality of services but low enough to prevent overpricing.

The advent of dynamic pricing models in various sectors, facilitated by advances in technology and data analytics, has brought new challenges and perspectives to the concept of price capping. Dynamic pricing, which allows for real-time price adjustments based on demand, supply, and other factors, can lead to prices for services like ride-sharing or airline tickets fluctuate widely (Chen et al., 2015). In these contexts, economists debate the appropriateness and effectiveness of price caps. While caps can protect consumers from extreme price spikes during high-demand periods, they can also interfere with the efficiency of price signals in allocating resources and managing demand (Borenstein & Rose, 1994).

2.2 Ceiling prices in electricity markets

In the day-ahead electricity markets, applying ceiling prices is critical in mitigating price volatility and protecting consumers from potential price spikes due to fluctuating demand and supply conditions. Typically, electricity markets are characterized by inelastic demand and a diverse supply portfolio, ranging from baseload power plants to intermittent renewable energy sources. This unique market structure often leads to significant price fluctuations, especially during peak demand periods or when renewable sources are less productive. To address these challenges, market regulators often implement price ceilings to cap the maximum price that can be charged for electricity in the day-ahead market (Wilson, 2018). This intervention

aims to prevent extreme price peaks, which can arise due to factors such as limited generation capacity, transmission constraints, or sudden spikes in demand. While such price ceilings can provide short-term consumer protection and market stability, they also carry the risk of disincentivizing investment in new generation capacity or in technologies that enhance grid flexibility and reliability. As a result, careful calibration of these price caps is essential to balance consumer protection and the long-term sustainability and efficiency of the electricity market (Borenstein & Bushnell, 2015).

In Türkiye day ahead electricity prices are determined and announced by Energy Market Regulatory Authority (EMRA). It is periodically announced in the market announcements of the EXIST announcements page. The market dynamics are considered, and the authorities have tried to find the most efficient ceiling price for the period.

2.3 Electricity markets

Electricity markets represent a complex and dynamic sector, integral to modern economies, characterized by unique challenges and opportunities. These markets operate on the principle of balancing supply and demand in real-time, necessitating a robust infrastructure and advanced forecasting techniques. Unlike traditional commodities, electricity cannot be stored easily, making its market highly sensitive to fluctuations in supply and demand. This characteristic introduces significant volatility, requiring incorporating diverse energy sources, including renewables, to ensure stability and sustainability.

Renewable energy sources have gained high importance in recent years. It has a significant share of investments in the energy sector. However, its volatile production profile also introduces new risks for the industry. The production performance of renewables mainly depends on external factors such as weather conditions. Because of this, for markets which have highly reliant on production from renewables, it is necessary to include wind and solar infeed in models regarding the electricity market (Wagner, 2014)

Market participants range from large-scale producers, such as nuclear and hydroelectric power plants, to smaller renewable sources, like wind and solar farms, alongside consumers increasingly engaging in energy production through distributed generation. Regulatory frameworks play a crucial role in shaping these markets, aiming to foster competition while ensuring reliability and affordability of supply. Additionally, the growing emphasis on decarbonization and sustainable practices drives innovation in this sector, leading to the development of smart grids and the integration of energy storage technologies. As a result, electricity markets are not only platforms for energy trading but also crucibles for technological and regulatory advancements, reflecting broader shifts in global energy policies and consumer preferences.

Providing a continuous and reliable electrical system is one of the most important goals of countries because of electricity's indispensable nature in daily life; this is only possible with proper planning and management of the power supply system, which requires efficient modeling and forecasting capabilities. (Boru İpek, 2021).

The Turkish electricity market is distinguished by its strategic geographical position, bridging Europe and Asia, facilitating its energy corridor role. The market is characterized by significant renewable energy potential, notably in hydro, wind, solar, and geothermal resources, aligning with the government's incentives to reduce reliance on imported fossil fuels. However, Turkey remains heavily dependent on energy imports, particularly natural gas. Rapid demand growth due to economic and population expansion necessitates substantial investments in new infrastructure. The market is undergoing transformative liberalization and regulatory reforms to foster private investment and enhance competition. Concurrently, Turkey faces environmental and regulatory challenges, balancing energy demands with sustainability commitments. Integrating advanced technologies, such as smart grids and energy storage, is pivotal. Additionally, Turkey's efforts to augment regional and international energy connectivity underscore the uniqueness of its electricity market.

Türkiye's electricity market has a comprehensive and intricate framework accommodating a blend of conventional and renewable energy sources. This market is governed by regulatory bodies emphasizing sustainability, efficiency, and reliability. The Turkish electricity market was liberalized in 2001 to facilitate competition and encourage private-sector involvement. Türkiye has many entities that generate, transmit, distribute, and retail electricity.

As a result of the liberalization process, three primary entities are formed in the market: the state-owned transmission company (TEİAŞ), the state-owned distribution company (TEDAŞ), and the state-owned power generation company (EUAŞ). Many private sector companies also generate electricity, distribute it, and retail it.

The Market Operator oversees and facilitates the day-ahead electricity market, designated for trading and balancing electricity. It functions a day before the scheduled power delivery date, ensuring efficient transactions and stability in the electricity supply.

In the Day-Ahead Market, participants can submit various types of orders, including hourly, daily, and flexible, for specific periods. These orders comprise both volume and price data, with the latter exhibiting centesimal sensitivity. Prices for these orders are expressed in Turkish Lira (TL/MWh), and the volume is denoted in terms of 'Lot', where one Lot represents 0.1 MWh. This structure allows for detailed and precise trading, catering to the diverse needs of market participants while maintaining a high level of specificity and flexibility in the market.

Participants in the Day-Ahead Electricity Market can submit orders hourly and daily for specific time slots. These orders are characterized by their composition of volume and price information, which varies across different hours. This framework facilitates precise and flexible market operations, allowing participants to tailor their bids to the dynamic conditions of the market.

Single hourly orders are structured with a maximum of 64 steps, comprising an equal distribution of 32 purchases and 32 sales. The pricing structure for these orders is required to be in ascending order. Importantly, for a given price step, an order cannot simultaneously be valid for both a purchase and a sale. A linear interpolation method is applied to construct the supply-demand curve. This technique interpolates values between two consecutive steps of price and volume, aiding in creating a more accurate representation of market dynamics.

Profile block orders allow participants to submit varying amounts for each hour within a designated time frame. These orders must cover at least three consecutive hours and are restricted such that the volume for consecutive hours cannot exceed threefold. The entire profile block order covers a specific period and is subject to complete execution or rejection. Additionally, market participants are limited to submitting a maximum of 50 block orders for each trading day, ensuring a balanced and manageable order volume within the market framework.

Market participants have the option to submit up to six flexible orders for either the same or different time slots within a delivery day. These flexible orders have a minimum timeline of 8 hours and can extend up to 24 hours, though their validity is limited to a maximum of 4 hours within this span. Notably, flexible orders are not tied to any specific hour, offering participants greater adaptability in aligning their orders with fluctuating market conditions and individual trading strategies.

The market operator is responsible for determining the valid buy or sell quantities for each offer, known as the matching quantity, which must comply with specific conditions outlined in regulations. These quantities are calculated to increase daily market surplus, ensuring equal supply and demand for each hour, and must adhere to regulatory rules.

Offers are subject to constraints: hourly buy or sell offers depend on the Market Clearing Price (MCP) value. Block offers covering multiple periods might be acceptable in some periods based on the MCP but not others. A sell block offer is accepted if its price is lower or equal to the average MCP of covered periods; similarly, a buy block offer is accepted if its price is higher or equivalent. Flexible sell offers are accepted if their price is lower or equal to the highest announced MCP.

Paradoxically accepted offers, which deviate from these norms but are still accepted, can lead to discrepancies in settlement amounts.

The objective of the problem is to maximize the daily market surplus, which equals the sum of producers' and consumers' surpluses. A participant's producer surplus is the difference between the amount they receive for their matched sell quantity and the amount they offered to sell. The consumer surplus is the difference between the amount offered for a buy quantity and its actual cost.

2.4 Electricity price prediction

Harmonization and deregulation have been applied in electricity markets worldwide within the last two decades (Weron, 2006). Historically, the electricity market was vertically integrated with power utilities, and most state-owned firms have acted in this market. With trade liberalization, this picture has changed, and state-owned entities have started to get replaced with private sector companies that aim to maximize their profits. Because of this, the number of market participants increased in the sector, and the competition level became intense. Consequently, they need decision support systems to minimize risks and increase their margins (Bunn, 2004). Therefore, the importance of accuracy in day-ahead electricity price forecasting increased because, with this facility, participants can set their production plans efficiently and develop their bidding strategies to maximize their profits (Li & Becker, 2021).

As electricity markets became more integrated, the factors that needed to be considered in electricity price prediction increased, making accurate forecasts even more challenging (Li & Becker, 2021).

Applying classical financial models for MCP prediction is challenging because the electricity market is quite different from financial markets; the supply and demand sides are far more complicated, and end customers' demand is highly inelastic. Also, the demand must be balanced to avoid electricity shortages and or overload in the electricity system. On the other hand, on the supply side, volatile and inflexible electricity plants are acting in the same system; all of these factors make electricity price prediction highly complex (Lehna et al., 2022)

The production performance of renewable energy generation facilities depends on many external factors; because of this, they are adding incremental difficulties in predicting electricity prices. When we have a more significant proportion of electricity produced from renewables, the complexity will also increase, and therefore, it will be harder to predict the electricity price (Lehna et al., 2022). When we consider this factor, we can say that it is becoming more difficult because renewable energy generation has gained importance in the global energy mix, and it continues in this way:

Electricity utilities cannot pass their costs to their retail customers, and because of this, they are too vulnerable to price changes; this is confirmed in the California electricity crisis of 2000-2001 (Joskow, 2001). It is expensive to sell/buy in balancing power markets because of over or under-contracting; these costs can lead to significant financial losses or bankruptcy. Price volatility can reach extreme levels in electricity markets, which can be much more severe than any other commodities; because of this, market participants have to hedge both volume risk and price movements (Weron, 2014). Any participant in the electricity market who can forecast electricity market clearing prices with a reasonable level of accuracy can follow successful bidding strategies and adjust their production or consumption

schedule to reduce the risk or maximize the profits in the day-ahead trading market (Weron, 2014).

2.4.1 Types of models

Gürtler and Paulsen (2017) classify the models used for MCP prediction into four categories: Simulation, heuristic, computational intelligence, and statistical methods.

Simulation models are applied for conditions where the drivers of price creation, like demand and supply side dynamics, are known. They are based on fundamental price drivers, such as the behavior of market participants and technical market restrictions.

Heuristic models are so-called naïve methods. Moving the average of past values and using the price of one previous day and hour can be examples of heuristic models. The results created with heuristic methods can be used as a benchmark when evaluating the performance of more sophisticated techniques. They assume the same dynamics will be in place with the previous observations. There is seasonality in MCP, and because of this, using one week or one day before same-hour prices makes sense, but this is a random walk with one possible state.

Because of the nature of electricity consumption, we have a seasonality both within days and weeks. For instance, the electricity consumption behavior of people in the same region on Sunday is similar to the one we had on the previous Sunday. On the other hand, the consumption behavior is quite similar to the ones that happened in the same hour between Tuesday and Wednesday. Because of this nature, electricity prices can have similarities between the same year, same month, same day, same weekday, etc. The similar-day method is a prevalent method that is applied

based on searching a day & hour of prices with similar characteristics to the day subject to the prediction (Shahidehpour et al., 2003).

A Naïve method of benchmarking is introduced by Uniejewski et al. (2016); with this simple method, researchers can compare the results of their models with benchmark results selected from past data. If their prediction cannot outperform the naïve benchmark, it is said that the prediction is not successful enough (Pinhão et al., 2022). The procedure is simple enough to get the previous week's same-day results to apply to the recent week's same-day, same-hour results.

Another way to be used as a benchmark is exponential smoothing; it is more often used for load forecasting (Taylor, 2010). Although it has limited use for MCP prediction, it is worth considering that it is the exponentially weighted average of past observations. Cruz et al. (2011) applied double seasonal exponential smoothing in their study and got slightly better results than ARIMA. Also, they performed better than the naïve method. It is shown that dynamic regression models performed better than exponential smoothing, but for a single hour of a day, exponential smoothing performed better than all the other alternatives.

Statistical methods based on both previous electricity prices and previous or current versions of exogenous factors, such as production consumption amounts and weather information, used the mathematical combination of these data to predict current electricity prices. The first statistical methods for electricity price prediction originated from statistical methods for load forecasting: they first emerged from substituting prices for loads (Weron, 2014). A powerful aspect of statistical techniques is that it is possible to adapt physical explanations to model components; in this way, it is becoming easier for system operators and engineers to interpret the

model's behavior (Weron, 2014). On the other hand, it is impossible to capture non-linear relationships using statistical models; this is a drawback of statistical models, but the prediction performance is comparable with non-linear-enabled models (Weron, 2014).

As discussed in the studies of Ziel and Weron (2018), there are univariate and multivariate frameworks in MCP prediction. This is called a univariate framework when the prediction is conducted with a single price series over a 24-hour or 48-hour period. If one step-ahead prediction is applied to 24 different time series, this is called a multivariate framework. In their studies, Weron and Misiorek (2008) used a univariate framework for Nord Pool's day-ahead electricity price; following this study, Kristiansen (2012) applied the multivariate framework to the same data and argued that a univariate framework works better in terms of prediction precision. On the other hand, it is discussed in the study of Crespo Cuaresma et al. (2004) that a multivariate framework performs better than an univariate framework regarding prediction accuracy.

The multiple regression method is widely used in MCP prediction to understand better the interactions between several independent or predictor factors and a dependent or criterion variable. With regression models, it is desired to gain insight into the connections between several independent or predictor factors and a dependent or criterion variable (Pinhão et al., 2022). Least squares are the main component of multiple regression: model fit occurs where the sum of squares of the differences between observed and predicted values is minimized. The classical form of multiple regression assumes the relationship between variables is linear. Regression models are often used in literature for MCP prediction, usually combined

with more sophisticated methods. When regression models include lagged electricity prices as regressors, they are called auto-regression methods (Weron, 2014).

The standard time series model that considers the random nature and time correlations is named the Autoregressive Moving Average model. The current value of the price Xt is linearly expressed in terms of its past p values (autoregressive component) and its q past values of the noise (moving average part) in the ARMA(p, q) model. ARMA models consider the time series data under process to be stationary. If it is not stationary, related transformations should be conducted before prediction. This can be easily achieved using differencing. In their study, Box et al. (1976) applied AR and MA parts of the ARMA model and explicitly applied differencing in the formulation (ARIMA). Sometimes, the simple differencing is insufficient to make the series stationary, even if applied more than once. In some cases, the seasonal nature of data requires differencing with longer lags; this is the case with electricity load and price forecasting. These processes are known as Seasonal ARIMA (SARIMA) models, considering the seasonal component of the data.

Autoregressive type models are crucial for electricity price prediction using time series models. Some studies use the SARIMA method, but in most literature, it is used with exogenous variables (SARIMAX). In their studies, Misiorek and Weron (2005) used different autoregression methods to predict electricity prices in the California market. They have seen that the AR model with 24, 48, and 168 hours lag and each hour of the day is modeled separately performed better than a single SARIMA model specified for all hours.

Conejo et al. (2005) proposed a wavelet-transformed ARIMA technique in another study. It consists of the decomposition of price series using discrete wavelet

transformations (DWT), approximation series using ARIMA to predict 24 hours, and applying the inverse wavelet transformations to get the actual prices for the next 24 hours. In general, the performance of wavelet-transformed ARIMA is better than standard ARIMA; in four weekly test samples, the mean weekly errors are reduced.

Haldrup and Nielsen (2006) applied a more econometric approach to the question and found it very convenient to apply long memory and fractional integration in Nord Pool area data between 2000-2003. This can be explained by the dependency of the Nord Pool market on hydropower electricity generation. In classical empirical findings, the river flows, and reservoir levels exhibit long memory, which is critical for hydropower plants. In their studies, they applied ARFIMA models to Nord Pool prices and used them to make predictions.

Electricity prices are affected by many exogenous factors, such as weather conditions, generation capacities, and load profiles. For electricity price prediction, time series models including the exogenous factors are named with ARX-Type models, such as ARX, ARMAX, ARIMAX, and SARIMAX, are versions of AR, ARMA, ARIMA, and SARIMA including exogenous factors, respectively.

Because of the nature of the electricity prices, time series models with exogenous factors are extensively used in the literature. For instance, in Nogales et al. (2002), ARX and ARMAX methods are utilized to predict the electricity prices of California and Spain. The results of these modes were better in terms of Mean Weekly Error compared to ARIMA and ARIMA-E (ARIMA with load variable as an explanatory variable) proposed by Contreras et al. (2003). Although they include the same variables, the ARIMAX method outperformed the ARIMA-E function with more than %40. This can be explained by the inclusion of load features into two

models. In the ARIMA-E model, it is included just as an explanatory variable; on the other hand, in the ARIMAX model, it is tied with the autoregressive part of the model.

In their study, Misiorek and Weron (2005) implemented 24 different ARX models to predict each hour of the day separately. In this model, they have used California ISO's day-ahead electricity price forecasts as an exogenous variable. Also, they employed three dummy variables to give the effect of weekly seasonality on the model. They have seen that this new approach is working better than creating a single (S)ARIMA model for 24 hours proposed by Contreras et al. (2003) and slightly worse than the ARMAX and ARX models proposed by Nogales et al. (2002). The first week of April 2000 in the California Electricity market was the only comparable period commonly used in these studies.

In another case study, Weron and Misiorek (2006) calibrated ARX models to spike processed prices and saw that results improved for the first week of April compared to the other models. When the prices became more volatile, spike-calibrated models started not working that much precisely. This experiment shows that spike-processed ARX and ARMAX models work better in markets where prices are more stable than markets where we see price volatility.

Computational or artificial intelligence models are based on the nonlinear or hidden patterns in the data. Input-output relationships are complex in these models; training algorithms can identify these relationships, which are hard for other models to capture.

In energy research, machine learning (ML) is increasingly utilized in various areas, such as energy policy, market analysis, and environmental policy, aiding in (1)

incorporating more relevant and higher frequency information, such as weather forecasts—often omitted by conventional energy economics models, (2) managing the complexity of variable relationships—which the standard parametric models may not well capture, and (3) performing sensitivity and heterogeneous effects analysis (Athey & Imbens, 2017; Chernozhukov et al., 2017; Jedrzejewski et al., 2022; Stock & Watson, 2017).

The benefits mentioned above are accompanied by a drawback: reduced clarity regarding the significance of the incorporated variables and an increased possibility of misinterpreting the outcomes. Therefore, employing innovative analytical techniques requires thorough justification and careful assessment (Madadkhani & Ikonnikova, 2024).

While all the features collectively play a role in the prediction process, assessing the significance of each feature in the decision-making process is challenging. This complexity arises from numerous correlations among the variables, necessitating a comprehensive consideration of their interactions to gauge their impact accurately. Shapley values, originating from game theory, were introduced to distribute gains among multiple participants in a cooperative setting equitably. Equity implies that each participant's contribution is acknowledged in determining the overall gain. Consequently, participants are remunerated for their individual gain and collaborative input within the group dynamic. In their studies, Tschora et al. (2022) and Xie et al. (2022) used SHAP values to assess the role of the input features in influencing the model's predictions

2.5 Impact of ceiling prices on price prediction

The imposition of censors, or limitations, on the output of predictive models significantly influences their effectiveness and accuracy across various domains. In predictive analytics, censors typically represent predefined constraints or boundaries within which the model's outputs must fall. While these limitations are often necessary for regulatory compliance, ethical considerations, or practical feasibility, they can also skew the predictive accuracy. For instance, in financial forecasting, censors might cap predicted values to prevent unrealistic market expectations, but this could lead to underestimation of market potentials or risks. Similarly, censored predictions in healthcare might fail to fully capture the range of possible patient outcomes, affecting treatment plans and resource allocation. The key challenge lies in balancing the need for realistic and responsible predictions while acknowledging the full spectrum of possible outcomes. Advanced statistical techniques, such as censoring-aware machine learning models, are being developed to address this challenge. These models are designed to recognize and adjust for the effects of censors, thereby producing more nuanced and accurate predictions that reflect both the constraints and the underlying data trends. The impact of censors, therefore, extends beyond mere limitation; it necessitates a deeper understanding and integration of these constraints into the predictive modeling process to ensure that the insights generated are both useful and realistic within the defined boundaries.

CHAPTER 3:

METHODOLOGY

The limiting impact of ceiling prices on market clearing prices is undeniable for hours when we see the ceiling price is touched. On the other hand, the ceiling could impact all hours of the day. To analyze this effect, a comprehensive machine learning model comprising data sources of production, consumption, supply-demand curve important points, weather conditions, marginal prices of some essential production types, and many other features are constructed, and ceiling price information of the trading day is also integrated into this model. The output feature we want to predict using these features is the market clearing price. This methodology is employed to take advantage of prediction models to explain the relationships between dependent and independent variables. The following section will introduce the data used to build this machine learning model.

In the realm of predicting market clearing prices, the Light Gradient Boosting Machine (LightGBM) model demonstrates notable proficiency due to its distinctive features and methodologies. A pivotal aspect of LightGBM is its Gradient-based One-Side Sampling (GOSS), which strategically focuses on more informative data points, thereby enhancing the model's efficiency and accuracy in complex market scenarios. This is complemented by its inherent capability to handle categorical features effectively. It is particularly advantageous in energy markets where we must feed hour, day, month, and year information as categories. Moreover, LightGBM is commendably resilient against overfitting, a common challenge in predictive modeling, thereby ensuring the reliability and generalizability of its market price predictions. In our example, our data does not have a huge volume; because of this,

overfitting handling was necessary in our decision. The model's adaptability is further evident in its customization potential, allowing it to be finely tuned to align with specific energy markets' unique dynamics and characteristics. The model's speed and performance are unparalleled, mainly due to its histogram-based learning approach, enabling rapid and efficient processing of large and complex datasets commonly encountered in market price prediction.

We have selected dates between 5 July 2023 and 31 July 2023 as our test period to predict market clearing prices. This recent day has the market clearing price of 2700 TL/MWh, and the proportion of hours where the ceiling price is seen is %10.7, which is a fair proportion when we compare this with other periods of ceiling prices. Starting from one previous day of this period, two years of data are selected as our train data to capture the impacts of the seasonality of 2 years of data. With the data described in this section, we have built a successful model that accurately predicts market clearing prices for this period.

We aimed to analyze the impact of features on market clearing price, but machine learning models like LighGBM are so-called black-box models that need additional tools to understand the model's behavior. In the domain of machine learning, SHapley Additive exPlanations (SHAP) has emerged as a pivotal tool for interpreting complex models, underpinned by its foundation in cooperative game theory, which ensures consistent and fair allocation of feature contributions to model predictions. This characteristic is particularly beneficial in providing detailed and individualized explanations of model outputs, a crucial aspect in applications requiring precise understanding at the instance level, such as personalized medicine or credit risk assessment. Furthermore, SHAP extends its utility beyond local explanations, offering global interpretability by aggregating SHAP values across

datasets. This feature enables a broader understanding of feature influences, facilitating more effective feature engineering and model refinement. The granularity of insights SHAP provides is noteworthy, as it identifies influential features and quantifies their impact and directional influence on model predictions. Additionally, SHAP's ability to capture and elucidate complex feature interactions is precious in unpacking the decision-making processes of sophisticated models, where interactions can significantly sway predictions. Complementing these technical advantages, SHAP's suite of visualization tools, such as summary and dependence plots, enhances interpretability, making it accessible to diverse stakeholders. The SHAP tool has been selected to interpret our machine learning model for the aforementioned reasons.

After model behavior analysis, we can select the most important features from the model. For the next step of our analysis, which is the impact of ceiling prices on the other determinants of ceiling prices, we have built a statistical model that employs the most important features of the machine learning model. We have built a Tobit model to understand the impact of ceiling prices on the determinants of market clearing prices. The Tobit model, named after economist James Tobin, is a type of regression model designed to estimate linear relationships between variables when there is censoring in the dependent variable. The Tobit model is a powerful tool when the standard assumptions of linear regression are violated due to censoring in the dependent variable. In our example, it is used to predict the market clearing price from the selected important features that determine the clearing price. To understand the impact of censors on the coefficients of the equation, a linear regression model employing the same features of Tobit is created as a control model. These two

models' coefficients and intersection points are compared to understand the impact of ceiling prices on the determinants of market clearing prices.

3.1 Data

Most of the time, data that can be used for market clearing price prediction is publicly available. For instance, there are more than 140 data sets in the Turkish Energy Market (EXIST) Transparency platform. The data can be reached using file imports or web service API tool connections. Feature engineering is one of the most important aspects of electricity price prediction in this sense. Because it requires a business understanding of the electricity market and being able to evaluate the candidate features contributions to the model. In this project, we have created new features that can be fed to the model using different data sources. Data is key to the success of the model. Because of this, feature selection is carefully done for this project.

3.1.1 Consumption data

Consumption is one of the most critical drivers of MCP. Participants bid to optimize their consumption and production portfolios in the day ahead electricity market. Generally, we see high prices when we see high consumption amounts; otherwise, vice versa.

In Figure 6, you can find a plot of daily consumption and daily MCP means for dates after 2022-01-01. In this view, we can see the monthly seasonality effect of consumption. We are seeing higher consumption amounts in months that we have very high and very low temperatures because, in these months, we need to make

temperatures that we live warmer or cooler. In other words, consumption is higher in the summer and winter months because of temperature effects. It is lower in the spring and autumn periods.

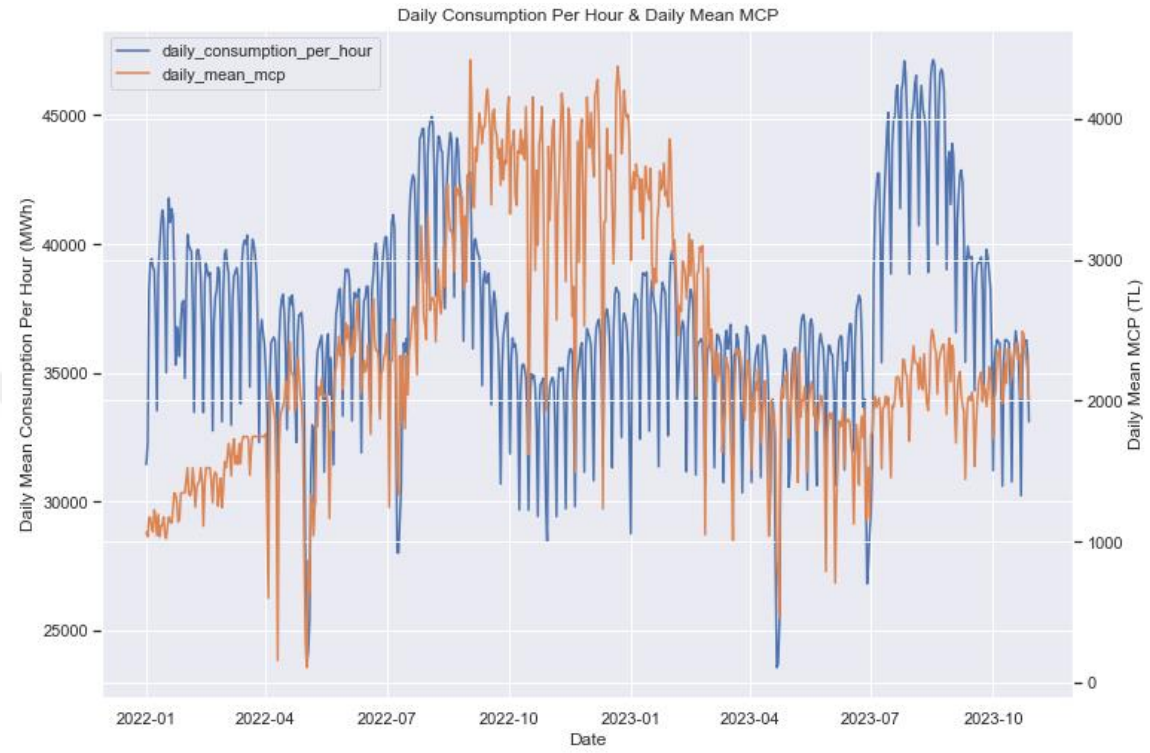

Figure 6. Daily consumption per hour and daily mean market clearing price

Plus, data includes another weekly seasonality. On working days of the week, we see higher consumption. On the other hand, we see lower consumption on weekends because industrial consumption is lower.

Figure 7 is another example of three weeks of data between 2023-02-13 (Monday) and 2023-03-05 (Sunday). We see the lowest consumption days on Sundays (2023-02-19, 2023-02-25 and 2023-03-05). The second lowest consumption days are Saturdays, and we are seeing higher consumption in weekdays:

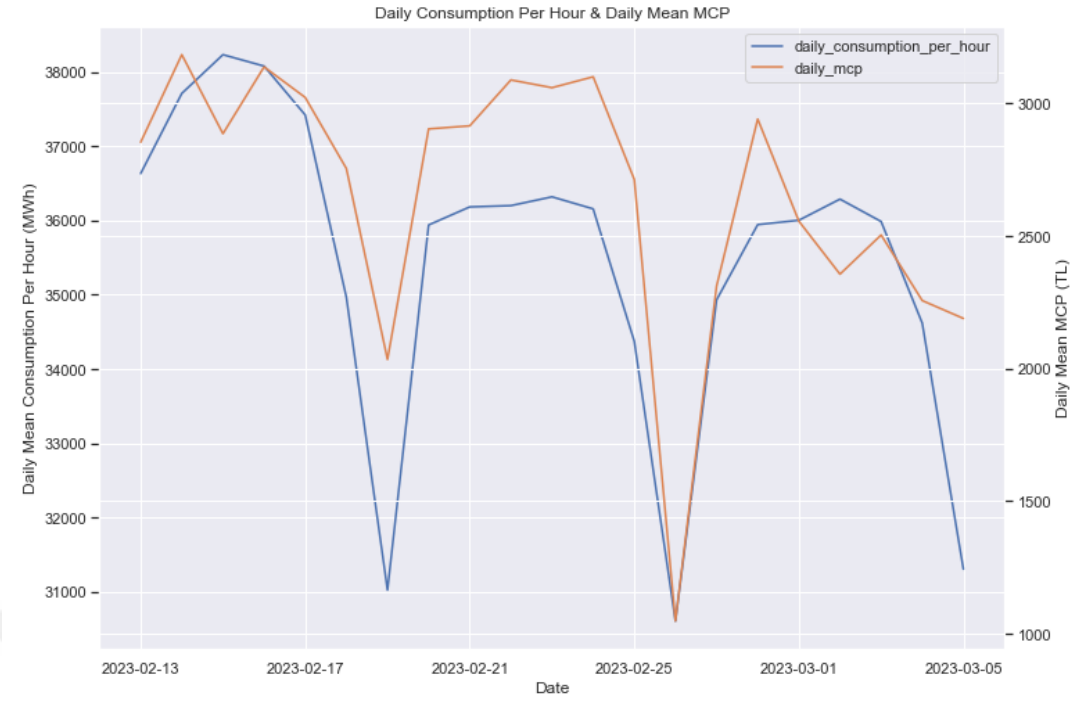

Figure 7. Daily consumption per hour and daily mean market clearing price

On the other hand, consumption is also affected by intra-day seasonality. We have peak hours when we see higher amounts of daily consumption: noon or afternoon. The consumption is lower during night hours because most people sleep during these hours.

You can find an example from the day of 2023-02-01 where we can see all hours of the day clearly in Figure 8.

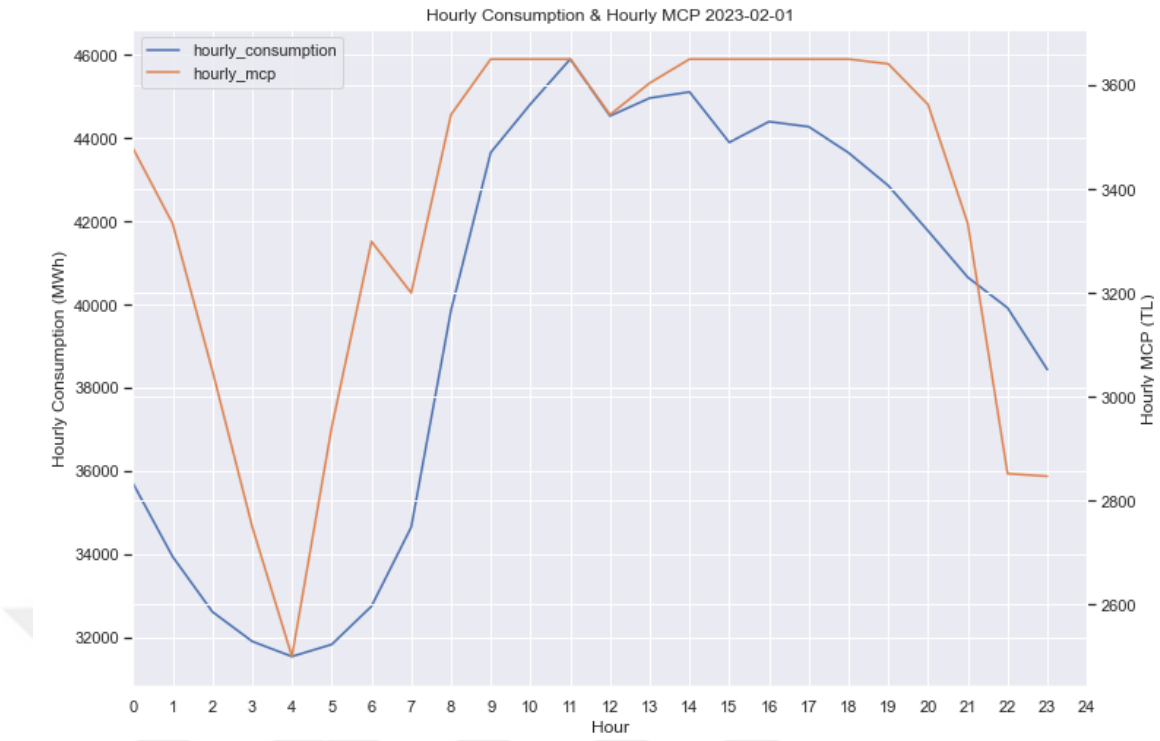

Figure 8. Hourly consumption and hourly market clearing price for a specific date

The plot above shows that consumption has its lowest values in night hours (4 AM with 31531.37 MWh in this example). On the other hand, we see the highest amount at 11 AM, with 45900.64 MWh. At these highest and lowest values, we have seen the highest and lowest MCPs in the day: 2500.01 TL lowest, 3650.00 TL highest (This is the ceiling price of the period).

3.1.2 Production data

Consumption in a day should be met with production to sustain the balance in the market. Because of this, it is very critical to include production amounts in the prediction model. Electricity cannot be feasibly stored in the current technological structure. Because of this, the produced loads should be consumed by consumers. When facilities produce amounts higher than their plans, this will create an

imbalance in the market, and this is subjected to penalties by transmission system operators. Optimizing production amounts is one of the most critical tasks of electricity production facilities.

Production plans can have different characteristics depending on the production type of the facility. For instance, some base load facilities need to make sure that they are producing amounts that are determined by the system operator. On the other hand, most renewable energy production facilities cannot decrease their production amounts. Their production is significantly dependent on the weather conditions. If there is high-speed wind, wind facilities do not reduce their production amounts even if the consumption is low (Of course, they can shut down when wind speed reaches dangerous levels.)

Production data has different characteristics because it includes production amounts from many different source types. It is affected by the weather conditions, but this time, wind speed or sun exposure levels play an essential role in the production. Also, water levels in rivers and barrages are critical because of the hydropower electricity production.

We are seeing similar seasonality to what we have seen in consumption. This is because production needs to follow consumption and should not exceed consumption.

The actual production amounts are used as a feature in the model to capture the effects of production.

In the Figure 9, we see that the relationship between production & MCP and the relationship between consumption & MCP are quite similar. This is because of the balance in the market.

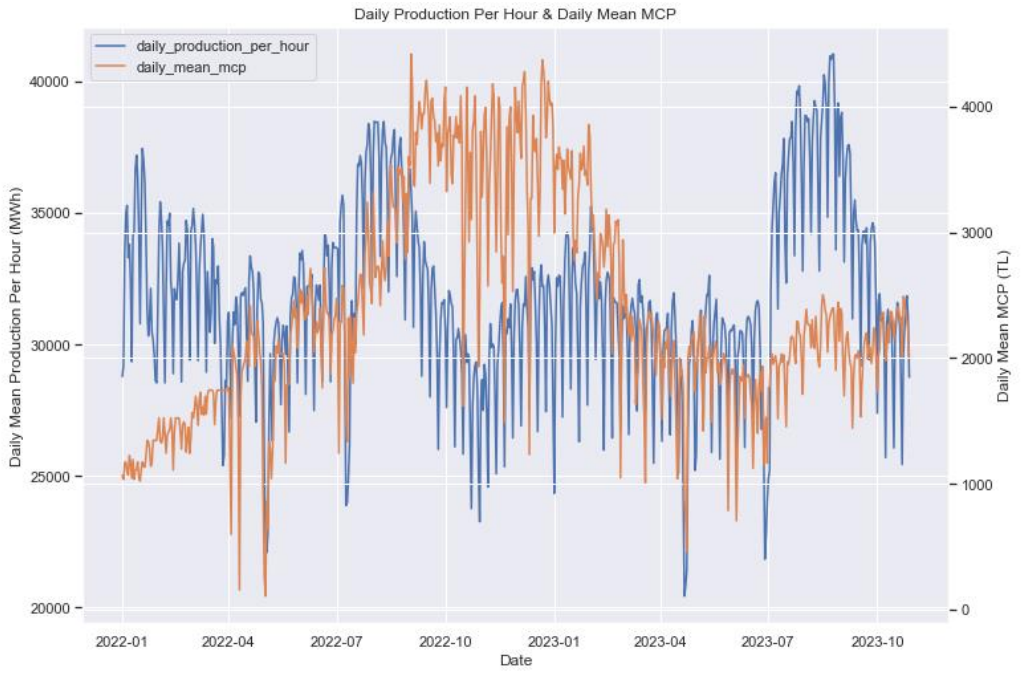

Figure 9. Daily production per hour and daily mean market clearing price

We are also seeing a similar characteristic of production in weekly period in Figure 10. To meet the higher demand in working days, we are seeing higher amounts of production for working days; on the other hand, we are seeing lower amounts of production in weekend periods.

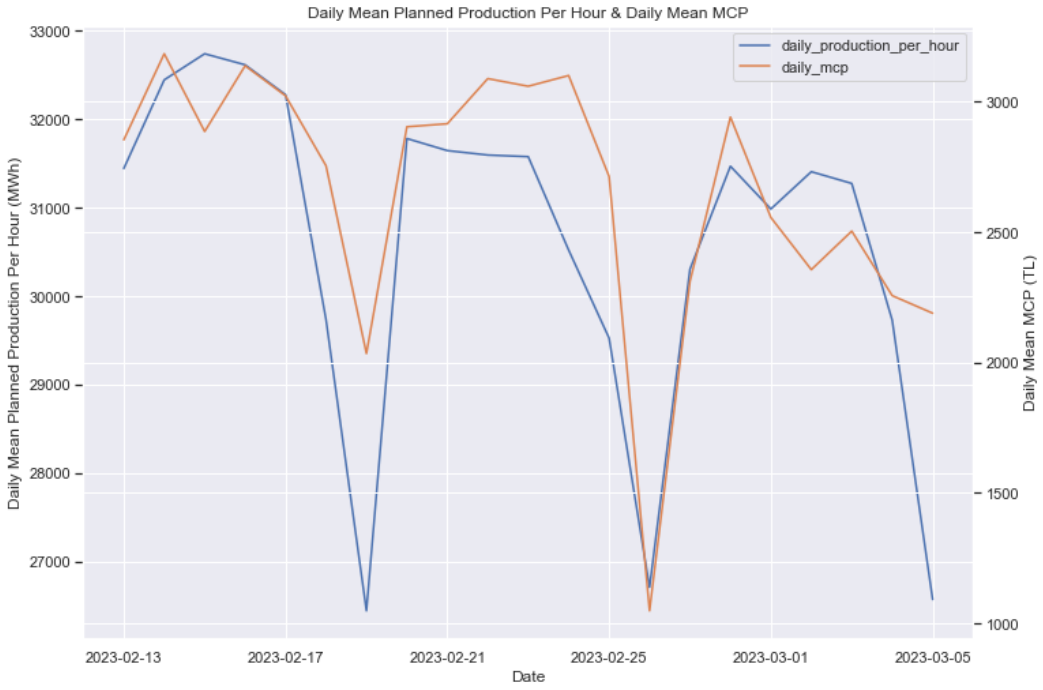

Figure 10. Daily mean production per hour and daily mean market clearing price

In Figure 11, hourly production also has intra-day seasonality; to meet the higher demand in peak hours, the production facilities plan similar high production amounts. Productions are lowered for low-consumption hours of nights, too.

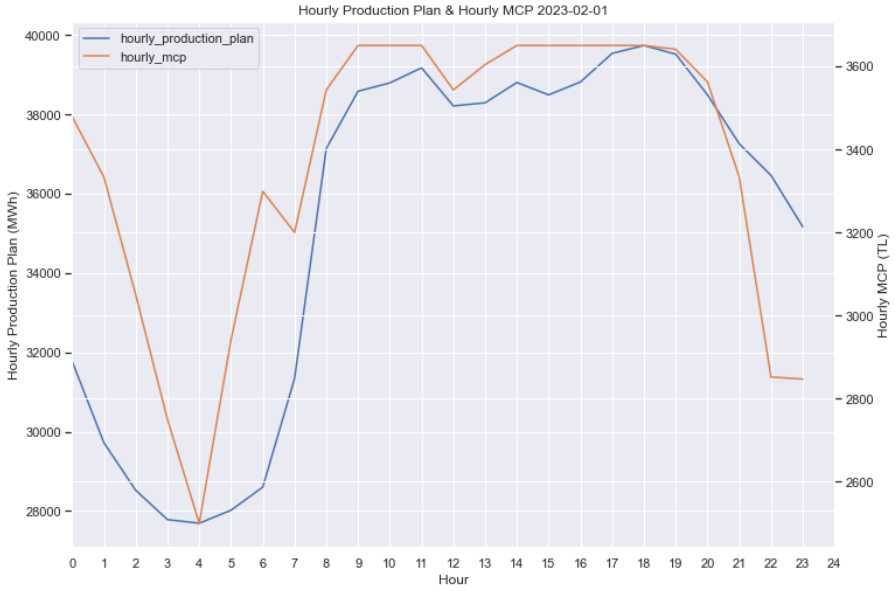

Figure 11. Hourly production amounts and market clearing price for a specific date

Although production and consumption curves seem very similar at the aggregate level, they differ in more granular data, especially at the company level. Moreover, this creates the logic of MCP; it is the price that we will see in points where supply and demand curves intersect. Because of this, it is dependent on many factors that can affect both supply and demand.

3.1.3 Production data for important sources

After including aggregate production data in the model, it is also essential to have some of the critical energy sources' production data in the model. These features aim to give the model the ability to reflect the dynamics of different sources of energy production. Renewable energy production sources have one of the most critical impacts on the market; the cost of energy produced in these facilities is comparatively lower than other energy sources. Therefore, they can bid for meager prices because of this cost efficiency. It is known that market participants of renewable energy sources are eager to give price-independent sale bids in the market because they have to sell what they have produced in the market because of the non-storable feature of electricity; it must be consumed when it is produced.

3.1.3.1 Production in renewable energy resources support mechanism

Turkey has implemented the Renewable Energy Resources Support Mechanism (YEKDEM in Turkish) to encourage the development and utilization of renewable energy sources within the country. The YEKDEM is a feed-in tariff system that incentivizes renewable energy producers, particularly in wind, solar, hydropower, and geothermal sectors. Under this mechanism, eligible renewable energy projects

are guaranteed a fixed tariff for a specific duration, typically ten years, providing investors with a stable and predictable income. The scheme aims to attract private investments, enhance energy security, and contribute to environmental sustainability by reducing reliance on fossil fuels.

As of their agreement with the government, YEKDEM-supported facilities are giving price-independent sale orders in the day ahead electricity market. Because of this, they have a decreasing impact on the market clearing price.

YEKDEM-supported facilities' production amounts are published in the Transparency platform with the details of licensed and license-exempt types.

To simplify the model, the amounts from wind, sun, river, dammed, biomass, and geothermal facilities from both licensed and license-exempt types are aggregated into one feature that will give the total amount which is produced by the YEKDEM-supported facilities, which gives price-independent sale orders in the market.

The methodology of the YEKDEM-supported facility's contribution total with the example of wind-sourced production can be found in Figure 12.

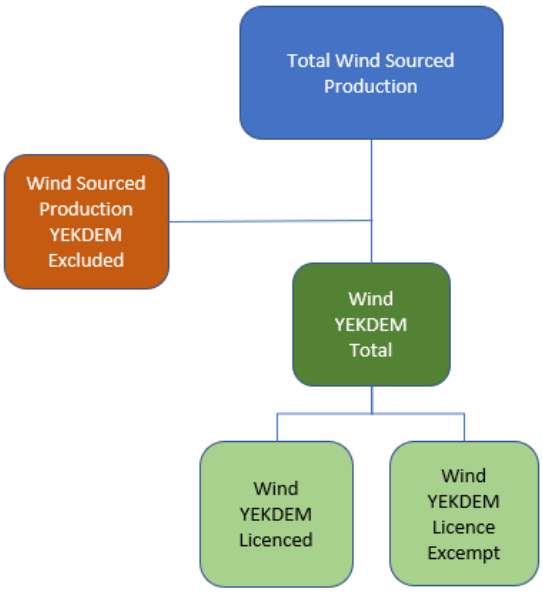

Figure 12. Methodology used to calculate YEKDEM-supported facilities

In wind-sourced production, YEKDEM licensed wind-sourced production and YEKDEM license-exempt wind production are summed up to include the total amount of wind-sourced YEKDEM production. This feature is created for wind, solar, river, dammed, biomass, and geothermal sourced productions and summed up to the total YEKDEM sourced production amount. On the other hand, the total value of the resource-based electricity amount given to the system hourly by the settlement units within a settlement period by production types is also published in the Transparency Platform. This data consists of the total production amounts for the specific production type. We have the total, and YEKDEM supported amount for each production type; by subtracting the YEKDEM supported amount from the total amount, we create the total production amount except the YEKDEM support mechanism for each production type in YEKDEM. After this part, the significance of YEKDEM excepted amount is checked and used in the model accordingly.

As expected, the amount of production from YEKDEM-supported facilities has a highly negative correlation with MCP in Figure 13:

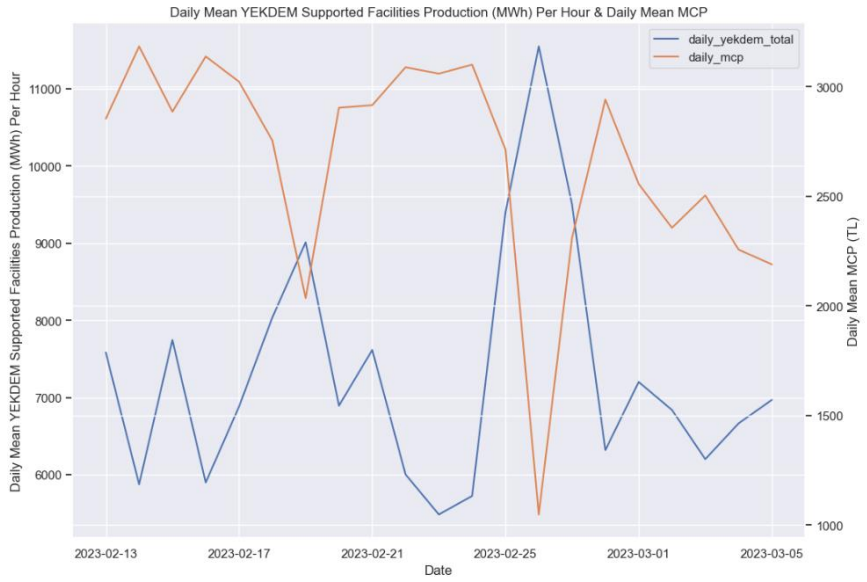

Figure 13. Daily mean YEKDEM supported facilities production per hour and daily mean MCP

3.1.3.2 Wind sourced production amounts

Wind is one of the most essential sources of energy for Türkiye. In 2022, %10.6 of energy produced was from wind sources. Because it is not dependent on any other input price, the cost associated with the electricity produced is considerably lower than other sources. Because of this, we can say there is a negative correlation between wind-sourced electricity production and MCP. An important amount of wind-sourced production is subject to the YEKDEM; because of this, they are already included in the features of YEKDEM productions. In this feature, we only feed the wind amounts without YEKDEM support in Figure 14.

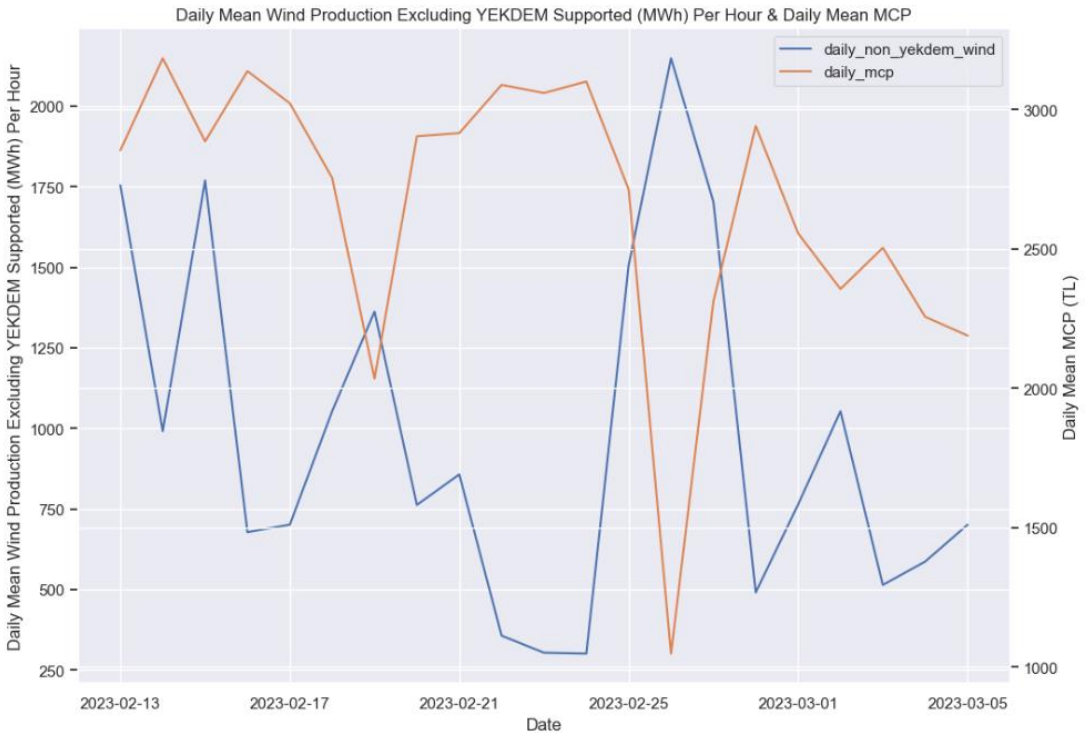

Figure 14. Daily mean wind sourced production per hour excluding YEKDEM and daily mean MCP

It can be seen from the above graph that although these facilities are not supported by YEKDEM and, therefore, they do not have to give price-independent sale orders by contract, for dates where we see higher amounts of wind energy production, we have considerably lower MCP.

3.1.3.3 Dammed hydroelectricity production

Hydroelectricity consists of %20.3 of electricity produced in 2022 in Türkiye. It is the top renewable energy source in Türkiye's installed energy capacities. It is also very cost-efficient because there is no input cost associated. On the other hand, most of the hydroelectricity production facilities are still owned by EÜAŞ (The Electricity

Generation Corporation), a state-owned company. We can also say that bids from hydropower companies tend to decrease the MCP for these reasons.

As with other renewable energy sources, we will include only the amounts YEKDEM does not support in this feature in Figure 15.

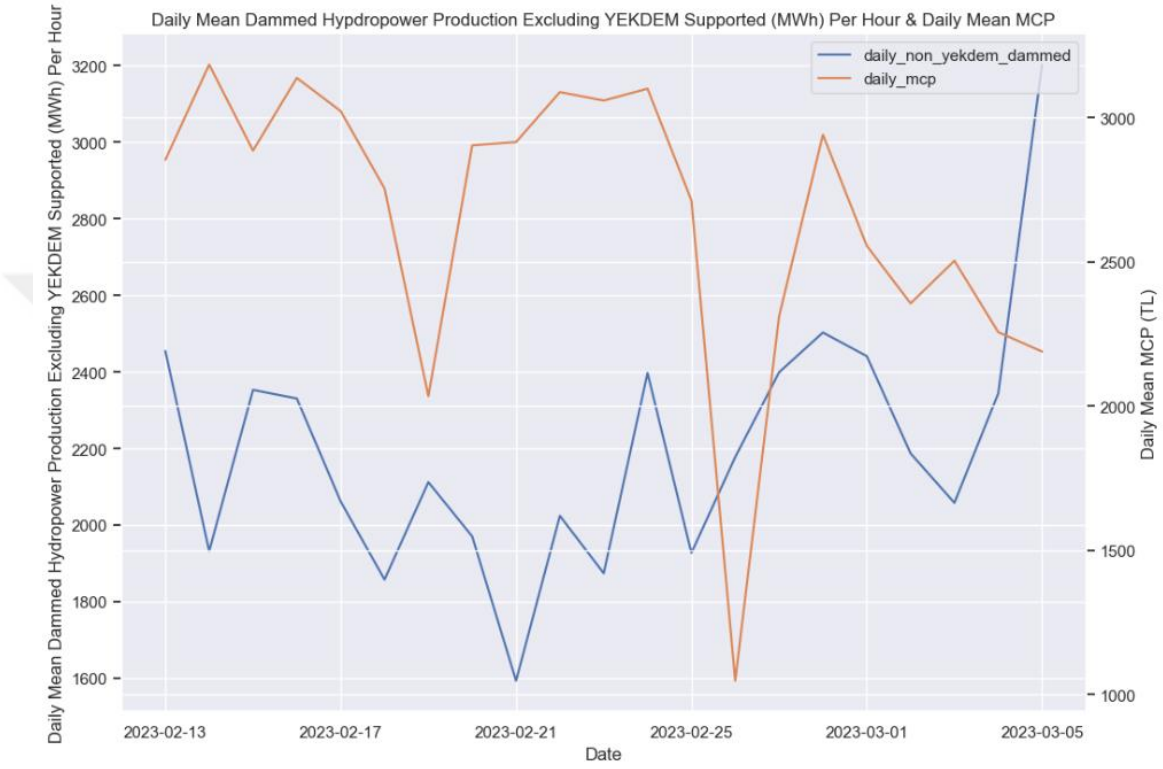

Figure 15. Daily mean dammed hydropower production per hour excluding YEKDEM supported and daily mean MCP

We see a negative correlation with MCP in hydropower facilities like wind production. One essential feature that makes dammed hydropower facilities unique is that they do not have to generate electricity according to the weather conditions with the help of barrages. Operators can decide not to produce electricity even when they have the potential to generate it. This gives EUAŞ the ability to balance the electricity market when it is needed. For instance, they may decide not to generate &

sell electricity in times when market prices are low. They may decide to generate electricity at high price periods to stabilize the market prices.

3.1.3.4 Natural gas sourced production amounts

Natural gas is a fossil fuel composed primarily of methane (CH4). It is extracted from underground reserves and can be used as a fuel for various purposes, including electricity generation. Natural gas-sourced electricity production was consisting of %22.9 in 2022 in Türkiye.

Natural gas production is considered more stable because it is not dependent on volatile weather conditions. On the other hand, electricity production from natural gas is more dependent on the price of natural gas as it is being used as input to the production.

Currently, Türkiye is a natural gas importer country where %99.6 of consumption is imported. It is sourced from Russia, Iran, Azerbaijan, and other countries. Besides, new natural gas reserves have been discovered in the Black Sea region. The extraction and transfer infrastructure are still in progress; after it gets started being used, this dependency will be weaker.

Natural gas-sourced production is crucial to sustaining the country's energy security. When weather conditions are inconvenient for renewable energy sources, production facilities like natural gas are essential in meeting the energy demand.

Because production cost depends on the natural gas used, it is expensive to generate electricity. Therefore, it has a positive correlation with MCP. The more we have to run natural gas facilities to generate electricity, the higher electricity prices we will see in the market, as can be seen in Figure 16.



Figure 16. Daily mean natural gas production amount per hour and daily mean market clearing price

3.1.3.5 Imported coal-sourced production amounts

Coal-sourced electricity production consists of the highest proportion of Türkiye's energy mix, with %34.6. Coal-sourced production also depends on the price of the material used as input for production. According to the data from the General Directorate of Mining and Petroleum Affairs Türkiye, as of December 2022, coal-based production capacity is taking %21 of Türkiye's installed power capacity. Among this, local coal-based capacity consists of %11 of overall capacity. Because of this, production from coal is less dependent on imported coal, global coal prices and exchange rate than natural gas.

Like natural gas, coal-based production is also planned to meet the consumption. It is also not dependent on weather conditions; it is less volatile than renewable sources.

In recent years, most production capacity investments have been made for renewable energy sources in Türkiye; the dependency on weather has increased. When weather conditions are convenient for renewable energy sources, the demand can be met cost-effectively. On the other hand, to sustain energy security, some coal and natural gas facilities are incentivized to offer their capacity when needed. We can say that when we see higher amounts of electricity produced from coal-sourced production, we see higher prices in the market.

Imported coal-based production is used in our model to include price drivers from this vital source of production, which has costs associated with global coal prices and exchange rates.

In Figure 17, we can see that there is a positive correlation between imported coal-based production and MCP. This is because the sell offers from these sources will have higher prices. After all, they have higher production costs.



Figure 17. Daily mean imported coal sourced production amount and mean market clearing price.

3.1.3.6 Domestic coal-sourced production

Another type of coal-sourced electricity production is domestic coal. Türkiye has domestic coal production capabilities, too. Most coal-sourced electricity production plants are located near the areas where coal is mined. There are three main types of coal in the production mix: lignite, asphaltite, and stone coal.

Here are the important coal production locations of Türkiye:

Zonguldak: Zonguldak, located on the Black Sea coast, has been a significant coal mining region in Türkiye for many years. The province has several coal mines, and it has historically been a major coal production center.

Soma: Soma, situated in the Manisa Province in western Türkiye, is another important coal mining area. The Soma coal basin is known for its lignite coal deposits.

Afşin-Elbistan: Afşin-Elbistan, located in the Kahramanmaras and Adiyaman provinces, is one of the largest lignite coal reserves in Türkiye. It is a crucial energy production center hosting mining operations and power plants.

Tunçbilek: Tunçbilek, in the Kütahya Province, is another area with coal mines and is part of Türkiye's coal production landscape.

Domestic coal is much cheaper than imported coal; because of this, these facilities have much lower marginal costs than imported coal facilities. To simplify the model, the total amount of lignite, asphaltite, and stone coal is fed to the model because they share similar marginal costs.

In Figure 18, you can find the plot of mean domestic coal-based production per hour and daily mean MCP; the correlation between this total amount and MCP is not clearly seen in the data.

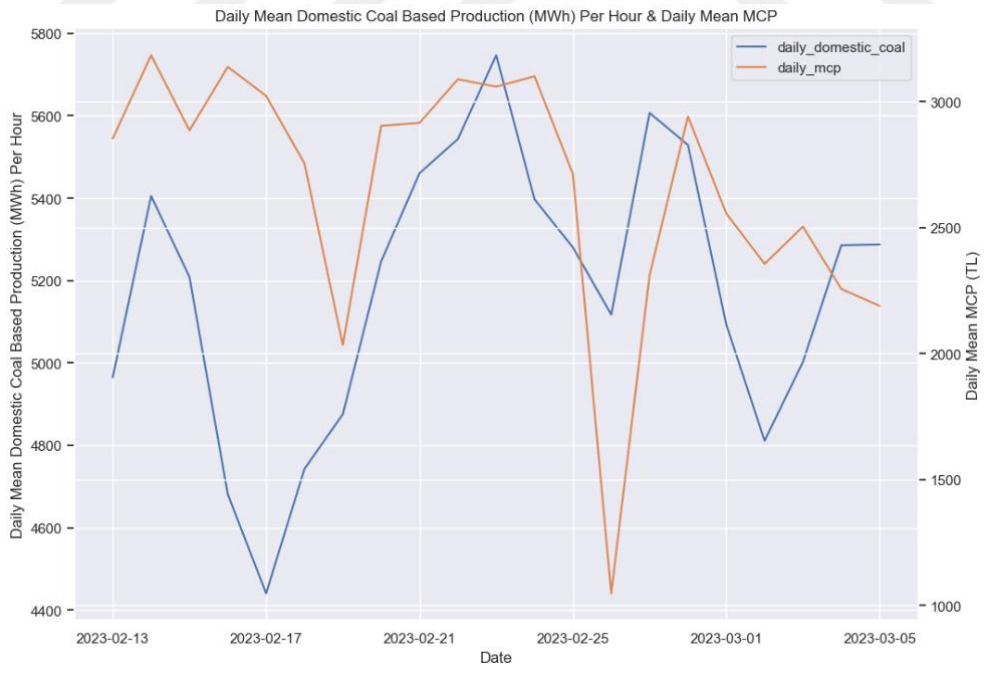

Figure 18. Daily mean domestic coal-based production per hour and daily mean market clearing price

3.1.4 USD/TRY exchange rates

As mentioned in the natural gas and imported coal-based production part, the exchange rate is vital in driving the MCP. Because of this, the exchange rate is also fed into the model to give the model the ability to capture macroeconomic effects that will impact the costs of production facilities.

The effect of exchange rates quickly passes on to the production costs of fossil-based electricity production facilities. Although companies can hedge their risks associated with exchange rate spikes, applying these measures to all participants is impossible. On the other hand, it is bringing additional costs. In countries like Türkiye, where we see exchange rate volatility much after 2018, inflation is mainly derived from exchange rate changes.

In Figure 19, it is seen that MCP follows an overall trend in exchange rates. This is because of the exchange rate pass-through that was mentioned previously.

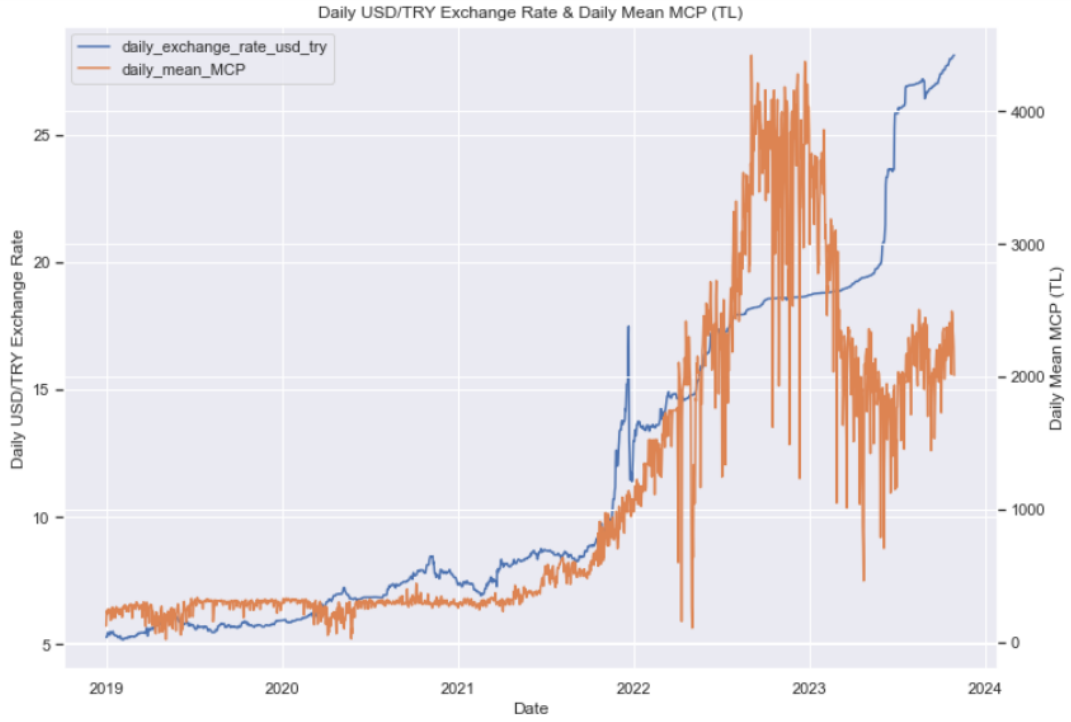

Figure 19. Daily USD/TRY exchange rate and daily mean market clearing price

3.1.5 Natural gas DRP (daily reference price)

As mentioned in previous sections, natural gas-sourced production plays a vital role in the energy market of Türkiye. When considering the price volatility and pass-through rates, it is important to include natural gas daily reference price in our prediction model.

The responsibility for establishing the platform on which the natural gas market will be operated and conducted has been assigned to EXIST. The OTSP operates through the Continuous Trading Platform (STP) developed by EXIST, allowing continuous trading. In the natural gas market, it provides market participants the opportunity to trade for a gas day from 08:00 one day prior to 14:00 one day after. Additionally, the Transmission Company primarily conducts balancing gas purchases and sales on the STP to ensure system balance. Through OTSP, daily

reference prices (DRP), which vary according to market conditions and supply and demand balances in Turkey, will also be shared with all stakeholders of the Turkish energy market via STP.

In general, we also see that MCP is following changes in the DRP of natural gas in Figure 20. It shows how effective natural gas prices are on electricity prices.



Figure 20. Daily mean gas reference price and mean market clearing price

3.1.6 Working day information

Electricity demand is affected by whether the consumption day is a working day. If it is not a working day, the consumption is limited because most industrial consumption is not in the picture. Although day-of-the-week information is fed to the model as a categoric feature, it does not cover non-working days other than

weekends. Because of this, the model could not understand holidays other than weekends. A binary feature is created and named as is_working_day to meet this requirement. If it is a weekend or Türkiye's public holidays, this feature gets the value of 0; if it is a working day, it gets 1.

In the period of Figure 21, we have 1 May Labor and Solidarity Day and Ramadan Feast (2-3-4 May). Although we see lower MCP on weekends, we also see lower MCP on public holidays from 1-4 May.



Figure 21. Is working day dummy feature and daily mean market clearing price

3.1.7 Consumption weighted daily temperatures

Weather information is essential for the model predictions. Renewable energy sources are very effective in MCP creation because their production costs are meager compared to fossil-sourced production. Furthermore, they are very dependent on weather conditions.

Inspired by Gokgoz & Filiz (2017), the consumption-weighted daily temperature feature is created with a similar strategy with more recent data.

We have historical data on weather conditions that can be used for the model. The source of the weather information used in this model is www.worldweatheronline.com. However, weather data is location-specific; on the other hand, we have a common MCP that is trying to be predicted. This is creating the room for feature engineering for weather data. It is known that energy consumption increases when the temperature is very low or very high because of the need for cooling or warming. This relationship between temperature and consumption creates a common feature that can be used for the whole market.

EPDK (EMRA Energy Market Regulatory Authority) monthly announces the electricity consumption information of 81 cities in Türkiye. This feature calculates the proportion of each city's electricity consumption for each month. Moreover, each city's daily temperature information is multiplied by the corresponding consumption proportion calculated. Multiplied figures are summed, and the final temperature information weighted by the consumption proportions is created.

Equation 1 shows the calculation of this daily feature:

$$\tau_m = \sum_{i=1}^{81} w_i t_i \quad (1)$$

i stands for the number of the city in Türkiye; there are 81 cities in Türkiye. w represents the consumption weight of the city, t is for the city's temperature, m is for the day, τ represents the consumption-weighted temperature for Türkiye that will be used for the model.

Consumption valus of highest consuming cities can be found in Table 1.

Table 1. Top 10 Cities Having Highest Proportion of Energy Consumption for May 2023

Cities	Lightning	Public/Private/Other	Household	Industry	Agricultural Activity	Total	Proportion
İstanbul	35,945.29	1,336,509.07	1,064,122.31	989,451.46	663.55	3,426,691.67	0.17
İzmir	16,805.35	314,198.11	332,857.58	550,871.54	34,898.55	1,249,631.13	0.06
Ankara	21,926.57	388,772.10	420,316.30	354,605.97	20,273.38	1,205,894.32	0.06
Bursa	10,231.58	169,711.85	195,353.08	672,266.02	7,718.56	1,055,281.09	0.05
Kocaeli	6,899.91	123,118.45	120,859.96	743,047.84	1,493.81	995,419.97	0.04
Konya	12,166.66	152,268.04	137,302.47	193,141.57	237,678.71	732,557.45	0.03
Gaziantep	6,515.02	87,963.60	123,639.34	486,370.07	19,308.67	723,796.69	0.03
Antalya	15,408.28	346,696.98	191,210.24	124,871.93	14,584.59	692,772.01	0.03
Tekirdağ	4,313.87	69,122.20	68,688.70	487,291.80	1,000.65	630,417.22	0.03
Adana	9,656.26	110,542.21	137,643.20	301,996.03	18,338.02	578,175.72	0.02

Source: EMRA, Electricity Market Sector Report (2023)

Using 81 cities' data consumption, weighted temperature data is created.

In Figure 22, the classical seasonality of temperature levels in Türkiye can be seen.

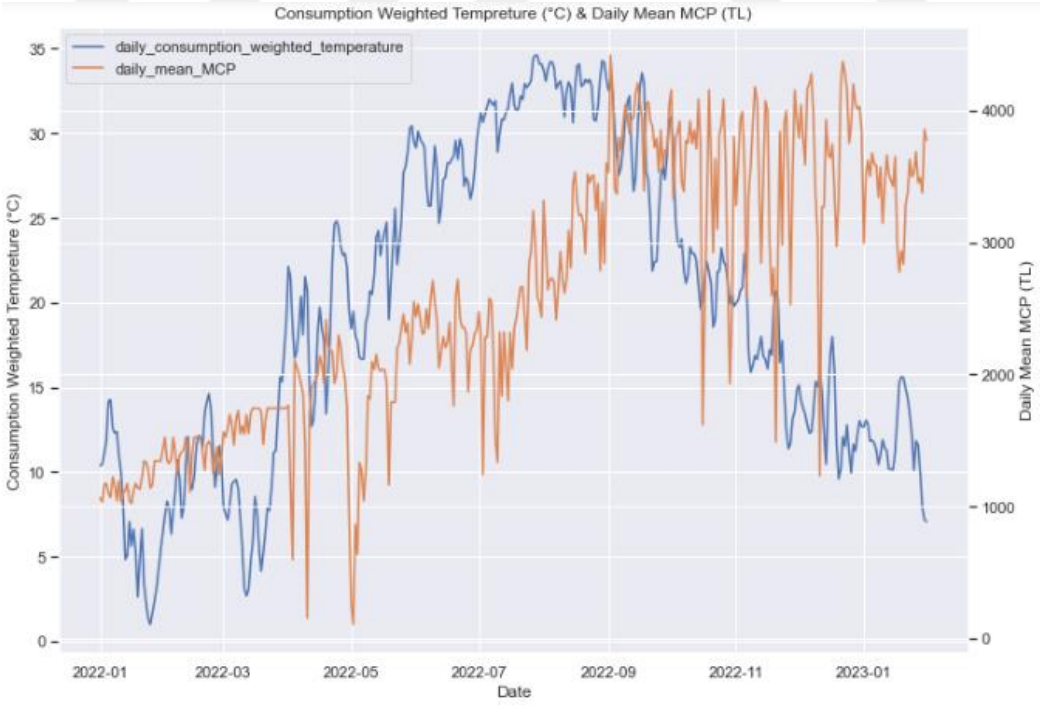

Figure 22. Consumption weighted temperature and daily mean market clearing price

Feeding temperature type information may not give the meaning of the feature. Because of this, the mean value of this temperature is calculated, and its absolute difference with daily values is calculated. In this way, extreme weather conditions get similar values in the data.

In Figure 23, it can be seen that on the days of winter and summer, we have higher amounts of deviation from the mean of this feature. On the other hand, we see lower amounts in spring and autumn because people do not need to make their living areas warmer or cooler during these periods.

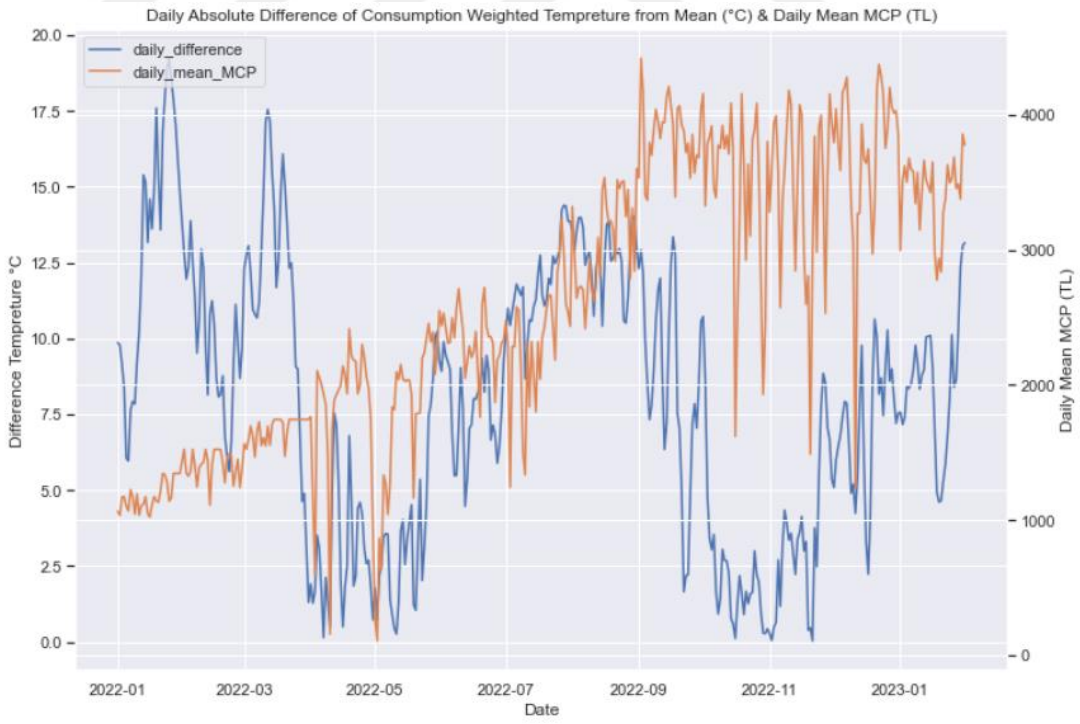

Figure 23. Daily absolute difference of consumption weighted temperature from the mean and daily market clearing price

3.1.8 Wind production weighted wind speed data

Weather information is not limited to temperature details. On www.worldweatheronline.com, we also have wind speed information. However, we also have a similar problem of selecting a city of the wind speed. This problem is also addressed using a methodology similar to the previous feature.

Wind speed information does not affect the consumption side, but it affects the production side with wind-sourced production facilities. Top 10 cities of wind production capacities can be found in Table 2.

Table 2. Top 10 Cities of Wind Production Capacities

City	Live	In Progress	Licensed	Pre-licensed	Total Operation
İzmir	1,757	18	0	0	1,775
Balıkesir	1,359	23	0	0	1,382
Çanakkale	891	202	0	222	1,315
İstanbul	830	68	1	50	948
Kırklareli	394	29	0	475	898
Manisa	702	1	10	0	712
Edirne	204	12	0	295	510
Aydın	401	19	0	84	504
Sivas	150	0	80	250	480
Hatay	428	38	0	0	467

Source: Enerji Atlası, Wind Farm Potential Data by Province (2023)

Using this data for 81 cities, each city's wind-sourced production proportions are calculated. Moreover, using this proportion, wind speed information of each city is weighted, and final wind speed information for the whole country is created.

Equation 2:

$$\sigma_m = \sum_{i=1}^{81} w_i s_i \quad (2)$$

In this equation, i is for each city, w is the proportion of each city in wind-sourced electricity production, s stands for the wind speed of the city in day m, and σ stands for the wind production weighted wind speed information.

After the creation of the final feature, it is directly used in the model. It's relationship with market clearing price can be seen in Figure 24.

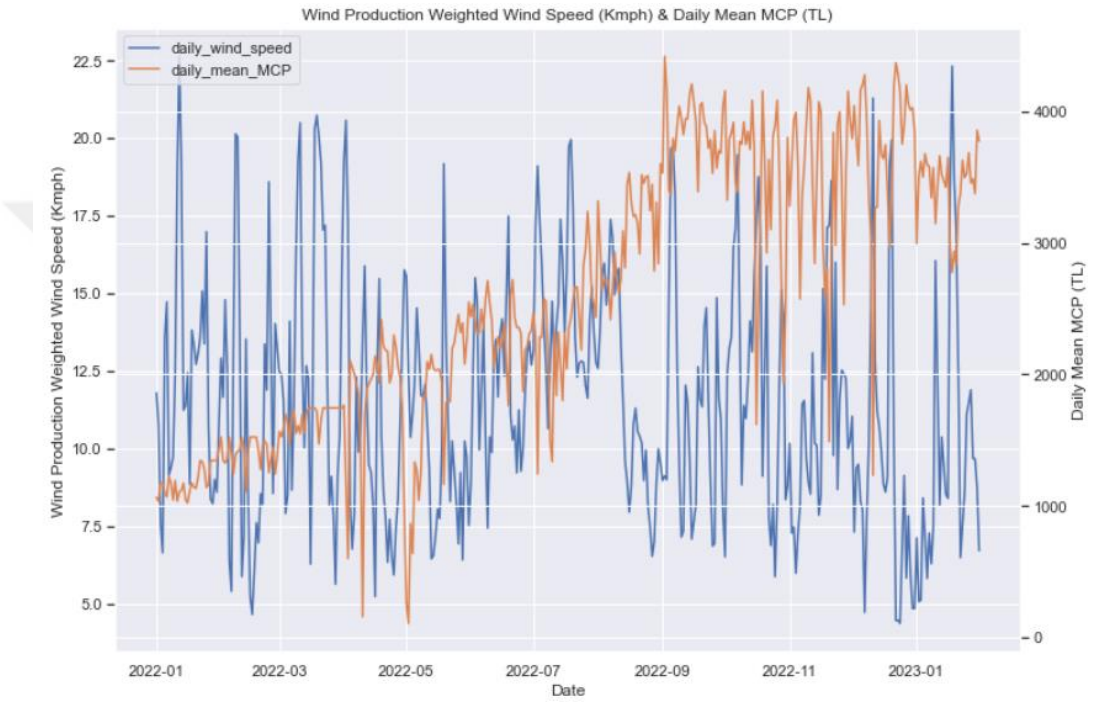

Figure 24. Wind production weighted wind speed and daily mean market clearing price

This data has very high volatility as per the nature of wind speed. It is impossible to see a direct monthly or weekly seasonality in this data, but it effectively covers the impact of wind information for the model.

3.1.9 Solar capacity weighted sun hour information

Cloudiness is also critical for the electricity system because a considerable amount of electricity production comes from solar energy sources. The issue of location-based data is handled using a similar methodology to the previous two features.

A sun hour refers to a metric that quantifies the level of sunlight intensity on the Earth at a specific moment, which can be employed in the generation of solar power, considering variables like climate and weather conditions. Because of this, the solar energy production capacities of each city are used as input to the weighting.

EPDK (EMRA Energy Market Regulatory Authority) announces the unlicensed solar production information in monthly reports. The weighting data is taken from this source and presented in Table 3.

Table 3. Top 10 Cities with Highest Unlicenced Solar Power Generation

Cities	Solar	Proportion
Konya	640.68	0.073420443
Ankara	550.82	0.063122695
Gaziantep	479.76	0.054979384
Şanliurfa	429.33	0.049200223
Kayseri	396.43	0.045429959
İzmir	365.53	0.041888891
Manisa	272.22	0.031195781
Kahramanmaraş	269.81	0.030919601
Afyonkarahisar	264.84	0.03035005
Elazığ	238.17	0.02729373

Source: EMRA, Electricity Market Sector Report (2023)

Using this data for 81 cities, each city's solar-sourced production proportions are calculated. Moreover, using this proportion, the sun hour information of each city is weighted, and the final sun hour information for the whole country is created.

Equation 3:

$$\rho_m = \sum_{i=1}^{81} w_i h_i \quad (3)$$

In this equation, i is for each city, w is the proportion of each city in solar-sourced electricity production, h stands for the sun hour of the city in day m, and ρ stands for the wind production weighted wind speed information.

After the creation of the final feature, it is directly used in the model. In Figure 25, it can be seen that sun hour information is highly seasonal. On summer days, it gets its highest amounts; on winter dayrls, it gets the lowest values.

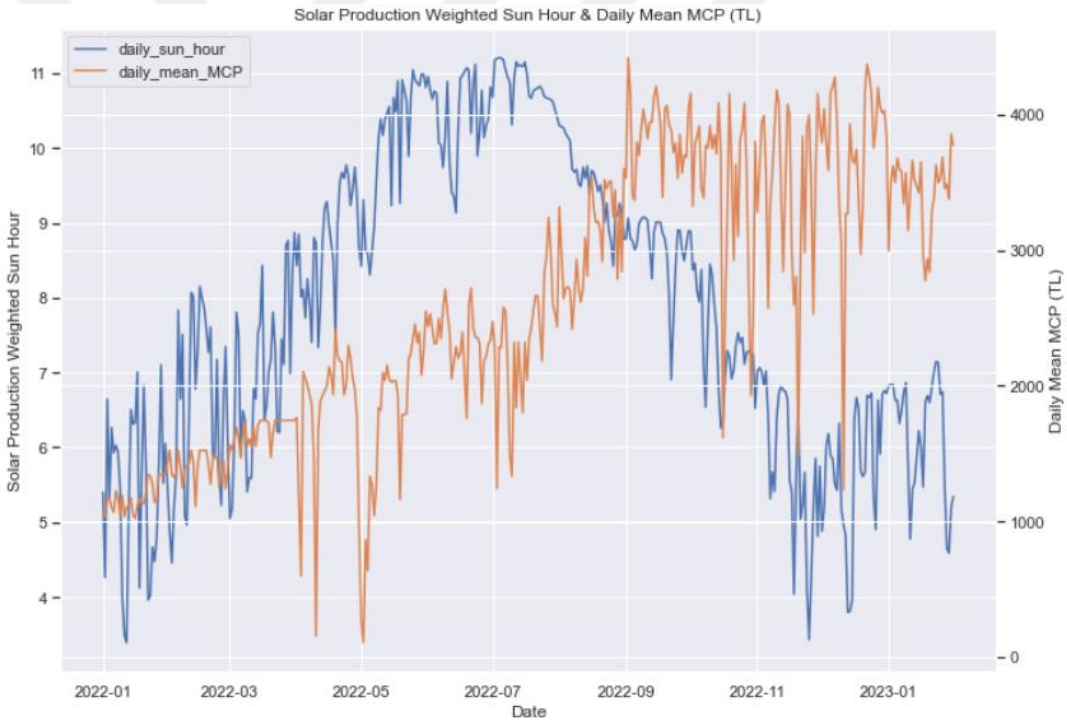

Figure 25. Solar sourced production weighted sun hour and daily mean market clearing price

3.1.11 Price independent sales orders

In the day-ahead electricity market, participants give their offers with price intervals. These intervals start from 0 and end with the ceiling price. The hourly sale offers with 0 prices are called price-independent sales offers. Most of the time, renewable energy production facilities make these offers. The reason behind this is that their cost of production is meager, and they have to sell what they have produced on the market. Also, facilities that are supported by the Renewable Energy Resources Support Mechanism (Yenilenebilir Enerji Kaynaklarını Destekleme Mekanizması YEKDEM) are required to give price-independent sale offers in the day-ahead market.

This is the starting point of the merit order mechanism, and it is very critical for the prediction of MCP. We will have a lower MCP when we have more for this type of offer as can be seen in Figure 26.

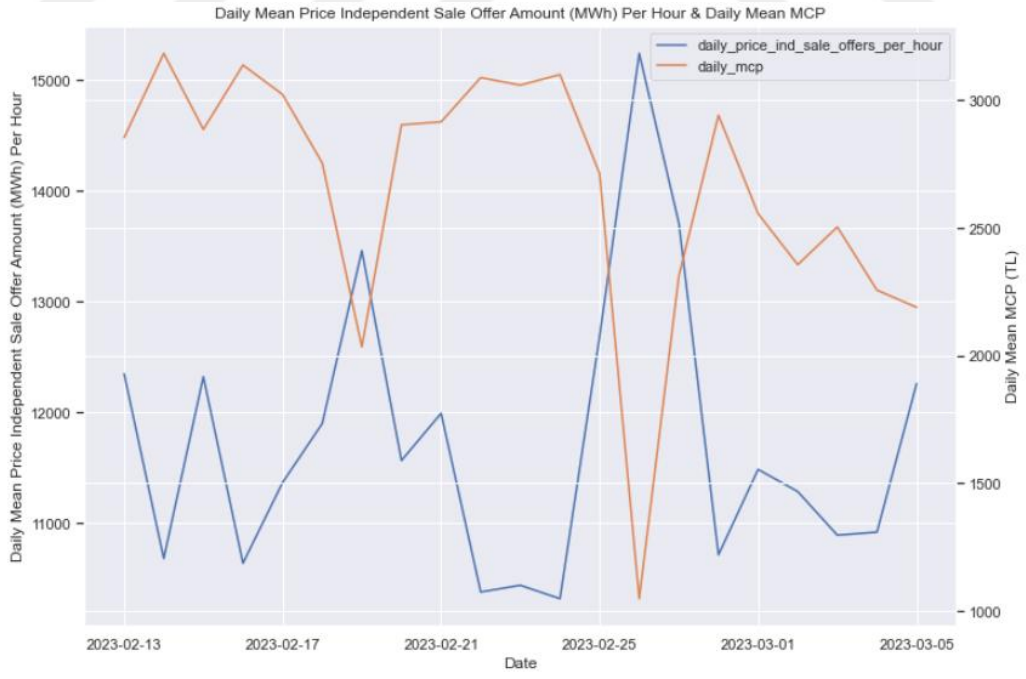

Figure 26. Daily mean price independent sale offer amount and daily mean market clearing price

3.1.12 Price independent bid order

The price-independent bid amount is the sum of the bids submitted hourly in the day ahead market without any price step. These bid orders are given by the market participants who are in critical need of electricity.

As expected, as we see more price-independent bid offers in the market, we will come up with higher clearing prices, please see Figure 27. Because of this, it is vital to have this feature in the model to reflect this participant group's behavior in the market.



Figure 27. Daily mean price independent bid order amount per hour and daily mean market clearing price

3.1.13 Features from supply demand curves

This model aims to understand the impact of ceiling prices on market clearing prices. If the purpose was to predict the market clearing price one day before it is announced, we could not use features from the supply and demand curves. Because

of this, it is critical to include the important points of supply-demand curves in the model.

In supply demand curves, created for each hour of the day, order sets are displayed by adding a block and flexible bids to hourly bids at each price step. The amounts where supply-demand curves start and end are marked with red circles in the Figure 28. The price axis of these points is already known: the price of the starting point of the supply curve and the ending point of the demand curve are with a price of 0. On the other hand, the starting point of demand and the ending point of supply curves are also common: the ceiling price is determined for the day.

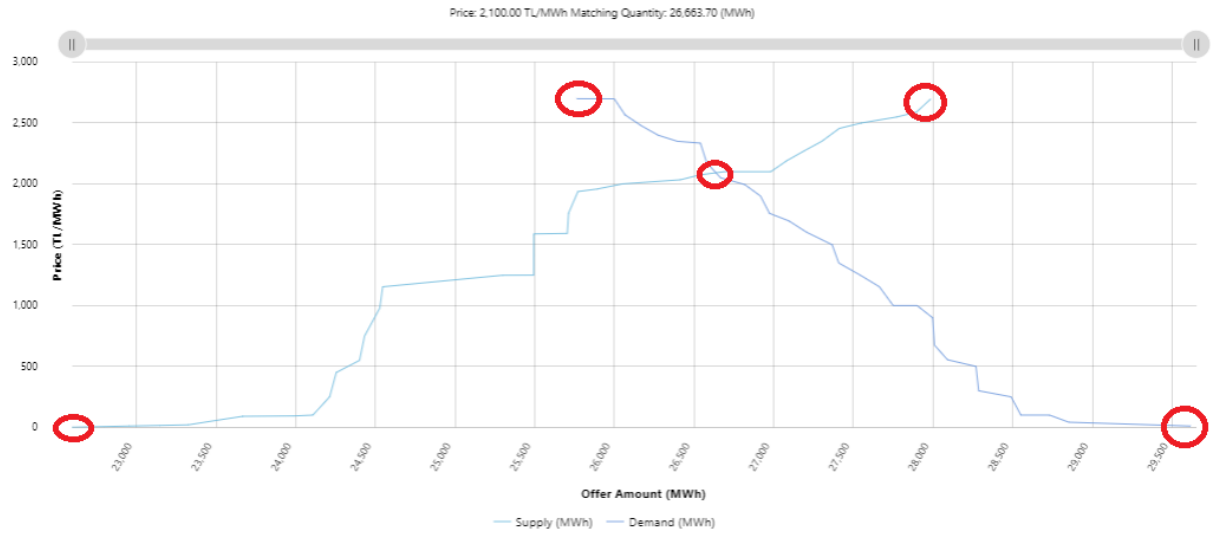

Figure 28. Supply demand curve of 17 July 2023 14.00

They are critical to allowing the model to cover market participant behaviors for a specific hour. Because we know that the intersection of these curves will create the market clearing price of the hour it is important to feed these critical points to the model.

3.1.13.1 Minimum and maximum demand points

As we see more demand in the market, we will see higher prices. This also holds for the starting and ending points of the demand curve, please see Figure 29.

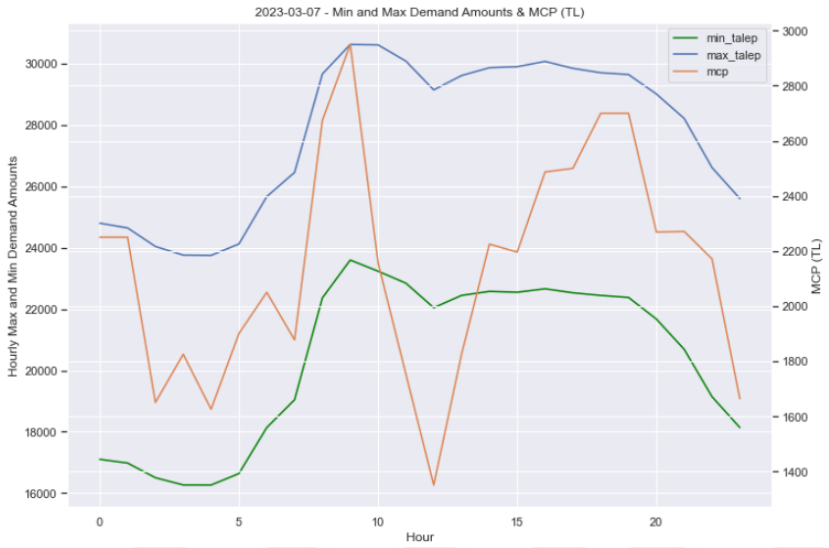

Figure 29 Minimum and maximum demand amounts & market clearing price for a specific date

3.1.13.2 Minimum and maximum supply points

We have a negative correlation between supply points and the MCP, please see Figure 30.

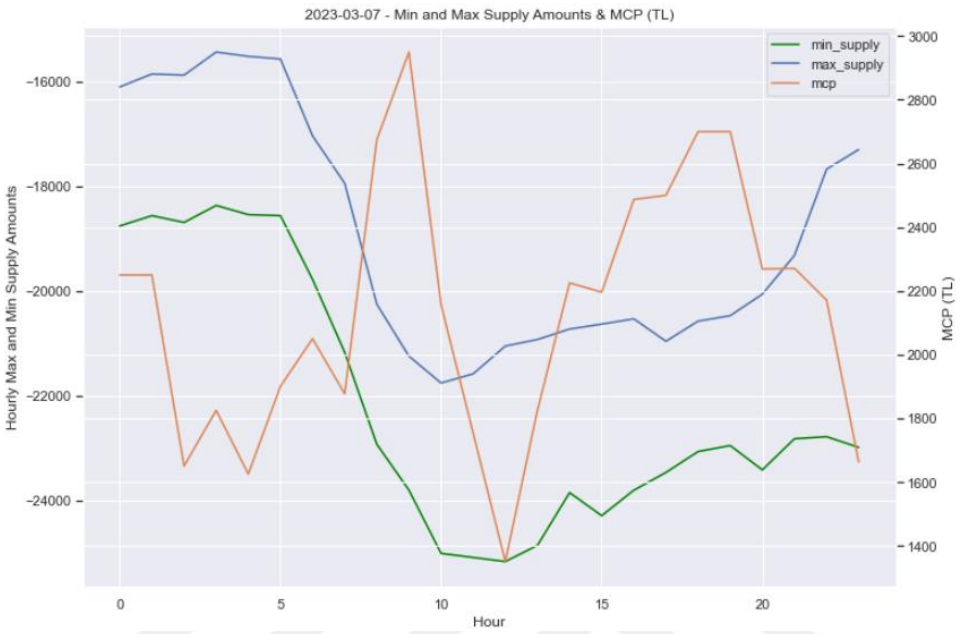

Figure 30 Minimum and maximum supply amounts & market clearing price for a specific date

The intersection point of the demand curve with the supply curve will be between these minimum and maximum supply amounts.

3.1.13.3 Matching amounts

We are trying to predict the price at the intersection of curves. Because of this, another critical point in the supply-demand curve is the amount we see at the intersection. Again, this feature can be applied to models aiming to understand the price drivers. It is impossible to have this amount at hand before the trade date.

We will see higher MCP where we see higher amounts of matching amounts in the market, as it is shown in the Figure 31.



Figure 31 Daily mean matching amounts per hour and daily mean market clearing price

3.1.14 Merit order features from supply demand curves

Merit order is a fundamental concept in the context of electricity markets, serving as a mechanism for determining the dispatch order of available generating units based on their marginal costs. In essence, it represents a hierarchical arrangement of power plants according to their production costs, with the least expensive generators given priority in the dispatch process. The utilization of merit order in market clearing price creation is integral to achieving economic efficiency within the electricity market. As the market operator seeks to satisfy demand at the lowest overall cost, the merit order facilitates the selection of generators to minimize the total production costs required to meet the prevailing demand. Consequently, the market clearing price is established

by the marginal cost of the last unit dispatched, ensuring that all generators receive a price equal to their production costs while providing a transparent and economically rational framework for resource allocation in the electricity market.

Figure 32 sequences renewables, nuclear, coal, and gas facilities according to their production costs per MWh. This hypothetical supply curve shows higher market clearing prices as demand moves toward costly electricity production facilities.

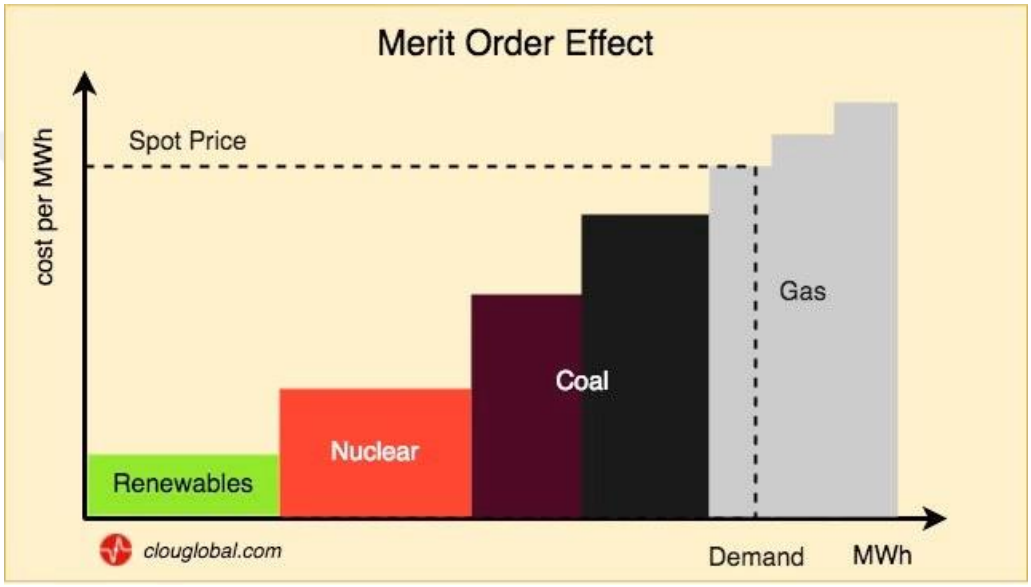

Figure 32. Merit order representation

Source: clouglobal.com

We are seeing a similar step-like pattern in the real supply curve of DAM. The marginal prices of critical energy sources can be used as features for the prediction model because they are quite critical to determining the MCP.

3.1.14.1 Imported coal-fired plants marginal prices

Imported coal-fired plants are quite crucial in the energy mix of Türkiye. Most of the time, regarding marginal costs, they come after renewable and domestic coal plants.

To calculate the marginal cost of one MWh production for coal-fired plants, we first need to calculate the MWh amount that can be generated using 1 Tonne of imported coal. The imported coal type in Türkiye is mostly anthracite type. Because of this, the calorific value of anthracite is used in this calculation. According to The Engineering ToolBox (2010), imported coal has a calorific value of 30,080 kJ/kg. It is making 30,080,000 kJ per Tonne. We have this formula of conversion from kJ to MWh:

- megawatt-hours = kilojoules ÷ 3,600,000

Using this formula, we can create 8.35 MWh per Tonne value.

The formula below calculates the TL cost per MWh created using imported coal.

- TL/MWh = (TL/Tonne) / (MWh/Tonne)

This value is the hypothetical cost of the plant, which has %100 productivity value.

The average productivity value for imported coal plants in Türkiye is assumed to be %40 in this study. Moreover, the marginal cost of the plant is finally calculated with the below formula:

- Marginal cost (TL/Mwh) = Hypothetical Cost (TL/MWh) / Productivity Value

Operation and maintenance costs are not considered in marginal cost calculations in this study. On the other hand, ARA CIF Coal price is used as the spot coal price in this study.

3.1.14.2 Efficient and less efficient natural gas-fired plant marginal costs

In merit order, in most cases, the facility type following the imported coal plants is the natural gas-fired plants.

In Türkiye, natural gas-fired plants purchase their natural gas from BOTAŞ (Pipeline Petroleum Transportation Joint Stock Company) with the announced tariff each month. A specific tariff is applied for natural gas-fired electricity generation facilities in the tariff table; this value is used for the marginal cost calculation.

In this calculation, operation and maintenance costs are not considered again.

The marginal cost of natural gas plants can be calculated using the following:

- Marginal Cost (TL/MWh) = BOTAŞ Price (TL/MWh) / Productivity

There are two main types of natural gas-fired plants in Türkiye. New technology effective ones have a productivity of %57; other standard production facilities have a productivity of %41. These productivity values calculate marginal costs for both effective and non-effective natural gas-fired plants.

Figure 33 plots the marginal cost of imported coal, efficient natural gas, and non-efficient natural gas-fired facilities' marginal costs.

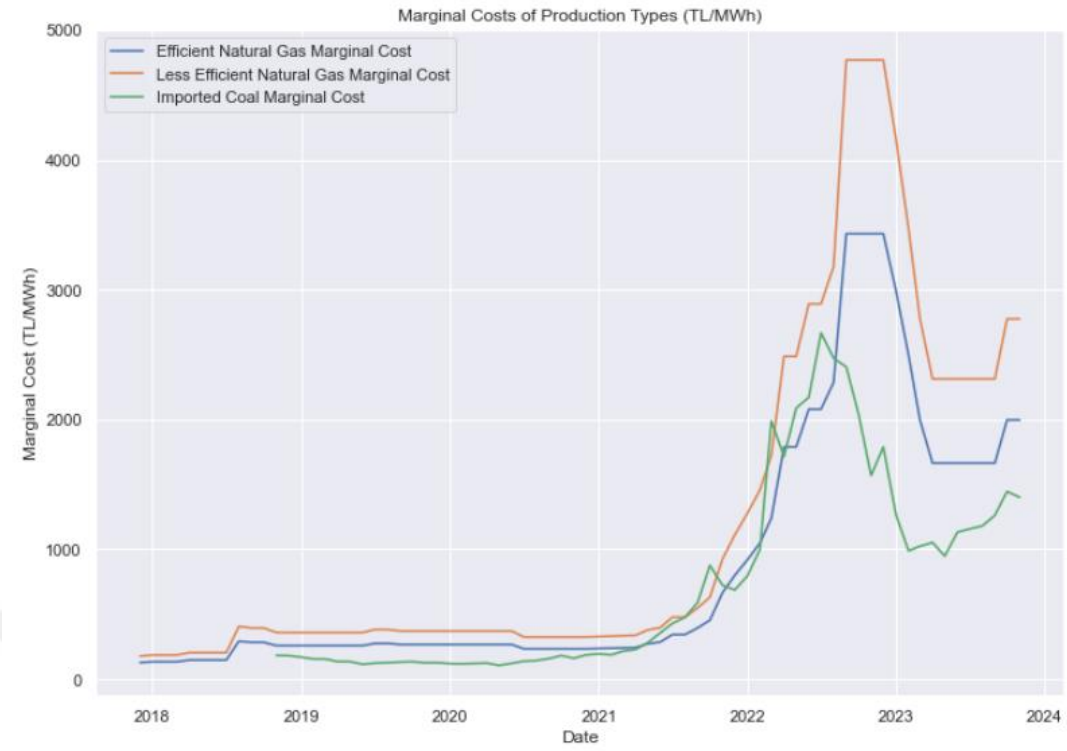

Figure 33 Marginal costs of production types

3.1.14.3 Supply amounts of types of production facilities

After generating the marginal costs of coal and natural gas-fired production facilities, we can get the supply amounts given as sales offers in the market. For instance, let us consider that the marginal cost of less effective natural gas plants is greater than that of effective natural gas plants, and the marginal cost of effective natural gas plants is greater than imported coal-fired plants. This is the most common case in marginal cost data.

- Supply amount between the price of 0 and the marginal cost of imported coal-fired plants will be a supply of facilities less costly than imported coal; these can be renewables and domestic coal-fired plants.
- The amount between the marginal cost of imported coal-fired plants and effective natural gas plants will be the supply amount from coal-fired plants.

- The difference between effective natural gas plants and less effective natural gas plants will be the amount that effective natural gas plants can supply.
- The amount above less effective natural gas plants includes the supply of less effective natural gas plants and facilities that are costlier than less effective natural gas plants.

The graph above shows that the marginal cost orders may change occasionally. For instance, although in most cases we have the order of less effective natural gas > effective natural gas and imported coal-fired plants, we see differences in these prices between mid-2021 and mid-2022. To avoid negative amounts or unmeaningful comparisons, for every single hour,

- The supply difference between 0 and the first marginal cost
- The supply difference between first and second marginal cost
- The supply difference between the second and third marginal cost
- The supply amount greater than the third marginal cost

are calculated and used in the model.

3.1.15 Ceiling price of the day information

This study aims to understand the impact of ceiling prices on electricity prices. Because of this, it is quite important to include information on the ceiling price of the day in the model. This information is mined from the market announcements in the Transparency Platform. It is determined periodically by the Energy Market Regulatory Authority (EMRA) in Türkiye. The relationship between ceiling prices and MCP can be seen in Figure 34.

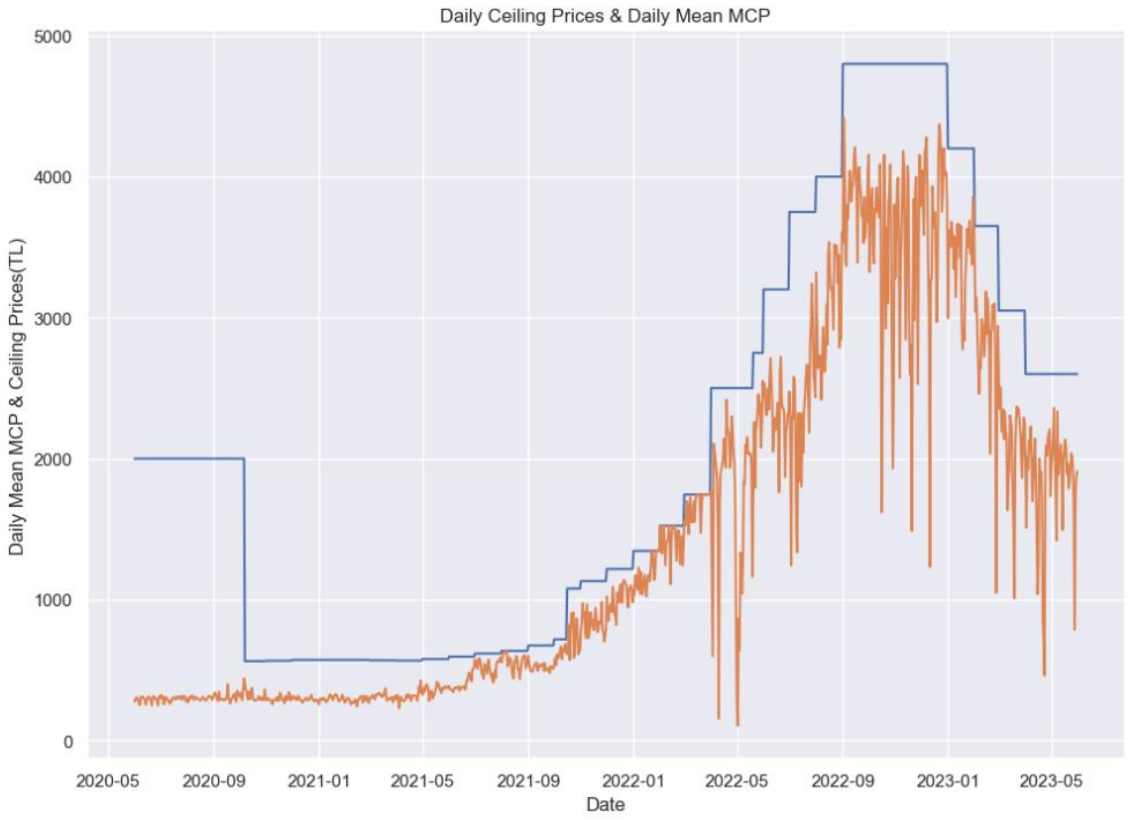

Figure 34. Daily ceiling prices and daily mean market clearing price

In Figure 34, mean daily MCP rarely touches the daily ceiling price; this happens when we see the situation where MCP equals ceiling prices at all hours of the day. The ceiling price is generally touched during the peak hours of the day because of high demand. In night hours, for instance, the ceiling price is rarely touched.

In Figure 35, the number of hours that MCP equals to the ceiling price for each day is shown. It can be seen that the ceiling price became quite effective in the market after July 2021, when we saw an energy crisis in the post-pandemic period. Before this period, the ceiling price was rarely touched and, therefore, was less effective.

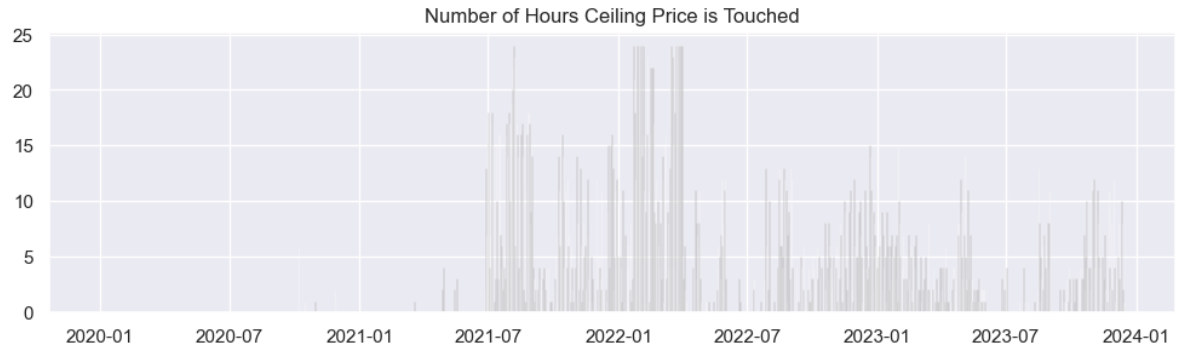

Figure 35. Number of hours ceiling price is touched

CHAPTER 4:

RESULTS

4.1 Preprocessing

The data pulled from different data sources are joined into one data frame, which has one hour in each row of the data frame. Some features like USD/TRY currency and consumption-weighted daily temperature are not hourly in data; these features are duplicated for each hour of the day. On the other hand, data pulled from these sources have very few missing values; because of this, there was no need to apply missing data handling techniques; for some edge cases where we do not have any data in sources, we dropped the whole row from the data frame. There were not many outliers in the data, so there was no need to handle the outliers differently.

On the other hand, categorical features are separated from other numeric features and marked as categorical features before integrating them into the LightGBM model.

4.2 Train-test split

This model aims to predict the market clearing price as accurately as possible by integrating ceiling price as a feature to the model to understand the impact of this feature on market clearing price; because this is one of the most recent periods that we see ceiling price announcement is selected as a test period. On the other hand, we included two years of data for the training period of the data. As a result, 2021-07-05 till 2023-07-04 is selected as the training period of our model. 2023-07-05 till 2023-07-31 is selected for the test period of our model. Please note that from 2023-07-04, the ceiling price in the day-ahead electricity market changed from 2600 TL/MWh to

2700 TL/MWh. Because we want to do some sensitivity analysis on the ceiling price feature, this period has been selected as the test period for our model.

4.3 Correlation check

Before finalizing the input data frame, the correlation of inputs is checked. The correlation matrix that is created with the final input data can be seen in Figure 36.

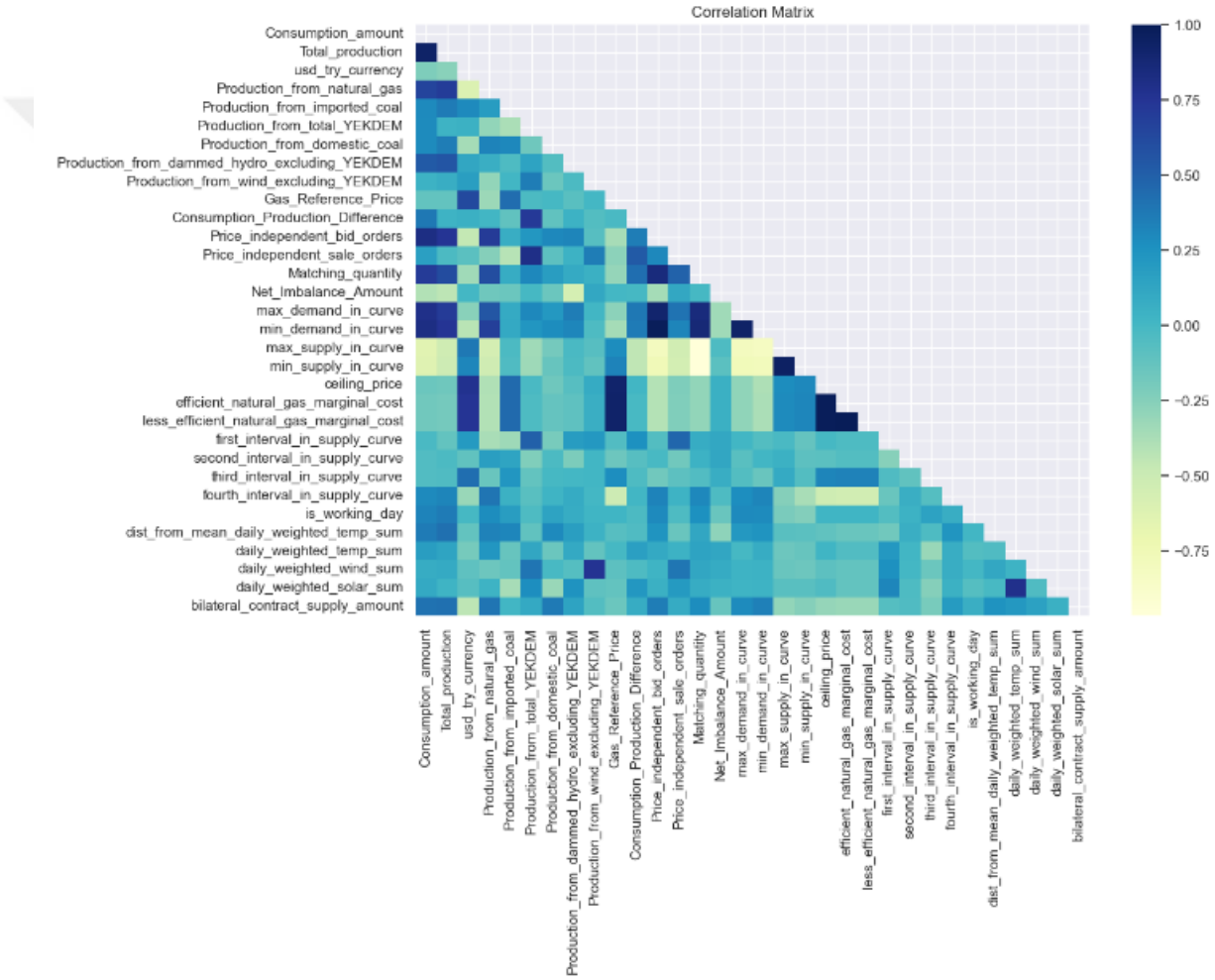

Figure 36. Correlation matrix of the model

4.4 Selected model and hyperparameters

In the pursuit of accurately predicting electricity market clearing prices and comprehensively understanding the impact of ceiling prices, choosing an appropriate machine learning algorithm plays a pivotal role. LightGBM was selected as the primary modeling method in this study due to its well-established capabilities in handling large-scale datasets, high-dimensional feature spaces, and complex relationships within the data. LightGBM, as a gradient boosting framework, is adept at handling non-linear relationships and capturing intricate patterns in electricity market data. Its efficiency in training speed and memory utilization proved advantageous, especially given the electricity market data's dynamic and intricate nature. Furthermore, LightGBM's ability to handle categorical variables and inherent support for dealing with imbalanced datasets ensured a robust modeling approach, essential for accurate predictions in complex market clearing prices subject to ceiling constraints. By leveraging the strengths of LightGBM, this research aims to contribute valuable insights into the nuanced dynamics of electricity markets, shedding light on the implications of ceiling prices on market clearing price predictions.

LightGBM operates as a gradient boosting framework, utilizing an ensemble of decision trees to make accurate predictions on complex datasets. Unlike traditional decision tree algorithms, LightGBM employs a novel approach known as 'leaf-wise' tree growth, which involves growing the tree vertically and adding nodes to the leaves with the highest information gain. This method accelerates the training process, making LightGBM particularly efficient for large datasets and high-dimensional feature spaces. Additionally, LightGBM employs a histogram-based learning method, which discretizes continuous features into discrete bins, reducing

memory usage and computational complexity. Its ability to handle categorical features without needing one-hot encoding further increases efficiency. Furthermore, LightGBM implements a 'Gradient-based One-Side Sampling' strategy to address imbalanced datasets, enhancing its robustness in predicting outcomes in scenarios where certain classes may be underrepresented. By combining these innovative techniques, LightGBM emerges as a powerful tool for predictive modeling that is well-suited for capturing intricate patterns and relationships within the context of electricity market data.

The selected hyperparameters for the LightGBM model have been meticulously tuned through an extensive grid search to optimize predictive performance for the specific task of electricity market clearing price prediction. The 'objective' parameter is set to 'regression,' indicating the nature of the problem as a regression task where the goal is to predict continuous values. The choice of the 'mse' (Mean Squared Error) metric aligns with the regression objective, serving as the optimization criterion to minimize the squared differences between predicted and actual clearing prices. The boosting strategy is specified as 'gbdt' (Gradient Boosting Decision Tree), a method well-suited for capturing complex relationships in the data. The 'colsample_bytree' parameter is set to 0.8, controlling the fraction of features randomly selected to build each tree and promoting diversity among the individual trees. A learning rate of 0.1 regulates the step size in the optimization process, striking a balance between convergence speed and avoiding overshooting. A maximum depth of 7 levels and 200 estimators contribute to the model's complexity, ensuring a nuanced representation of the underlying patterns in the electricity market data. The 'num_leaves' parameter is set to 100, determining the maximum number of leaves in each tree and influencing the tree's capacity to capture fine-grained patterns.

Lastly, a 'subsample' of 0.8 ensures that each boosting iteration is trained on a random subset of the data, enhancing the model's generalizability. Together, these hyperparameters collectively tailor the LightGBM model to the intricacies of the electricity market, striking a balance between complexity, predictive accuracy, and computational efficiency.

4.5 Prediction on test period

In assessing the predictive performance of the LightGBM model, which is trained using two years of data before the test period from 5 July 2023 to 31 July 2023, notable outcomes underscore the model's efficacy in forecasting electricity market clearing prices. The R-squared (R2) score, a measure of the model's ability to explain the variance in the target variable, yielded a commendable value of 0.8009. This high R2 score signifies the substantial variability in clearing prices that the model successfully captures. Furthermore, the Mean Absolute Percentage Error (MAPE) was calculated at 0.0704, indicating a low average percentage deviation between predicted and observed prices. The Root Mean Squared Error (RMSE) further reinforces the model's accuracy, with a value of 165.38, reflecting the square root of the average squared differences between predicted and actual clearing prices. These results collectively highlight the model's robust performance, offering a precise and reliable framework for predicting electricity market clearing prices within the specified temporal context. The demonstrated effectiveness of the LightGBM model, as evidenced by these metrics, contributes substantively to the understanding of market dynamics and bolsters its applicability in the nuanced domain of electricity market forecasting. The actual values of MCP and predicted values are plotted in a scatter plot to show how the model is performing in Figure 37.



Figure 37. Expected and Predicted values with the model proposed

It can be seen that the regression line that we draw according to the predicted and expected values of PTF is near the x=y axis, meaning that our predictions are close to the actual values in most cases.

4.6 Feature importance plots

After building the model, we can get the feature importance plots of the models. Figure 38 is the split-based feature importance of the model:

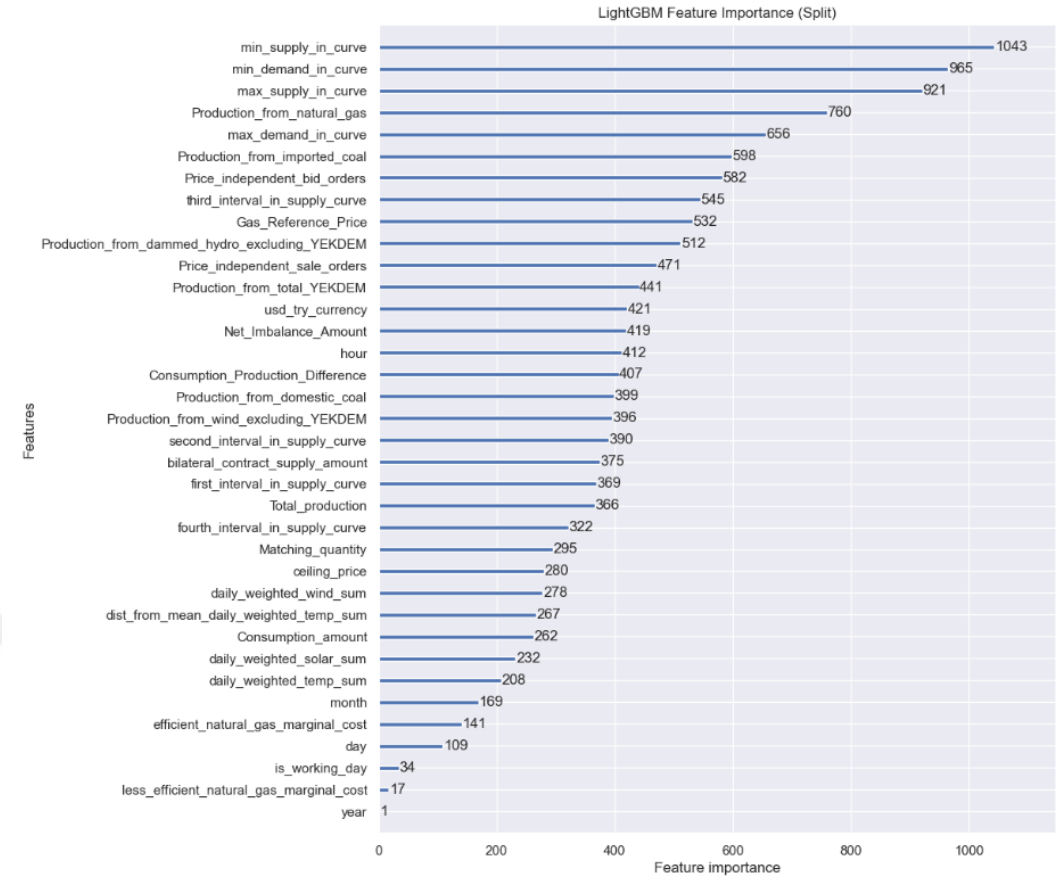

Figure 38. Split-based feature importance plot

Split-based and gain-based feature importance is essential for understanding the significance of input variables in machine learning models. Split-based importance measures the frequency with which a feature is utilized to divide the dataset during the construction of decision trees. Features with higher split-based importance are recurrently employed for partitioning the data, suggesting their pivotal role in the model's decision-making process. On the other hand, gain-based importance assesses the improvement in model performance attributed to each feature, often quantified by metrics such as information gain or Gini impurity reduction. A feature with high gain-based importance contributes significantly to the model's ability to accurately distinguish between classes or predict the target

variable. The differences between these two types of importance plots can offer valuable insights into the model's behavior: a feature might be frequently used for splitting. However, it may not necessarily contribute substantially to the model's overall predictive power, highlighting the nuanced interplay between feature relevance and performance enhancement in machine learning models.

In split-based gain importance, we see important points of the demand curve, production amounts from various sources, price-independent order amounts, and important reference prices included in higher ranks of feature importance. This plot does not accumulate the feature distribution in a single feature.

On the other hand, Figure 39 is the gain-based importance of the model:

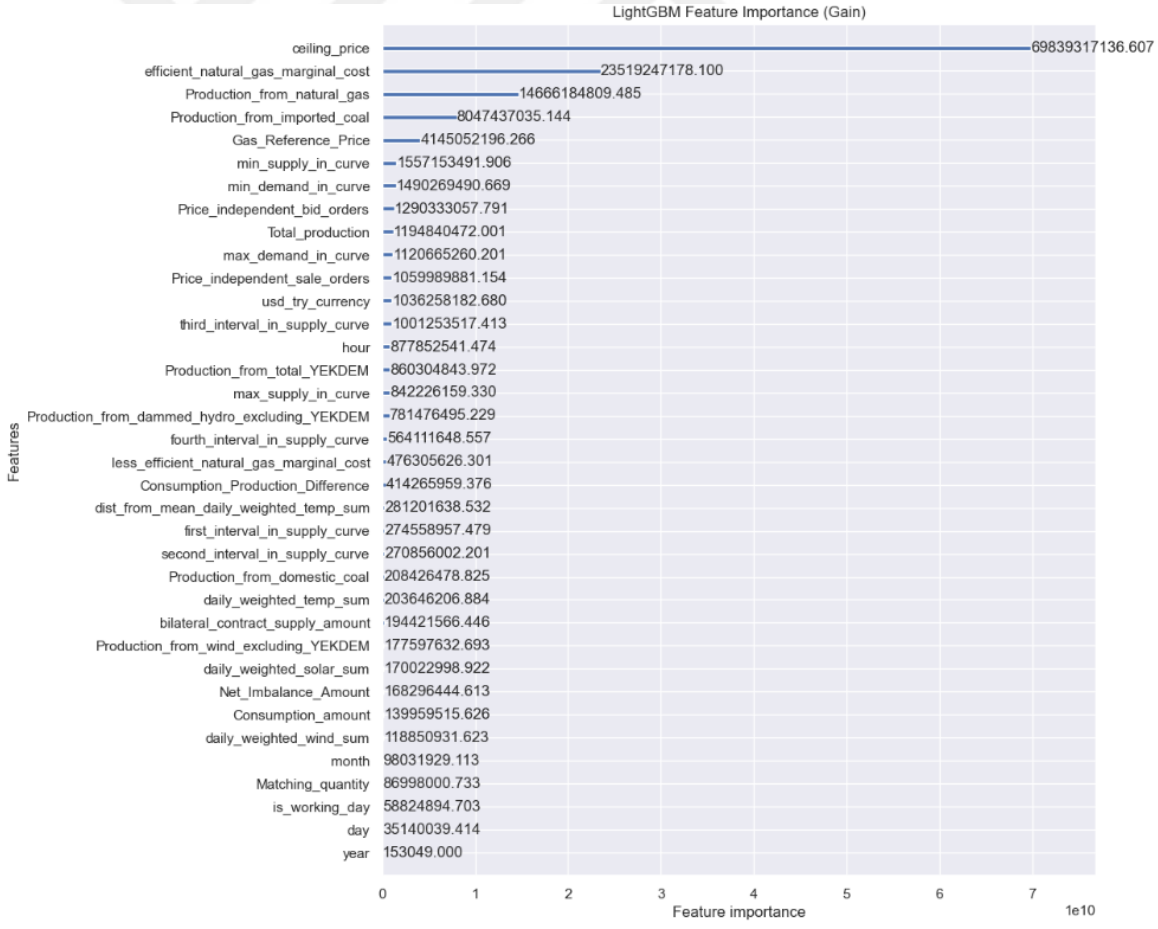

Figure 39. Gain-based feature importance plot

We see a very high dependence on ceiling price information in gain-based importance. That means that although the ceiling price is determined daily, its contribution to the model performance is significant. The ceiling price is followed by the marginal costs and production amount of natural gas sourced. These following features were common in split-based feature importance, too.

4.7 SHAP analysis

The high importance of the ceiling price is interesting because it does not change when we make the prediction; it is fixed at 2700 during this period. To understand the impact of each feature on the model prediction, we used the SHAP (SHapley Additive exPlanations) package in Python.

In the realm of interpretable machine learning, SHAP (SHapley Additive exPlanations) values have emerged as a powerful tool to elucidate the decision-making process of complex models. Originating from cooperative game theory, SHAP values assign each feature an importance value for a particular prediction, thereby providing a measure of the contribution of each feature to the output of the model (Lundberg & Lee, 2017). The methodology behind SHAP is grounded in the Shapley value—a concept from game theory that offers a fair distribution of payouts to players based on their marginal contributions to the total payout.

Figure 40 is the summary plot that we created using SHAP:

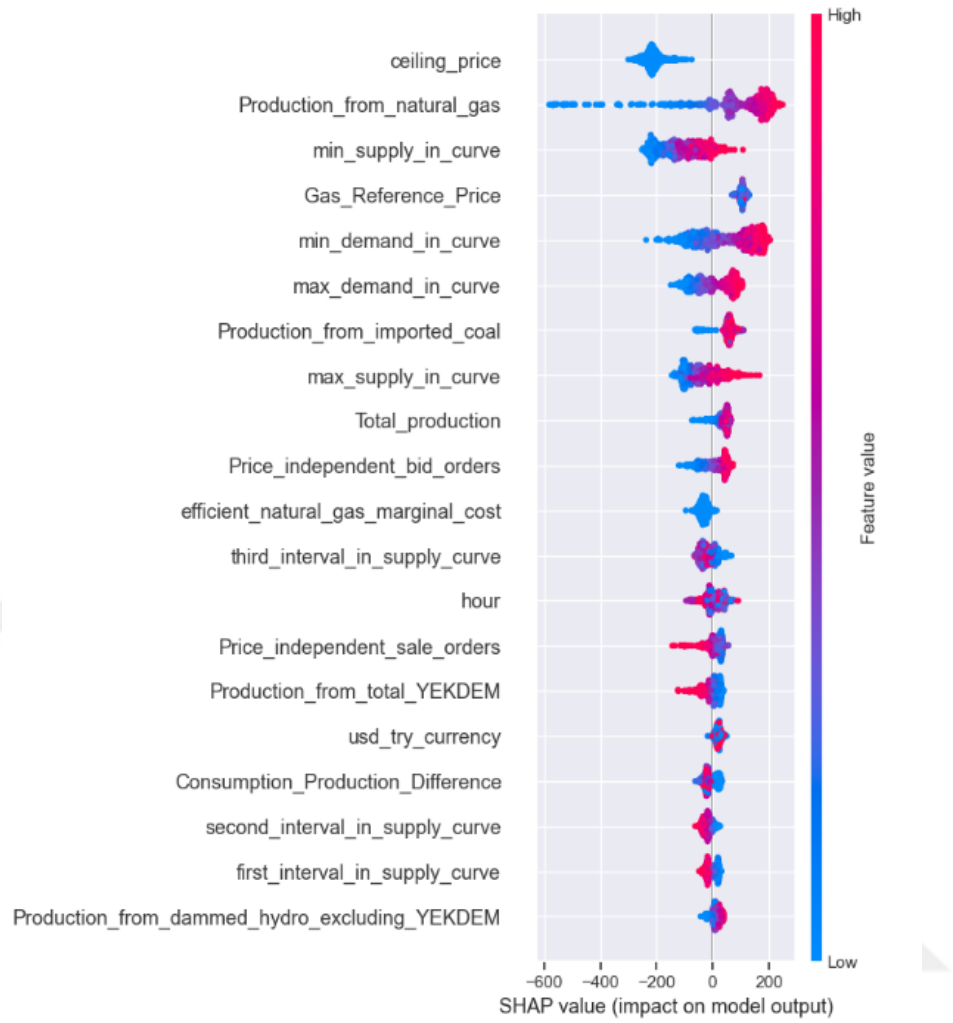

Figure 40. SHAP Summary Plot

SHAP values are particularly insightful because they account for the interaction effects between features, offering a granular and accurate depiction of feature attributions. This is achieved by computing the predictions with and without the feature across all possible combinations of features and then averaging these marginal contributions. The result is a consistent and locally accurate additive feature attribution method that satisfies properties of local accuracy, missingness, and consistency—criteria essential for reliable interpretations.

In a SHAP summary plot, each dot represents the SHAP value for a feature and an individual prediction, with the color indicating the feature’s value (high or low) and the position along the horizontal axis representing the magnitude and direction (positive or negative) of the feature’s impact on the model's prediction. Aggregating these points for each feature across all instances provides an interpretable snapshot of the feature's predictive importance and distribution of effects.

In our case, with ceiling price as a feature, we see that it has negative SHAP values. Therefore, the impact of this feature on the output label is negative. The feature values are also blue, indicating that as it has lower values it is decreasing the value of output MCP. On the other hand, the second important feature in the list, which is production from natural gas, has a comparatively scattered distribution. When it gets lower values, it decreases the MCP; when it has higher values, it increases the output feature.

When we consider another example of production from YEKDEM-supported facilities, we see that when it gets higher values, it decreases the output feature; when it gets lower values, it increases the output.

In pursuing a more interpretable machine learning model, SHAP (SHapley Additive exPlanations) dependency plots provide an insightful visualization of how individual feature values affect the model's predictions. The plot below illustrates the relationship between the 'ceiling_price' feature and the model's output, revealing the impact of the feature across its value spectrum. Each point on the plot corresponds to a 'ceiling_price' value for a particular data instance, with the SHAP value on the y-axis quantifying the influence of that feature value on the model's prediction—

positive SHAP values signify an increase. In contrast, negative values denote a decrease in the predicted outcome.



Figure 41. Shap dependency plot for ceiling price and the target variable

The color gradient represents the magnitude of the target variable, adding an additional dimension to the analysis by indicating how the feature value relates to the predicted outcome. A pattern discernible from the plot suggests that higher ceiling price values generally contribute to an increase in the model's predicted value, indicative of a positive correlation. This pattern is not just a simple linear relationship; the vertical dispersion of SHAP values at individual feature values suggests variability influenced by the interaction with other features.

For more granular level analysis, SHAP force plots are commonly used. Figure 42 is an example that we selected one of the predictions of the model:



Figure 42. First example of SHAP force plot

The SHAP force plot provides a compelling visual representation of the individual contributions of features to a specific prediction by a machine learning model. In the given plot, the base value, representing the mean prediction over the dataset, is established as a reference point from which deviations, influenced by feature values, are measured. The resultant predictive output is displayed to the right, indicating the cumulative effect after accounting for the individual SHAP values. Features exerting a positive influence, such as production from imported coal and natural gas, as well as gas reference price, are denoted in red, with their respective SHAP values indicating the magnitude of their contribution to an increase in the prediction. Conversely, features that apply downward pressure on the prediction, ceiling price, and minimum and maximum demand in curves are represented in blue, highlighting their significant negative impact. The juxtaposition of these forces vividly demonstrates the dynamic interplay of features that the model considers, underpinning the final prediction, which, in this instance, is lower than the base value due to the preponderance of negative contributions. This intricate dance of features, made accessible by the SHAP force plot, underscores the complex nature of model decision-making, offering a transparent lens through which the model's reliance on

specific features can be scrutinized and interpreted within an academic or applied research context.

Figure 43 is another example of SHAP force plots from our training data. It can be seen that this time, the predicted price is near the ceiling price. Although this is the case, we see ceiling price information as the first decreasing factor in the model prediction. On the other hand, we have increasing factors such as total production, production from imported coal, maximum demand in a curve, gas reference price, minimum demand in a curve, and production from natural gas.

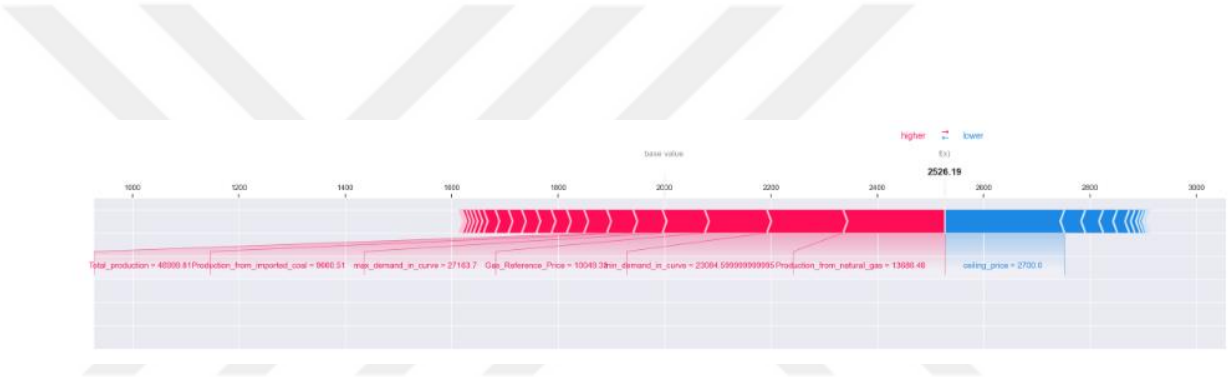

Figure 43. Second example of SHAP force plot

It is seen that this time, minimum and maximum demand values are included in increasing factors. This example makes reading the summary plot of SHAP easier. In most examples, the ceiling price is on the decreasing side of force plots. This is also seen in the summary plot with all blue values in the negative SHAP value area. On the other hand, we can see minimum and maximum demand values on both sides of the force plot. For instance, in the example that we have a lower predicted value, the maximum demand is 21,746.4, and the minimum demand is 18,301.8; in the higher predicted value example, we have the maximum demand as 27,163.7, and the minimum demand is 23,068.59. We see that the corresponding values are quite

different in these examples. These features have blue values in the negative SHAP value area and red values in the positive SHAP value area, making these features positively correlated with the output feature.

4.8 Tobit model built and analyze on determinants of market clearing price

Censored data is incomplete information commonly seen in various real-world scenarios. Imagine, for instance, a study on how much people spend watching TV daily. If the study only records times up to a maximum of 4 hours, any amount beyond 4 hours is unknown – this is an example of right-censoring. Conversely, if the study only records times above 30 minutes, any viewing time less than 30 minutes is unknown, which is left-censoring.

Tobit models are employed to analyze such censored data effectively. These models are specially designed to handle situations where the full range of data is not observed due to censoring. Consider a survey on household incomes in a city, where respondents are only asked to specify if their income falls in specific brackets, like "less than $20,000" or "more than $100,000". The exact incomes for many households are unknown, especially for those earning below $20,000 or above $100,000. A Tobit model can be used here to estimate the likely distribution of incomes across the entire city despite this censoring. It does so by considering both the partial information available (the income brackets) and the pattern of censoring (incomes beyond the specified brackets), thus providing a more accurate analysis than traditional methods, which might inaccurately assume complete data. Tobit models are, therefore, a valuable tool in fields like sociology or market research, where censored data is frequently encountered.

In our case, MCP data can also be considered censored because it has a ceiling price that makes it impossible to cross the threshold. In this section, the impact of ceiling prices on the other determinants of market clearing price will be analyzed using the Tobit model.

In the importance plots of the LightGBM model, we have seen that the features from the supply-demand curve are quite crucial for predicting market clearing prices. Because of this, we have selected these features to construct the Tobit model, which requires careful selection of features because it is a linear regression-based model.

In the matrix in Figure 44, the features used in this model and the correlations of these features within each other can be found:

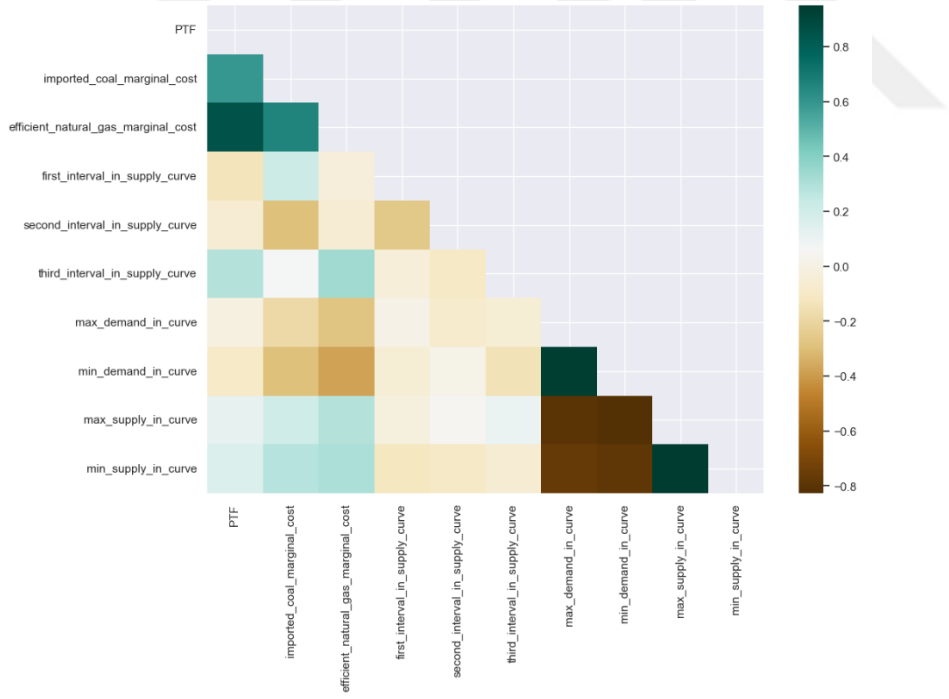

Figure 44. Tobit model correlation matrix

It can be seen that supply and demand amounts have a negative correlation; on the other hand, min and max values of supply and demand have strong correlations.

The py4etrics package's Tobit module is used to fit the Tobit model. Left or right censors are given statically in many packages to the Tobit model. With the py4etrics package, users can give the model in which instances the censored value is touched. This is important for our case because we have different ceiling prices in different periods. The censored flag reflects the hours when the ceiling price is touched; exogenous and endogenous factors are also fed to the model.

The intercept of the regression line is created by adding an intercept variable that takes the value of 1 statically in the model.

As a result of this, the model described in Table 4 is created.

Table 4. Tobit Regression Model Results

Tobit Regression Results

Dep. Variable:	PTF	**Pseudo R-squ:**	0.121
Method:	Maximum Likelihood	**Log-Likelihood:**	-102963.1
No. Observations:	17544	**LL-Null:**	-117179.3
No. Uncensored Obs:	13812	**LL-Ratio:**	28432.4
No. Left-censored Obs:	0	**LLR p-value:**	0.000
No. Right-censored Obs:	3732	**AIC:**	205946.2
Df Residuals:	17534	**BIC:**	206024.0
Df Model:	9	**Covariance Type:**	nonrobust

	coef	std err	z	P>|z|	[0.025	0.975]
intercept	-909.8013	33.026	-27.548	0.000	-974.532	-845.071
imported_coal_marginal_cost	0.3230	0.008	38.432	0.000	0.307	0.339
efficient_natural_gas_marginal_cost	0.8344	0.006	137.394	0.000	0.823	0.846
first_interval_in_supply_curve	-0.2919	0.009	-30.748	0.000	-0.311	-0.273
second_interval_in_supply_curve	-0.1083	0.010	-10.633	0.000	-0.128	-0.088
third_interval_in_supply_curve	0.1435	0.011	13.183	0.000	0.122	0.165
max_demand_in_curve	0.0471	0.003	16.111	0.000	0.041	0.053
min_demand_in_curve	0.4177	0.005	87.653	0.000	0.408	0.427
max_supply_in_curve	0.2670	0.009	30.354	0.000	0.250	0.284
min_supply_in_curve	0.1575	0.009	17.942	0.000	0.140	0.175
Log(Sigma)	6.0354	0.006	1003.955	0.000	6.024	6.047

It is seen that all of the variables provided to the model are significant by checking the p-value of each coefficient.

In this example, a linear regression model with the impact of censors is created. To cover the impact of ceiling prices in the regression model, we have built another linear regression model without the censoring impact. The intercept point of the regression line is handled in the same way as the Tobit model, please see Table 5.

Table 5. Standard Linear Regression Model Results

OLS Regression Results

Dep. Variable:	PTF	**R-squared:**	0.813
Model:	OLS	**Adj. R-squared:**	0.813
Method:	Least Squares	**F-statistic:**	8448.
Date:	Tue, 09 Jan 2024	**Prob (F-statistic):**	0.00
Time:	13:11:33	**Log-Likelihood:**	-1.3494e+05
No. Observations:	17544	**AIC:**	2.699e+05
Df Residuals:	17534	**BIC:**	2.700e+05
Df Model:	9		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
intercept	-1666.4836	36.190	-46.048	0.000	-1737.420	-1595.547
imported_coal_marginal_cost	0.2417	0.009	25.872	0.000	0.223	0.260
efficient_natural_gas_marginal_cost	1.0230	0.006	159.484	0.000	1.010	1.036
first_interval_in_supply_curve	-0.2335	0.011	-21.605	0.000	-0.255	-0.212
second_interval_in_supply_curve	-0.0228	0.011	-2.036	0.042	-0.045	-0.001
third_interval_in_supply_curve	0.0313	0.012	2.579	0.010	0.008	0.055
max_demand_in_curve	0.0368	0.003	10.820	0.000	0.030	0.043
min_demand_in_curve	0.1105	0.004	29.945	0.000	0.103	0.118
max_supply_in_curve	0.0952	0.010	9.879	0.000	0.076	0.114
min_supply_in_curve	-0.0127	0.009	-1.363	0.173	-0.031	0.006

Omnibus:	4862.987	**Durbin-Watson:**	0.471
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	18521.373
Skew:	-1.346	**Prob(JB):**	0.00
Kurtosis:	7.253	**Cond. No.**	4.04e+05

All of the features provided in the OLS model, except minimum supply in the curve, are also significant. In our linear regression analysis context, the coefficient of determination, commonly referred to as the R-squared value, is reported to be 0.813. This statistical measure is pivotal in interpreting the proportion of variance in the dependent variable that can be attributed to our model's independent variable(s). An R-squared value of 0.813 indicates that the predictors included in the model explain approximately 81.3% of the variation in our outcome variable. This high value

suggests a strong fit between the model and the observed data, implying that the independent variables selected significantly contribute to the variation observed in the dependent variable. However, it is essential to note that while a high R-squared value points to a model that accounts for a significant portion of the variance, it does not necessarily imply causation.

It can be seen that the coefficients of both models are quite different from each other. Comparing how the OLS and Tobit regression models predict the dependent variable (MCP) across the range of each independent variable can be achieved using the following:

1. Determining the minimum and maximum values of each independent variable used in the model.
2. Creating a sequence of x values from minimum and maximum values of independent variables, 100 points are selected in our example as a balance between having enough points for a smooth line in the plot and computational efficiency.
3. Computing the predicted y-values for the Ordinary Least Squares (OLS) regression model.

$$y_{ols} = k_{\,OLS\ intercept} + x * \beta_{\substack{selected\ OLS \\ feature}}$$

It uses the model's intercept and the coefficient for the current feature multiplied by the x-values. This essentially applies the linear equation y=intercept+slope×x for the OLS model.

4. Calculating Predicted Y-Values for the Tobit Model

$$y_{Tobit} = k_{Tobit\ intercept} + x * \beta_{\substack{selected\ Tobit \\ feature}}$$

Similar to the OLS calculation, this computes the predicted y-values for the Tobit model using its intercept and the coefficient for the current feature.

Figure 45 shows that the slope of curves is changed for most of the features used in the model. For instance, the ceiling price and second interval in supply curve features have very different slopes between Tobit and OLS models.

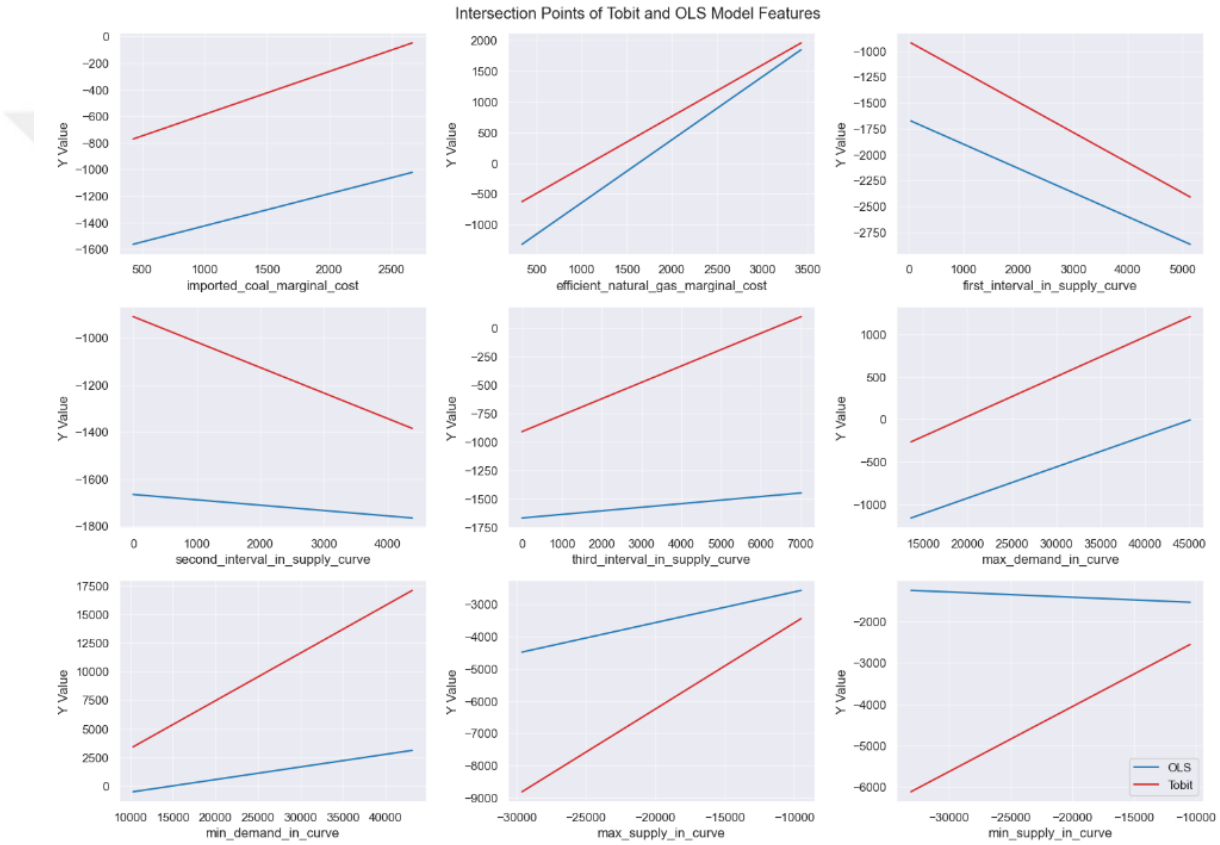

Figure 45. Intersection points of Tobit and OLS model features

Considering these changes in the slopes of most of the features, we can infer that ceiling price impacts prices in general; the impact is not limited to the hours during which we see MCP near ceiling prices.

4.9 Sensitivity analysis on MCP predictions

When getting predictions, we have manipulated the ceiling price feature to anticipate the numerical impact of ceiling prices on market clearing prices. By doing so, we can get the market clearing price of the scenario when having higher ceiling prices.

In the test period that we have worked on, between 05-07-2023 and 31-07-2023, we have a ceiling price of 2700 TL/MWh. In this scenario, we have the mean of MCP as 1917.46 TL/MWh. Figure 46 is the prediction histogram in this original scenario, where it has high accuracy:

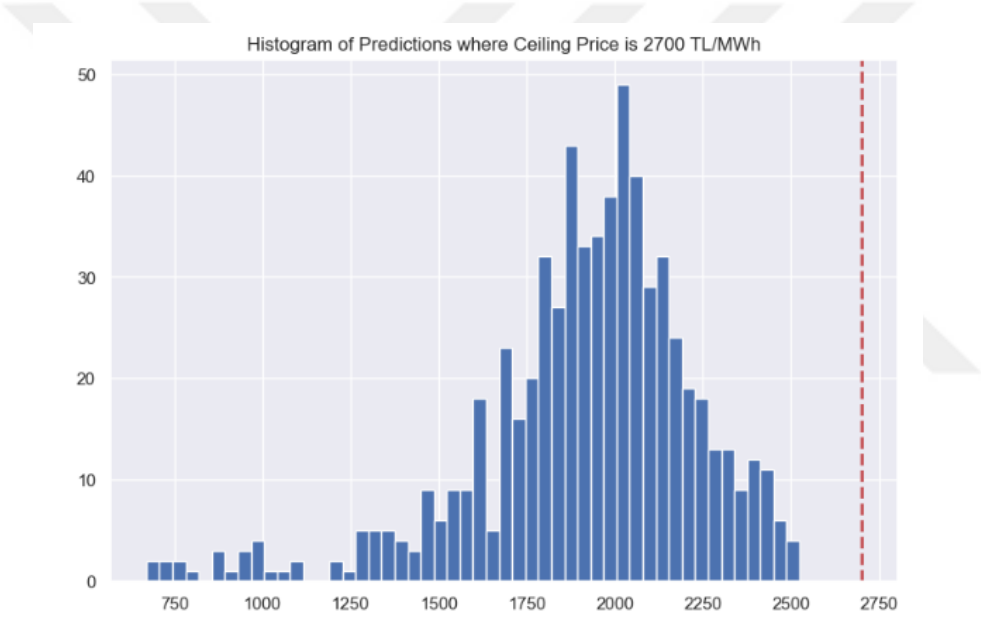

Figure 46. Histogram of predictions where ceiling price is 2700 TL/MWh

We have increased this ceiling price by 200 TL/MWh and get predictions in the same time frame; when we do so, the mean of MCP is increased to 2201.500. This increase is almost 1.5 times our increase in ceiling prices. Figure 47 is the histogram of this scenario:

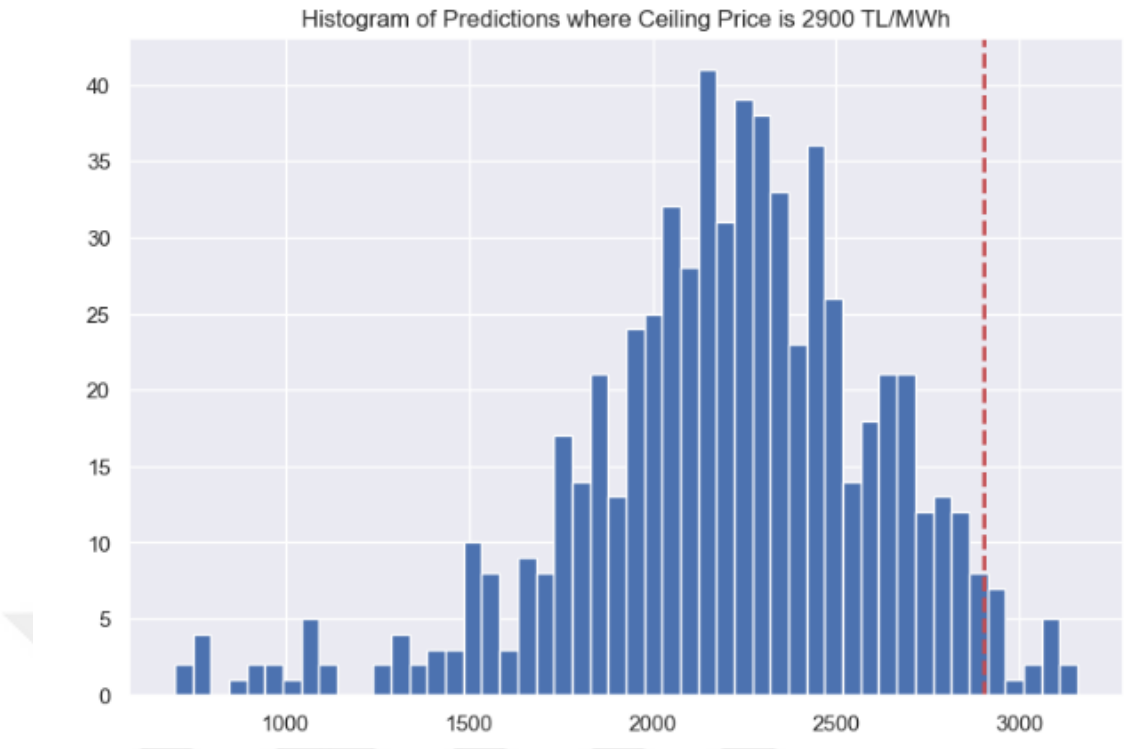

Figure 47. Histogram of predictions where ceiling price is 2900 TL/MWh

Our model fails to predict MCPs under the ceiling price in this prediction. In the histogram, in the lines where MCP is higher than the ceiling price, we see the limitation of our model in keeping predictions under the ceiling price.

In another scenario, we have increased the ceiling price to 3200 TL/MWh; this is 500 TL/MWh higher than the original scenario. When predictions are created, the mean of MCP is observed as 2431.46 TL/MWh. This increase in the mean is close to the increase that we did in the ceiling price. Figure 48 is the histogram of this scenario.

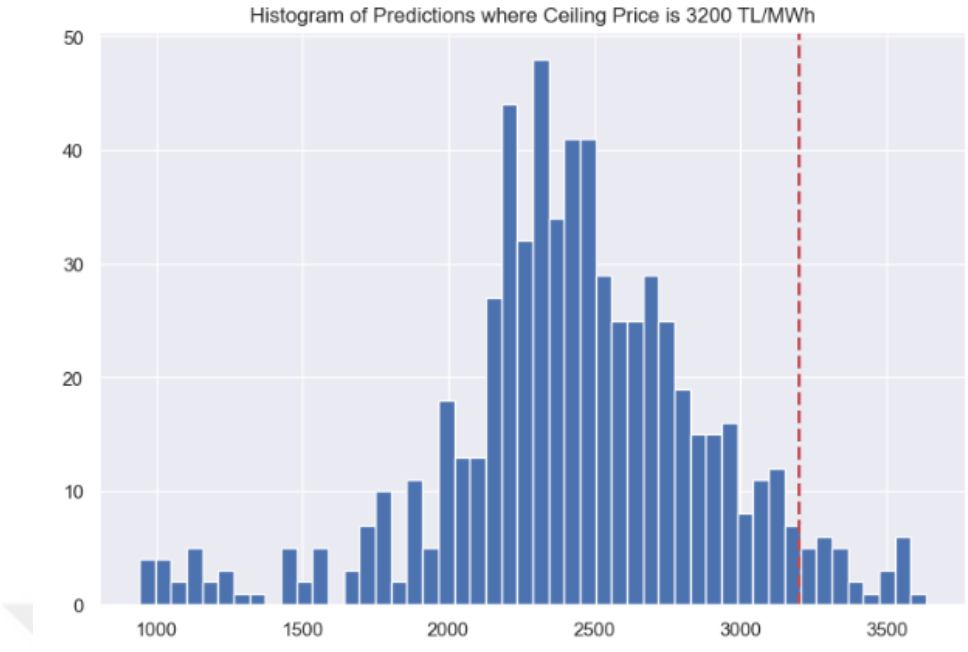

Figure 48. Histogram of predictions where ceiling price is 3200 TL/MWh

We see more records of disability to keep predictions under the ceiling price.

Other ceiling prices above 3200 TL/MWh have also been tried to get predictions, but the model could not return results for these trials. We can consider this level as the maximum level of ceiling prices for our model trained with historical ceiling prices. With this prediction model, the impact of ceiling prices on market clearing price can be anticipated as %26 by calculating the differences between means of scenarios of 2700 and 3200 TL/MWh and taking the proportion to 2700 TL/MWh scenario.

CHAPTER 5:

CONCLUSION

This study analyzes the impact of ceiling prices on day-ahead market clearing prices with three prediction models. The ceiling price feature is critical for the prediction model with both feature importance analysis and sensitivity analysis we did with this feature.

Ceiling prices are critical for electricity prices because they censor the data and should be handled carefully in electricity price predictions. As a result of the comparison between Tobit and linear regression models, it is seen that when we eliminate censors from the data, other features' behavior also changes. This evidences that the impact of censors is not only practical in hours when we see higher prices close to price ceilings, but it is also effective in the overall market behavior in all hours of trade. Overall behavior impact has been analyzed, and material has been found with the analysis of ceiling prices on the other determinants of the market clearing price.

With the analysis of the impact of ceiling prices on market clearing price, it is seen that ceiling prices are not only effective in the hours where we see market clearing price equal to ceiling price, but they are also effective in the other hours of the day other than hours where ceiling price is touched.

Another important outcome of the prediction models is that the ceiling price, which is of the highest importance in the models, has a static value for the prediction period of 2700 TL/MWh. It is exciting to see this kind of importance in the model. Although it has different values in the training data, we have a very static distribution

in the prediction area for each prediction hour. This shows that the ceiling price is synchronized with the other features used in the model in training data. We can say that the ceiling price of July has been determined by effectively covering the dynamics in 2 years of training data.

On the other hand, it is seen that merit order in the supply curve is also quite critical in the determination of market clearing price. The essential features followed by ceiling prices are efficient natural gas marginal cost, production from natural gas, production from imported coal, and gas reference price; each is related to the merit order mechanism we have in supply curves. The intersection points that will occur in supply and demand curves are dependent on the merit order features. It is commonly seen that intersections generally occur at the point where we see natural gas plants' marginal costs in the supply curve.

Although we have more features from the supply curve, the ones we have for the demand curve are also significant. The intersection point of these curves depends on the demand dynamics. Minimum demand in the demand curve and price-independent bid orders are some of the examples that are important for the model.

The multi-seasonal feature of electricity prices is also essential in these models. The daily, weekly, and yearly seasonality impacts are seen in the model prediction. These features are fed into a model with categorical features. However, it is essential to note that the other features we incorporate into the model include the impacts of seasonality. Because of this, we do not see these categorical seasonality features in the top ranks of feature importance.

It is also important to say that electricity prices are being affected by many factors, and in order to predict electricity prices accurately, features must be created

and selected carefully. Feature engineering is essential for creating meaningful features for the model and market clearing price prediction success. Although there are many freely available data sources about electricity markets, it is necessary to create meaningful features from these available sources. From this perspective, we can understand the importance of field knowledge on the MCP predictions. Creating informative features for the model requires comprehensive knowledge of the market dynamics. This knowledge must be followed by effective technical implementation into data that are inputs of the model.

This study aimed to understand the relationship between ceiling price, other features, and MCP. Because of this, we have not created a model to predict the following day's clearing prices. We have used the actual values seen in the market as prediction inputs to the model (X_test). Because of this, our task did not reflect the challenges faced by people who try to predict the next day’s electricity price experience. We were able to get successful prediction results with exact test inputs equal to real data. If we were in the shoes of predictors working on the next day's or next week's prices, our tasks would be complex, including predicting the test inputs to the model because they did not happen at the time of prediction. These models that predict the following hours are generally designed to have scenarios because the test inputs of models are also very hard to predict.

CHAPTER 6:

LIMITATIONS AND FUTURE WORK

As mentioned previously, the model proposed in this study aims to understand the impact of ceiling prices on MCP. However, in industry, predictors aim to predict the next hour's price as accurately as possible. Further studies show the impact of ceiling prices and how predictors should handle these prices, which can also be analyzed. If it is impactful in the model with an analysis aim, it will also be impactful for the models that aim to predict the following hours. In the second and third models, we have seen that ceiling prices impact the other features that we use in modeling; this should be handled in the prediction of the next hour's test data.

In this study, the method of creating a prediction model is employed. However, stochastic methods are also commonly used in these kinds of impact analyses. These methods, which use probabilistic models to predict and interpret data, are particularly effective in dealing with the uncertainties and variabilities inherent in electricity markets. Using stochastic models, researchers can simulate various market scenarios under different ceiling price levels. This approach helps understand how these caps affect market dynamics, such as supply and demand equilibrium, price volatility, and investment incentives. Additionally, stochastic methods enable examining the ripple effects of ceiling prices on related sectors, such as renewable energy development and carbon emissions. For these reasons, stochastic models should also be used to analyze the impact of ceiling prices in future steps.

We used linear regression-based models in the second and third models in which we analyzed censors' impact on MCP prediction. These models were quite

successful in the interpretability of our results. On the other hand, they cannot cover the factors from many different sources to predict the MCP accurately. Because of this, this study can be extended to more complex models, such as machine learning models integrated with censor adjustments. With this kind of model, the impact of predictions can be more accurately analyzed.

This study analyzes the impact of ceiling prices on market clearing prices, considering that it creates a right censor in the data with the maximum price. However, the starting point of the market clearing price is 0; in other words, it could not take values below 0. This is another censor in the data that can be further analyzed. In some day-ahead electricity markets, negative pricing is allowed. Negative pricing happens when the supply of electricity exceeds demand. This situation is more likely to occur when electricity demand is low (like overnight or during holidays) and when there is a high production from renewable sources like wind and solar power. Some traditional power plants, such as nuclear, cannot be shut down or have their output adjusted quickly. So, during low demand and high renewable generation periods, these plants continue to generate power, contributing to the oversupply. Producers might prefer to pay to offload their excess electricity rather than shut down their plants. Shutting down and restarting plants, especially non-renewable ones, can be costly and time-consuming. Negative prices are a signal that these generators need to reduce their output. Excess electricity can pose a risk to the stability of the power grid. Negative prices can act as a mechanism to balance the grid, encouraging large consumers to increase their usage or generators to reduce their output.

REFERENCES

Abbott, T. A. (1995). Price regulation in the pharmaceutical industry: Prescription or placebo? *Journal of Health Economics*, *14*(5), 551–565. https://doi.org/10.1016/0167-6296(95)00022-4

Arnott, R. (1995). Time for revisionism on rent control? *The Journal of Economic Perspectives*, *9*(1), 99–120.

Athey, S., & Imbens, G. W. (2017). The state of applied econometrics: causality and policy evaluation. *Journal of Economic Perspectives*, *31*(2), 3–32. https://doi.org/10.1257/jep.31.2.3

Borenstein, S., & Rose, N. L. (1994). Competition and price dispersion in the U.S. airline industry. *Journal of Political Economy*, *102*(4), 653–683.

Boru İPek, A. (2021). Prediction of market-clearing price using neural networks based methods and boosting algorithms. *International Advanced Researches and Engineering Journal*, *5*(2), 240–246. https://doi.org/10.35860/iarej.824168

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control*. John Wiley & Sons.

Bunn, D. (2004). *Modelling Prices in Competitive Electricity Markets | Wiley*. Wiley.Com. https://www.wiley.com/en-us/Modelling+Prices+in+Competitive+Electricity+Markets-p-9780470848609

Chen, L., Mislove, A., & Wilson, C. (2015). Peeking beneath the hood of Uber. *Proceedings of the 2015 Internet Measurement Conference*, 495–508. https://doi.org/10.1145/2815675.2815681

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., & Newey, W. (2017). Double/debiased/neyman machine learning of treatment effects. *American Economic Review*, *107*(5), 261–265. https://doi.org/10.1257/aer.p20171038

Contreras, J., Espínola, R., Nogales, F. J., & Conejo, A. J. (2003). ARIMA models to predict next-day electricity prices. *IEEE Transactions on Power Systems*, *18*(3), 1014–1020. Scopus. https://doi.org/10.1109/TPWRS.2002.804943

Crespo Cuaresma, J., Hlouskova, J., Kossmeier, S., & Obersteiner, M. (2004). Forecasting electricity spot-prices using linear univariate time-series models. *Applied Energy*, *77*(1), 87–106. https://doi.org/10.1016/S0306-2619(03)00096-5

Enerji Atlası. (2023a). *Wind Farm Potential Data by Province*. https://www.enerjiatlasi.com/ruzgar-enerjisi-haritasi/turkiye

EPDK. (2023b). *EPDK Electricity Market Sector Report*. https://www.epdk.gov.tr/detay/icerik/3-0-23/elektrikaylik-sektor-raporlar

Glaeser, E. L., & Luttmer, E. F. P. (2003). The misallocation of housing under rent control. *American Economic Review*, *93*(4), 1027–1046. https://doi.org/10.1257/000282803769206188

Gürtler, M., & Paulsen, T. (2017). Forecasting performance of time series models on electricity spot markets: A quasi-meta-analysis. *International Journal of Energy Sector Management*, *12*(1), 103–129. https://doi.org/10.1108/IJESM-06-2017-0004

Haldrup, N., & Nielsen, M. Ø. (2006). A regime switching long memory model for electricity prices. *Journal of Econometrics*, *135*(1–2), 349–376. https://doi.org/10.1016/j.jeconom.2005.07.021

Jedrzejewski, A., Lago, J., Marcjasz, G., & Weron, R. (2022). Electricity price forecasting: The dawn of machine learning. *IEEE Power and Energy Magazine*, *20*(3), 24–31. https://doi.org/10.1109/MPE.2022.3150809

Joskow, P. (2007). *Regulation of Natural Monopoly* (pp. 1227–1348) [Handbook of Law and Economics]. Elsevier. https://econpapers.repec.org/bookchap/eeelawchp/2-16.htm

Joskow, P. L. (2001). California's electricity crisis. *Oxford Review of Economic Policy*, *17*(3), 365–388. https://doi.org/10.1093/oxrep/17.3.365

Katz, M. L., & Rosen, H. S. (1998). *Microeconomics (3rd ed.)*. McGraw-Hill.

Kristiansen, T. (2012). Forecasting Nord Pool day-ahead prices with an autoregressive model. *Energy Policy*, *49*, 328–332. https://doi.org/10.1016/j.enpol.2012.06.028

Lehna, M., Scheller, F., & Herwartz, H. (2022). Forecasting day-ahead electricity prices: A comparison of time series and neural network models taking external regressors into account. *Energy Economics*, *106*, 105742. https://doi.org/10.1016/j.eneco.2021.105742

Li, W., & Becker, D. M. (2021). Day-ahead electricity price prediction applying hybrid models of LSTM-based deep learning methods and feature selection algorithms under consideration of market coupling. *Energy*, *237*, 121543. https://doi.org/10.1016/j.energy.2021.121543

Madadkhani, S., & Ikonnikova, S. (2024). Toward high-resolution projection of electricity prices: A machine learning approach to quantifying the effects of

high fuel and CO2 prices. *Energy Economics*, *129*, 107241. https://doi.org/10.1016/j.eneco.2023.107241

Mankiw, N. G. (2018). *Principles of Economics (8th ed.)*. Cengage Learning.

Misiorek, A., & Weron, R. (2005). *Forecasting spot electricity prices with time series models*. https://www.academia.edu/29565250/Forecasting_Spot_Electricity_Prices_with_Time_Series_Models

Nogales, F. J., Contreras, J., Conejo, A. J., & Espinola, R. (2002). Forecasting next-day electricity prices by time series models. *IEEE Transactions on Power Systems*, *17*(2), 342–348. https://doi.org/10.1109/TPWRS.2002.1007902

Pinhão, M., Fonseca, M., & Covas, R. (2022). Electricity spot price forecast by modelling supply and demand curve. *Mathematics*, *10*(12), 2012. https://doi.org/10.3390/math10122012

Shahidehpour, M., Yamin, H., & Li, Z. (2003). *Market Operations in Electric Power Systems: Forecasting, Scheduling, and Risk Management*. John Wiley & Sons.

Stock, J. H., & Watson, M. W. (2017). Twenty years of time series econometrics in ten pictures. *Journal of Economic Perspectives*, *31*(2), 59–86. https://doi.org/10.1257/jep.31.2.59

Taylor, J. W. (2010). Triple seasonal methods for short-term electricity demand forecasting. *European Journal of Operational Research*, *204*(1), 139–152. https://doi.org/10.1016/j.ejor.2009.10.003

Tschora, L., Pierre, E., Plantevit, M., & Robardet, C. (2022). Electricity price forecasting on the day-ahead market using machine learning. *Applied Energy*, *313*, 118752. https://doi.org/10.1016/j.apenergy.2022.118752

Uniejewski, B., Nowotarski, J., & Weron, R. (2016). Automated variable selection and shrinkage for day-ahead electricity price forecasting. *Energies*, *9*(8), 621. https://doi.org/10.3390/en9080621

Wagner, A. (2014). Residual demand modeling and application to electricity pricing. *The Energy Journal*, *35*(2).

Weron, R. (2006). Front Matter. In *Modeling and Forecasting Electricity Loads and Prices* (pp. i–xii). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118673362.fmatter

Weron, R. (2014). *Electricity price forecasting: A review of the state-of-the-art with a look into the future | Elsevier Enhanced Reader*. https://doi.org/10.1016/j.ijforecast.2014.08.008

Weron, R., & Misiorek, A. (2006). Short-term electricity price forecasting with time series models: A review and evaluation. *HSC Research Reports*, Article HSC/06/01. https://ideas.repec.org//p/wuu/wpaper/hsc0601.html

Weron, R., & Misiorek, A. (2008). Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models. *International Journal of Forecasting*, *24*(4), 744–763. https://doi.org/10.1016/j.ijforecast.2008.08.004

Wilson, R. (2018). Market design in wholesale electricity markets. In handbook of market design (pp. 539-574). *Oxford University Press*.

Xie, H., Chen, S., Lai, C., Ma, G., & Huang, W. (2022). Forecasting the clearing price in the day-ahead spot market using eXtreme Gradient Boosting. *Electrical Engineering*, *104*(3), 1607–1621. https://doi.org/10.1007/s00202-021-01410-6

Ziel, F., & Weron, R. (2018). Day-ahead electricity price forecasting with high-dimensional structures: Univariate vs. multivariate modeling frameworks. *Energy Economics*, *70*, 396–420. https://doi.org/10.1016/j.eneco.2017.12.016