FACULDADE
DE ECONOMIA
UNIVERSIDADE
DO PORTO

Electricity Day-ahead Price Forecasting for MIBEL Market

Ana Sofia Pereira Brito

Dissertation

Master in Modelling, Data Analysis and Decision Support Systems

Supervised by:

Maria Eduarda da Rocha Pinto Augusto da Silva

Co-supervised by:

José Abilio de Oliveira Matos

2025

Acknowledgments

Firstly, I would like to express my sincere gratitude to my supervisors, Professor Maria Eduarda Silva and Professor José Matos, for all their help and support throughout this journey.

I would also like to thank Jorge Leite, employee at EDP, for providing me with the data I needed to complete my dissertation and for his help and advice.

To Catarina, Carolina, Miguel, Patrícia, Renato and Mafalda, thank you for always being by my side and helping me to be the best version of myself.

To my friends from my master's programme, Clara, Ricardo, João, Sara and Sofia, thank you for all the moments we shared, which made these two years a much more beautiful journey.

And finally, to my family, for always being my inspiration and support.

Abstract

The electricity market is characterised by its volatility and the continuous growth of renewable energy sources, making price forecasting even more crucial. This dissertation aims to forecast the daily price of electricity in the Iberian electricity market (MIBEL), based on the use of autoregressive models with the application of smoothing techniques, such as Elastic Nets. The methodology is based on the study by Ziel et al. (2016), which was previously applied to the Nordic and British markets, which have similar characteristics to the market under study. In addition, more advanced models, such as machine learning, will be used to compare the results and their accuracy in a specific market such as MIBEL. This work will present a state of the art demonstrating the various types of models used to predict electricity prices, as well as an explanation of what the MIBEL market is and how it works.

Keywords: Electricity Price Forecasting, MIBEL Market, Autoregressive Model (fARX), Elastic Nets, Renewable Energies

Resumo

O mercado da eletricidade caracteriza-se pela sua volatilidade e pelo crescimento contínuo das fontes de energia renováveis, tornando a previsão dos preços ainda mais crucial. Esta dissertação tem como objetivo prever o preço diário da eletricidade no mercado ibérico de eletricidade (MIBEL), com base na utilização de modelos autorregressivos com a aplicação de técnicas de suavização, tais como Elastic Nets. A metodologia baseia-se no estudo de Ziel et al. (2016), que foi anteriormente aplicado aos mercados nórdico e britânico, que têm características semelhantes ao mercado em estudo. Além disso, modelos mais avançados, como machine learning, serão utilizados para comparar os resultados e a sua precisão num mercado específico como o MIBEL. Este trabalho apresentará um estado da arte demonstrando os vários tipos de modelos utilizados para prever os preços da eletricidade, bem como uma explicação sobre o que é o mercado MIBEL e como funciona.

Palavras-chave: Previsão de Preços de Eletricidade, Mercado MIBEL, Modelo Autoregressivo (fARX), Elastic Nets, Energias Renováveis

Contents

Glossary

MIBEL- Iberian Electricity Market
EPF- Electricity price forecasting
DA- Day-ahead
QH- Quartly-hour
fARX- Full Autoregressive model
OMEL- Iberian market operator in Spain
OMIP- Iberian Market Operator in Portugal
EPEX- European Power Exchange
REN- Redes Energéticas Nacionais
ESIOS- Sistema de Información del Operador del Sistema
SDAC- Single day-ahead coupling
CET- Central European Time
SIDC- Single Intraday Coupling
DAEPF- day-ahead electricity prices
ARIMA- Autoregressive integrated moving average
ARX- Autoregressive with exogenous inputs
SARIMA- Seasonal Auto-Regressive Integrated Moving Average
GARCH- Generalized Autoregressive Conditional Heteroscedasticity
ARCH- Autoregressive Conditional Heteroscedasticity
SVM- Support Vector Machine
RF- Random Forest
NN- Neural Network
KNN- k-nearest neighbors
MLP- Multi-Layer Perceptron
SVR- Support Vector Regressor
MAE- Mean Absolute Error
WMAE- weekly weighted average absolute error
MAPE- Mean Absolute Percentage Error
RMSE- Root Mean Square
MWh- Megawatt-hour

List of Figures

List of Tables

Chapter 1

Introduction

1.1 General Context

Considering the complexity and volatility of the electricity market, Cerasa and Zani (2025), price forecasting is an essential process for all parties involved. Since electricity cannot be stored, it is crucial to balance production and consumption, making the price influenced by several variables, including weather and consumer demand.

Electricity price forecasting (EPF) enables companies to plan their operations in advance, helps to reduce the risks associated with price volatility, enabling agents to make more informed decisions and, with more accurate forecasts, helps to reduce their costs.

The electricity price forecast will be applied in the Iberian Electricity Market (MIBEL), which stands out as one of the main European markets, differentiating itself from other markets. MIBEL includes the countries of Portugal and Spain and what distinguishes this market from others is the high penetration of renewable energies, with a 34 per cent increase in solar installations and a 59 per cent increase in hydroelectric installations by 2023, contributing to lower electricity prices. Another unique feature is its interconnection between Portugal and Spain, with a high degree of price convergence. In 2023, the price difference between the two countries was less than 1€/MWh in 95 per cent of the hours, as the Figure 1.1 shows, extract from *Annual Report 2023* (2023).

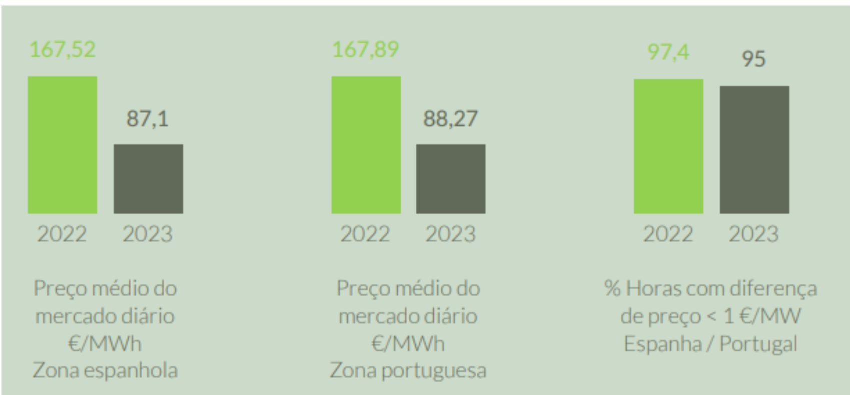

Figure 1.1: 2023 Day-ahead mean price, Source: OMEL

1.2 Thesis Prupose

The purpose of this dissertation is to apply the Full Autoregressive model (fARX) with selection and shrinkage Procedures, such as Elastic Nets described in the article Uniejewski, Nowotarski, and Weron (2016) to the MIBEL market for daily price forecasting. Finally, a decision tree-based model, XGBoost, will be applied to compare the prediction of the model under study.

The model was initially applied to the Nordic and UK markets. Several variables are taken into account by this autoregressive fARX model, which employs selection techniques like elastic nets to enhance the model's prediction. To get more accurate results, advanced forecasting techniques will be applied to MIBEL data. The model is being applied to this market given that the Nordic market is known for having a significant proportion of renewable energy sources, which may cause price volatility equivalent to MIBEL's.

1.3 Structure

This dissertation is organised into four chapters. The first will be the Introduction, where the problem will be framed and the objectives of the study will be defined. This is followed by the Contextualisation of MIBEL, which will describe how the market works and the process of electricity price formation. The Literature Review, the third chapter, will cover different models and methods that have already been used to predict electricity prices in various markets. Next, in Chapter 4, we'll describe in detail what the model consists of, how it's put together and how we're going to adapt it for the market under study. In Chapter 5, in the methodology section, we'll cover the whole process from creating the model to forecasting it, including the machine learning model chosen, which is XGBoost. Finally, we'll analyse the results and draw a conclusion, describing the performance of the two models and drawing final conclusions.

Chapter 2

Contextualizing MIBEL

MIBEL, the Iberian Electricity Market, according to *Mercado Ibérico da Energia Elétrica – MIBEL* (n.d.), was formally set up on 1 July 2007, an initiative coming from the collaboration between the governments of Portugal and Spain in order to integrate their electricity markets. The aim of this merger was to create a single electricity market that would benefit consumers in both countries and increase market efficiency, reduce costs and promote competition in the electricity sector.

In this journey, which began in 1998, some important milestones stand out, such as the signing of the collaboration protocol between the Portuguese and Spanish administrations in November 2001 and the signing of the agreement between the Portuguese Republic and the Kingdom of Spain in Santiago de Compostela in October 2004. The latter agreement established the legal framework for the creation of MIBEL, with OMEL, the Iberian market operator in Spain, being responsible for managing the daily and intraday market and OMIP, the Iberian market operator in Portugal, for the forward market.

The importance of the 22nd Luso-Spanish Summit, where the Regulatory Compatibility Plan was defined, held in Badajoz in November 2006, and the signing of the Agreement revising the Santiago Agreement in January 2008 and more recently in February 2014, with the coupling of MIBEL with the north-west European markets, should also be highlighted.

2.1 How does the MIBEL market work?

MIBEL, is a **Wholesale Market**, where producers and commercializers negotiate electricity before it reaches final consumers, ensuring a balance between production and consumption and taking into account the needs of the agents. This market can be divided into 4 submarkets. The first is the derivatives market, where electricity futures contracts are traded, which can be done through physical settlement when there is an actual delivery of electricity or through financial settlement when there is monetary compensation, but in this case without physical delivery of the energy. This submarket aims to protect agents against price fluctuations, ensuring greater stability and predictability. As stated before, the OMIP (Portuguese Electricity Market Operator) is responsible for managing this market, as it is more geared towards complex, long-term negotiations. The second is the spot market, which is managed by OMEL (Spanish Electricity Market Operator), as it is responsible for short-term management. This submarket is made up of the day-ahead market, which will be discussed in more detail below, which allows the purchase and sale of electricity for the following day.

Figure 2.1: Day-ahead market session, Source: OMEL

It is through this market that prices are formed through the demand and supply curve for each hour of the following day. The other market is the intraday market, where real-time adjustments are made after the day-ahead market, allowing for greater flexibility in the event of changes in supply or demand. The third submarket, the system services market, makes sure that electricity system operations are secure and stable in real time, ensuring that there is a balance between generation and consumption. And finally, the bilateral contract market, which allows direct negotiations between agents without the intermediation of market operators in situations where specific needs are met.

The day-ahead market and the intraday market, which are governed by OMEL, are the two most significant trading mechanisms within MIBEL. In order to balance supply and demand and make sure efficient production and consumption of electricity, these two markets are crucial. Both of these markets are interconnected. As previously stated, the day-ahead market is where electricity transactions occur, with market agents selling and buying electricity for the next 24 hours. The intraday market is used to adjust these plans following the daily market's close. Together, these two markets provide a trustworthy and open system for delivering electricity to consumers.

2.1.1 Day-ahead Market

One of MIBEL's main markets is the **day-ahead market**, also called single day-ahead coupling (SDAC), where electricity is traded one day ahead of time. It functions similarly to an auction, with producers indicating how much energy they plan to supply and at what price point, and buyers specify the amount of energy they need and what price they are willing to pay. The meeting point between supply, which comes from electricity producers, and demand, which comes from distributors, commercializers, and consumers, determines the schedules for electricity production and consumption. The Portuguese government only introduced this market in 2007, but Spain has been benefiting from it since 1998.

There is a set schedule for this market's sessions and results publication, illustrated in Figure 2.5. Electricity prices and quantities are established for the entire European Union during the daily market session, which begins at 12:00 CET every day, *Annual Report 2023* (2023).

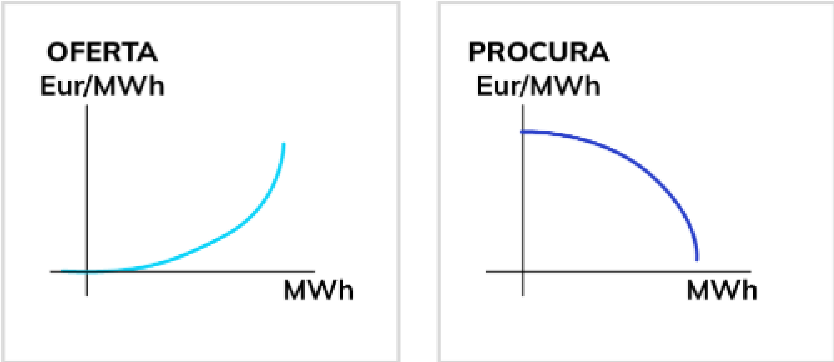

Figure 2.2: Offer and demand curve, Source: EDP

2.1.2 Price Formation

Price formation in this market is a daily process involving the submission of sell offers and buy bids for each hour of the following day. A brief description of the process will be given below.

The **offer curve**. is the combination of all the sales offers proposed by electricity producers for each hour of the following day. These are sorted in ascending order of price, forming a curve that represents the total amount of energy that producers are willing to sell for each type of price. The overall shape of an offer curve is represented in Figure 2.2. However, the price of the offers can fluctuate significantly depending on the type of energy they are selling. An example of an offer curve is depicted in Figure 2.3 for the 1st January 2025. Due to their extremely low production costs, nuclear power plants and special regime (renewable) stations - such as wind, solar, run-of-river hydro and biomass - generally have low margin costs, as indicated by the **lower part of the curve** in Figure 2.3. It is essential to highlight that Portugal's special regime of production—which includes wind and other renewable energy sources—does not directly participate in the production market. In this particular case, it is sold to a final supplier, who uses it to lower demand in their purchase proposals. Power plant with moderately variable costs, like combined cycle coal and natural gas plants, are located in the **middle of the curve**, Figure 2.3. These are slightly more expensive than the earlier models and are more susceptible to changes in the price of petrol, but can still be competitive in many situations.

Lastly, fuel oil power plants and adjustable hydroelectric power plants are located at the **top**, Figure 2.3. Because the opportunity cost of using water to generate electricity is influenced by market price expectations, these plants,in the case of hydroelectric plants, have higher costs. For example, if the expected price for a future period is higher, power stations may postpone production, causing the price of supply to increase. Both Portugal and Spain have diversified electricity production, with Portugal being more dependent on renewable sources.

Demand Curve

Unlike the supply curve, the **demand curve** aggregates all the purchase offers from consumers and market agents. A typical demand curve is represented in Figure 2.2 and the curve for 1st January 2025 is represented in Figure 2.3.

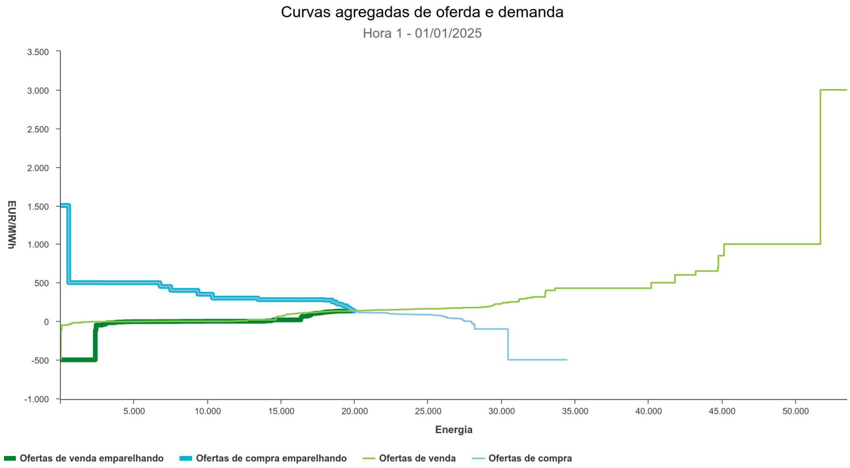

Figure 2.3: Aggregate supply and demand curves, Source: OMIE

Generally, the demand curve has a negative slope, i.e., the higher the price of electricity, the less consumers are willing to buy, reflecting the law of demand, which dictates that demand for a good decreases as its price increases.

The elasticity of demand varies according to the type of consumer, i.e., large industrial consumers have a higher elasticity because they have a greater capacity to adjust their consumption in response to variations in electricity prices, while small consumers have a lower elasticity because their consumption is less flexible.

Price Determination

The market operator develops their curves for each hour of the next day by aggregating the bids made by producers and consumers to sell and buy accordingly. The equilibrium price for that particular hour is determined by the intersection of the two curves. This price can be regarded as a marginal price, meaning that all transactions for that hour are settled at that price regardless of the price of the individual offers. It represents the value that strikes a balance between the amount of electricity that producers are willing to sell and the amount that consumers are willing to buy. This procedure is illustrated in Figure 2.3.

2.1.3 Market Splitting

The interconnection capacity shows the maximum amount of energy that can be transferred between the two countries in a specific period of time. The interconnection is the transmission lines that link the electrical systems of Portugal and Spain, allowing electricity to flow between the two countries and integrating the market.

However, there are times when the flow of energy needed to satisfy joint demand and supply exceeds the available interconnection capacity, and this is where the Market Splitting mechanism comes in. This market, which went into effect on July 1, 2007, is used to control this kind of congestion in the Portugal-Spain connection.

MIBEL typically functions as a single market, meaning it has a single supply and demand curve. However, when the balancing point is identified and it is concluded that the necessary energy flow exceeds the interconnection capacity, a congestion situation is identified. If this occurs, the markets are divided into two segments, one for Spain and one for Portugal, each of

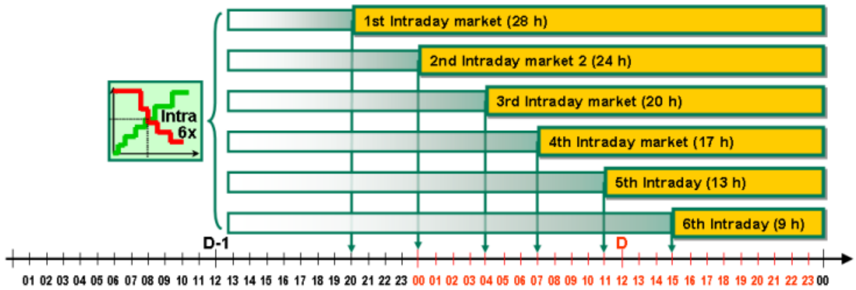

Figure 2.4: 6 steps of intraday market, Source: OMIE

	SESION 1	SESION 2	SESION 3
Opening of the Sesion	14:00	21:00	9:00
Clousure of the Sesion	15:00	22:00	10:00
Matching and publication	15:00	22:00	10:00
Programming Horizon (Hourly Periods)	24 hours (1-24 D+1)	24 hours (1-24 D+1)	12 hours (13-24 D)

Figure 2.5: 3 step intraday market, Source: OMIE

which has its own supply and demand curve. The balance point will be determined differently for each market by varying the interconnection capacity, which will lead to differing power pricing in Portugal and Spain. Market splitting aims to securely and efficiently monetize the use of interconnection capacity so that power reaches the most important areas even when capacity is limited.

As mentioned above, use of this market results in different prices. Energy moves from the nation with the lowest price to the one with the highest price, up to the limit of the interconnection capacity and indicates whether or not it is congested.

2.1.4 Intraday Market

The intraday market can be thought of as a balancing market that aims to adjust energy positions after the close of the day-ahead market based on their needs in real time.

Previously, this market was divided into 6 sessions, as shown in Figure 2.4, however, this structure was changed to 3 auction sessions (see Figure 2.5) at the European level, and a continuous cross-border market was introduced, or in other terms, single intraday coupling (SIDC), OMIE (n.d.). The reasons for this change were to harmonise the European electricity markets, following a model approved by them with regard to the management of interconnections. The introduction of this continuous market allows agents to manage energy imbalances more flexibly up to an hour before the energy is delivered. The concentration of the market to 3 sessions was intended to increase its liquidity since it has access to electricity markets in other areas of Europe, resulting in more efficient prices.

Thus, in these three auctions, the volume of energy and the price are determined by the intersection between the supply and demand curve, following the marginalist and market coupling models. In addition to the auctions, there is the continuous market, which operates uninterruptedly, giving agents the opportunity to adjust their positions up to an hour before the energy is delivered. However, cross-border trading on the continuous market is interrupted 20 minutes before the close of each auction session so that contracts can be entered on the horizon of that auction, with trading taking place only at a local level until the session's bids are closed. After the close of the first auction of the current day (D), trading begins on the continuous market for the following day (D+1), provided that the system operator has published the definitive daily programme for the following day.

Chapter 3

Literature Review

As mentioned earlier, forecasting day-ahead electricity prices (DAEPF) is crucial for several reasons; however, this is becoming an increasingly complex process due to various factors, such as price volatility and solar and wind energy production.

According to Mao et al. (2025), price complexity arises from the interaction of multiple factors, such as the dynamics between supply and demand, the volatility of fuel prices, the variability of renewable energy sources, the fact that it is not possible to store energy, and the uncertainty of the actions of market participants, among others.

Price forecasting models have therefore evolved and can be structured into three main categories:

- *Traditional models: such as ARIMA, SARIMA and GARCH, among others.*
- *Automatic learning models: such as Support Vector Machine (SVM), Random Forest (RF), and Neural Networks, among others.*
- *Hybrid models: these models can combine two or more EPF techniques, such as ARIMA with regularisation models.*

It should be emphasised that more complex models do not lead to greater performance in predicting the price of electricity, and sometimes simpler models can be more effective in studying historical patterns and their fluctuations.

This chapter discusses the various model types in greater detail in the sections that follow.

3.1 Traditional models

3.1.1 ARIMA and SARIMA

Classical ARIMA time series models have been used to forecast electricity market prices . These models can be adjusted to deal with seasonality and market volatility (SARIMA models), as well as deal with external variables, such as ARX (Forecasting Nord Pool), which includes exogenous variables, such as Nordic demand and Danish wind energy production, to improve forecast accuracy

Kristiansen (2012) Contreras, Espínola, Nogales, and Conejo (n.d.) In terms of their effectiveness, it can be concluded that the models are even quite accurate when it comes to forecasting prices, with errors of between 5 and 8 per cent in markets such as Spain and California. When studying this type of market, it can be seen that its characteristics, such as high price volatility, outliers and seasonality, affect the accuracy of the forecast and the inclusion of exogenous variables can improve the model's performance when there is a strong correlation between these variables and electricity prices.

It can be seen that the same model can give very different results depending on the nature of the market, for example, as the Spanish market is more volatile, the model had to be adjusted, requiring data from the previous 5 hours to predict the price of the next hour, without using differentiation, while for the California market,it was only necessary to use the previous 2 hours, but with hourly differentiation. In the case of the Nordic market, which is similar to the MIBEL market in terms of the high use of renewable energies, the use of models such as ARX can bring several benefits.

3.1.2 GARCH

Garcia, Contreras, van Akkeren, and Garcia (2005) Due to the fact that the electricity market time series have a non-linear behaviour and the ARMA model assumes a constant variance, it was necessary to build a new model to model the volatility of a time series. The model created is GARCH, Generalised Autoregressive Conditional Heteroskedasticity, an extension of the ARCH model, which aims to model the conditional variance, which is not constant over time.

The study found that this model proved to be a good predictor of electricity prices in the Spanish and California markets, with the average error in the Spanish market being around 9 per cent, while in California, before the crisis, it was around 5 per cent. However, this model had greater difficulty in predicting peaks, as in the case of the California market crisis, where the error rose to 36.54 per cent, and it can be concluded that GARCH is accurate in capturing price volatility but may have limitations in predicting peaks, such as the market crisis. Given that it can handle non-constant variances, the GARCH model has some significant advantages over the ARIMA model in markets with high volatility. However, putting this model into practice requires some statistical expertise, which makes parameter estimation more challenging. In contrast, ARIMA is much easier to use and comprehend.

3.2 Smoothing Models

In electricity price forecasting models, it is common to have a large amount of data, including a strong presence of explanatory variables, which makes forecasting challenging. In order to deal with these issues, regulation techniques have emerged, such as Ridge Regression, Least Absolute Shrinkage and Selection Operator (LASSO) and Elastic Nets, which help in the prediction of robust models and the selection of variables. These techniques make it possible to reduce the risk of overfitting and to identify the variables that contribute the most to your forecast in a world of extensive potential variables, which is common in electricity markets due to the diversity of various factors that influence prices.

Starting with **LASSO**, proposed by the article Uniejewski, Marcjasz, and Weron (2019), it introduces a penalty on the coefficients, which allows the most important variables to be selected automatically, forcing some coefficients to zero. As previously mentioned, this feature is suitable when talking about a high-dimensional database, where there can be hundreds of potential regressors, as is the case with the article, which has a version of 349 to 372 regressors, ranging from past prices to dummy variables.

The authors applied this model to the German market (EPEX) in order to improve predictive accuracy and were able to conclude that it outperforms simpler benchmark models such as Naive and ARX. They were also able to identify that the most important variables were the most recent intraday price, the day-ahead (DA) price corresponding to the same hour, DA prices from neighbouring hours and night DA prices, while the least explanatory were identified as day-of-the-week dummies and DA prices far ahead.

However, it is possible to conclude that this type of technique makes it possible to reduce the complexity of the model without it losing its performance, allowing for the construction of more simplified ARX models that are almost as effective, just by using variables that LASSO selects.

The authors Kath and Ziel (2018) also analysed the German market, but in this case they chose to study quarter-hourly (QH) prices, which are also characterised by high volatility due to the inclusion of renewable energies. They chose to use regression models such as **Ridge**, **LASSO** and, above all, **Elastic Nets** and concluded that the latter performed best, with significantly lower errors compared to simpler benchmarks.

Beyond the statistical comparison, we should emphasise a relevant point: small improvements in the performance of the models can generate significant profits in simple strategies based on forecasts, which in this case generate spreads of +0.7 €/MWh, representing millions in liquid markets. This evidence is important not only for the statistical exercise, but also to realise the impact it has financially and directly on market operations. These results are in line with the work of the authors of article Uniejewski et al. (2016), which is the article we're going to focus on later, where they used automatic variable selection and shrinkage methods in day-ahead forecasting in 3 different markets, Nord Pool, the UK and GEFCom2014.

This study showed that Elastic Nets, with an α=0.75, performed best, outperforming Ridge and LASSO. It was also possible to observe that Ridge, by applying an L2 penalty, managed to stabilise coefficients, but does not perform variable selection, while LASSO manages to force irrelevant coefficients to zero, although it can become unstable when there are strong correlations between variables. Elastic Nets manages to combine the strengths of each model while guaranteeing robustness and selection capacity, managing to automatically identify the most relevant predictors in a 107-variable model and improve predictive accuracy.

3.3 Machine Learning

In recent years, there has been a growing use of machine learning methods in electricity price forecasting, due to their ability to capture complex non-linear relationships and interactions between numerous explanatory variables. Among various algorithms, **Random Forest** stands out, which is based on emsembles of decision trees, and **XGBoost**, which is an optimised implementation of gradient boosting.

Díaz, Álvaro Romero, and Dorronsoro (2019) compared several models, including Ridge, k-nearest neighbours (KNN), multilayer perceptron (MLP), Support Vector Regressor (SVR) and Random Forest (RF), to forecast the hourly price in OMIE, using several basic variables, such as demand, wind and solar energy, temperature, calendar and price lags for Spain and France. This study shows that RF was the best performer, with an MAE of 3.92€/MWH for 2016 and 4.5€/MWh for 2017. However, it's important to note that normalising prices is usually chosen to reduce variance, but it destroys the relationship between price and demand, so in models like RF, non-normalisation is preferable.

Another model frequently used in electricity markets is **XGBoost**, and Bitirgen and Filik (2020) compares this model with ARIMA for day-ahead forecasting of hourly prices in the New York electricity market (NYISO). This market includes historical electricity prices and consumption of coal, natural gas, oil and uranium. The authors implemented a simple pipeline for XGBoost (exploration, importance calculation, feature selection, training and forecasting) and for ARIMA, identification, estimation, validation and forecasting, using evaluation metrics such as MAE; MAPE and RMSE for both models. The results of XGBoost were remarkable, as it outperformed ARIMA in all metrics (MAE 101.3 vs. 151.8; MAPE 306.9 vs. 480.7; RMSE 99.4 vs. 147.5). This study reveals a high MAPE, which could be explained by lower prices, where the value could explode. The study comes to the conclusion that XGBoost is more effective and computationally feasible than ARIMA in this specific scenario, but it also highlights the need to include important variables like load and weather or hybrid models to enhance peak detection and lower errors.

Chapter 4

Data and Exploratory Data Analysis

4.1 Data Description

The database used in this dissertation was created with data from two official sources, REN, Redes Energéticas Nacionais and ESIOS, Sistema de Información del Operador del Sistema. REN provided consumption data for Portugal and prices for both countries and ESIOS provided consumption data for Spain.

The period to be analysed will be from January 2023 to December 2024, on an hourly basis, and the variables chosen for both countries were:

- Electricity price, in €/MWh
- Electricity consumption, in MWh
- Renewable energies, such as photovoltaic and wind power

Before proceeding to the analysis, the data was pre-processed by replacing the negative price values with €0.01, all the variables were aligned with the same time zone to facilitate analysis and the hours of the day were set between 1 and 24.

The following exploratory analysis will aim to identify temporal patterns, seasonalities, differences between the two countries and identify relationships between variables that could be relevant for forecasting.

4.2 Hourly Patterns

The Figure 4.1 shows the average hourly price of electricity in Portugal and Spain. It can be seen that both countries have identical prices, with only a slight discrepancy in the second part of the day, between 11am and 6pm, which shows the integration of the MIBEL market. The Figure 4.1 shows that the lowest prices are in the early hours of the morning, between 2am and 6am, and in the afternoon, between 11am and 6pm, which coincides with lower demand for consumption during the night and also due to the increase in renewable energies in the afternoon, as the figure 4.2 shows. On the other hand, the highest values are recorded at the end of the day, between 8pm and 11pm, which can be explained by the increase in consumption in homes after working hours.

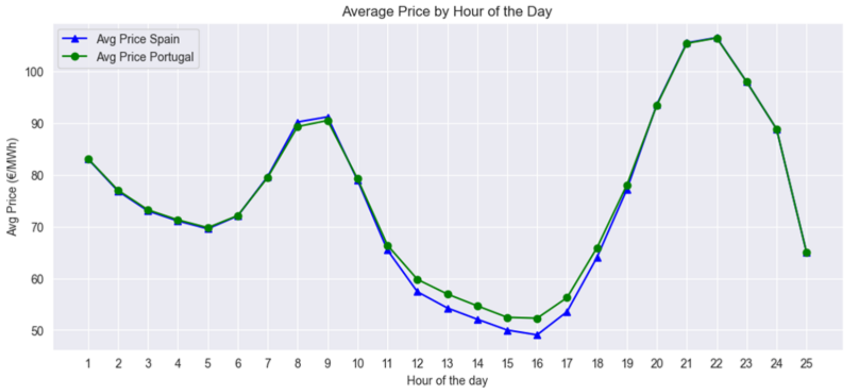

Figure 4.1: Average Price by Hour

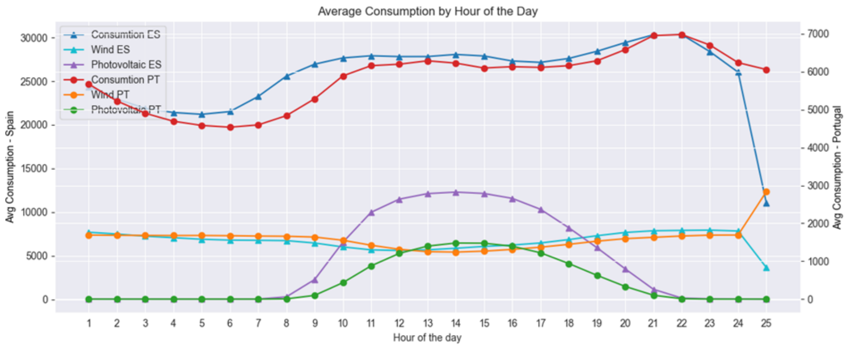

Figure 4.2: Average Consumption by Hour

The average hourly electricity consumption of Portugal and Spain can be seen in Figure 4.2, which shows the interaction with renewable sources, namely photovoltaic and wind power. As expected, it shows lower values between dawn and dusk, a progressive increase during the morning, and only at the end of the day does demand peak between 8pm and 11pm. This graph shows two scales, one for each country, due to the fact that Spain's consumption is much higher than that of Portugal.

One of the main sources for price formation is renewable energy and we can highlight the contribution of photovoltaic energy in the figure. It has a curve similar to the solar cycle, where we can see that it starts at 8am, peaks between 11am and 5pm and disappears after 9pm, around sunset. As previously mentioned, this is central to determining the price since at its peak of energy production we can see a reduction in the price of electricity, while at the end of the day, the absence of production is associated with a peak in the price.

4.3 Weekly Patterns

It was important to understand how price and consumption related during the week and the Figure 4.3 shows the average price per day of the week for Portugal and Spain. We can see that both countries show a consistent pattern and that prices tend to be higher on weekdays, especially on Tuesday, when they peak. From Wednesday and Thursday onwards there is a slight drop in prices, which is significantly accentuated at the weekend, reaching its lowest point on Sunday.

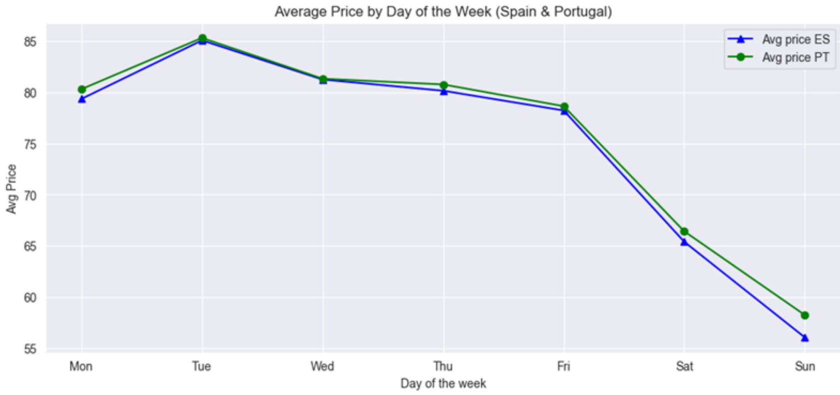

Figure 4.3: Average Price by day of the week

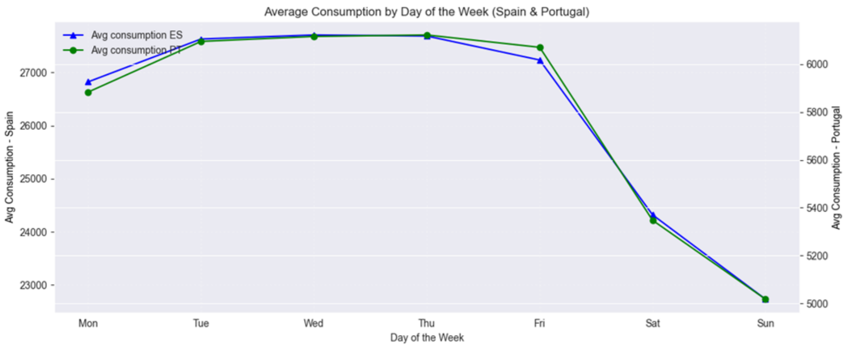

Figure 4.4: Average Consumption by day of the week

This difference in prices between weekdays and weekends may be related to the demand for energy during the week, associated with labour activity, while at the weekend it is reduced, resulting in a drop in prices.

The Figure 4.4 shows the average consumption per day of the week in both countries and, similar to the prices, both Portugal and Spain show higher values during working days, with a drop at the weekend. As previously mentioned, the value of consumption during the week remains stable, reflecting labour activity, and there is a clear reduction on Saturdays and Sundays. This can be explained by the decrease in economic activity and consequently lower intensity of energy use.

4.4 Monthly / Seasonal Patterns

Figure 4.5 reflects the monthly variation in average electricity prices in Portugal and Spain and reveals the strong harmonisation of the Iberian market. It can be seen that the highest values are recorded in winter, between December and February, and in summer, between July and September, coinciding with the periods of greatest demand and energy consumption, as shown in Figure 4.6. Once again, this behaviour can be reflected by the increased pressure on the electrical system, resulting from the increased use of air conditioning systems, both in the cold and hot months.

On the other hand, it is noticeable that in spring, between the months of March and May, there is a minimum. This can be explained by the period of lower energy demand, considering

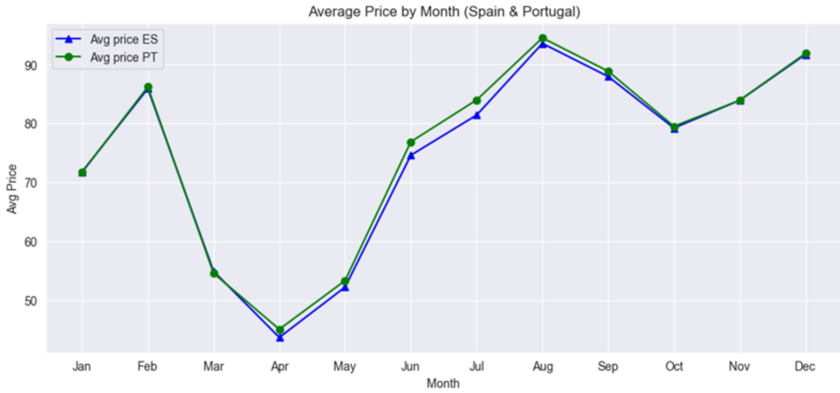

Figure 4.5: Average Price by month

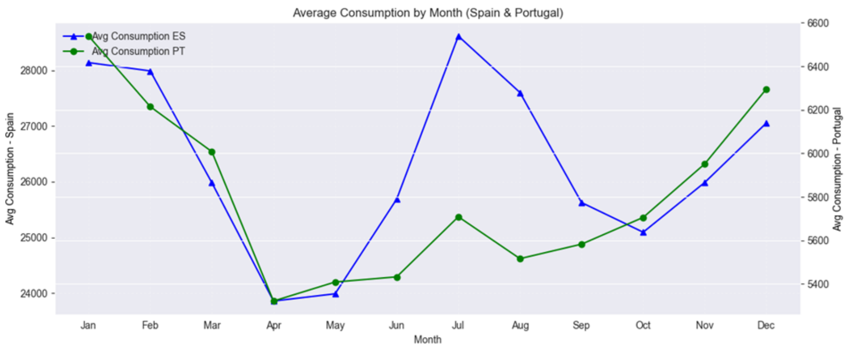

Figure 4.6: Average Consumption by month

temperatures are already moderate and heating systems may not be as necessary. From these graphs we can conclude that market prices are not only related to demand and supply, but also to seasonality and weather conditions.

Figure 4.6 shows the average monthly consumption in both countries, and as previously mentioned, consumption remains high in the winter months and in the summer months due to the need to use air conditioning systems. This graph is the first where we can see a discrepancy between the two countries because in the summer months, Spain shows a significant increase in energy consumption, while in Portugal the increase is milder. Similarly, it is in the spring months that the lowest values are recorded, which can be explained by the more favourable weather conditions, resulting in a reduction in pressure on the energy system. We can conclude that this correlation between the periods of greatest energy consumption, winter and summer, and price peaks, confirms that economic activity and climatic conditions are strongly interdependent in the context of the Iberian electricity system.

4.5 Correlation Analysis

Analysing correlations helps to identify potential reactions between the main variables of the Iberian market system. Firstly, it can be seen, by the Figure 4.7 that there is a strong correlation between the prices of Portugal and Spain, of around 0.99, which shows the high degree of integration of MIBEL, where both countries practise similar prices most of the time.

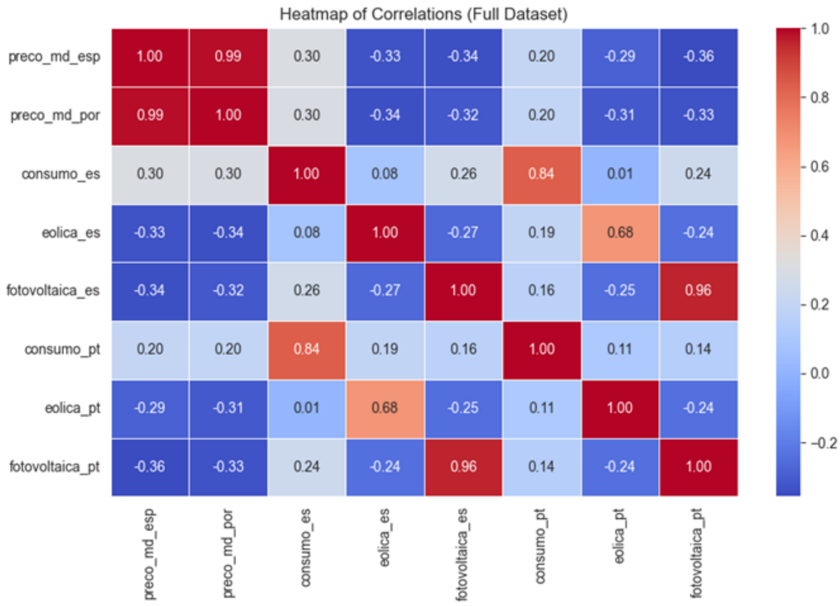

Figure 4.7: Heatmap of Correlations

Regarding consumption, there is a moderate positive correlation between Spain's price and its consumption, of around 0.3, which could lead one to understand that increases in demand tend to be reflected in higher prices. In the case of Portugal, the correlation between price and consumption is weaker, at around 0.2. However, consumption between the two countries shows a strong positive correlation of around 0.84, which corroborates the synchronisation of demand dynamics between the two countries.

As far as renewable energies are concerned, the negative correlations between the price and production of renewable sources corroborate what was already concluded in the previous graphs. The correlations between the price and production of renewable energy in each country, with -0.33 and -0.34 between wind and photovoltaic production in Spain and the Spanish price, and -0.31 and -0.33 in the case of Portugal, indicate that an increase in this production results in lower prices. The strong correlation between the same kind of renewable production in each country is another important finding. Wind and photovoltaic production in each nation have a correlation of 0.68 and 0.96, indicating that they are both strongly correlated and subject to comparable weather conditions.

Chapter 5

The model

The baseline of this dissertation will be on the fARX model proposed by Uniejewski et al. (2016), which is used to predict the price of electricity for each hour of the following day, taking into account multiple factors. This model manages to be more comprehensive than other basic autoregressive models such as Naive Benchmark and Autoregressive Expert Benchmarks, also mentioned in the article Uniejewski et al. (2016) including all the regressors that have significant predictive power. The main aim of this model is to aggregate the complex relationships between electricity prices, external variables and seasonality, thus creating a composite basis for future variable selection and reduction procedures.

Before we move on to the model itself, it's important to explain the variable we're going to predict, the price of electricity on the daily market. Firstly, we will use its logarithm in order to stabilise the variance, but instead of using this value alone, the article proposes refocusing the values using the formula 5.1,

$$p_{d,h} = \log(P_{d,h}) - \frac{1}{T}\sum_{t=1}^{T}\log(P_{t,h}) \tag{5.1}$$

Where $P_{d,h}$ represents the price on day d and time h, and T , the number of days in the calibration window. This formula corresponds to the logarithm of the price centred on the historical average for each hour so that the series has a zero average.

The model allows us to focus on variations around these expected values since daily seasonality, where, for instance, we see lower prices in the early morning and higher prices at peak times, is removed. Using centered pricing additionally offers the benefit of removing the model's intercept, thus ensuring that the estimated variables simply represent temporal dynamics and preventing structural variations across time periods from influencing the estimation.

The full autoregressive model can be explained by the following Formula 5.2 , where prices from the last days are used to capture temporal dependencies, and the averages, minimums, and maximums of prices from previous days are used to capture aggregate effects. The model also includes exogenous variables, such as wind power forecasts, which influence the price of electricity and also seasonal dummies, i.e., helps to identify which day of the week is and if it is a holiday or not, so that the model can adjust the forecasts to the specific consumption and price patterns of each day and date.

$$
\begin{aligned}
p_{d,h} = &\sum_{i=1}^{24} \left(\beta_{h,i} p_{d-1,i} + \beta_{h,i+24} p_{d-2,i} + \beta_{h,i+48} p_{d-3,i}\right) + \beta_{h,73} p_{d-7,h} \\
&+ \sum_{j=1}^{3} \left(\beta_{h,j+73} p_{d-j}^{\min} + \beta_{h,j+76} p_{d-j}^{\max} + \beta_{h,j+79} p_{d-j}^{avg}\right) \\
&+ \beta_{h,83} z_{d,h} + \beta_{h,84} z_{d-1,h} + \beta_{h,85} z_{d-7,h} + \beta_{h,86} y_{d,h} \\
&+ \sum_{k=1}^{7} \beta_{h,86+k} D_k + \sum_{k=1}^{7} \beta_{h,93+k} D_k z_{d,h} + \sum_{k=1}^{7} \beta_{h,100+k} D_k p_{d-1,h} + \varepsilon_{d,h}
\end{aligned}
\tag{5.2}
$$

5.1 The 107 variables

We will now describe in detail what the variables are in order to implement the model, according to the article.

Firstly, the most abundant category, with seventy-three variables, is the electricity past prices, which capture historical prices on different days and at different times. They use the logarithmic transform to make the price series more symmetric. In other words, it comprises:

- Day $d - 1$ (yesterday): including the twenty-four hours of the previous day
- Day $d - 2$ (the day before yesterday): also including 24-hour prices
- Day $d - 3$: Same as $d - 1$ and $d - 2$
- Day $d - 7$: with only one variable, which will be the price at the same time last week.

The authors consider that these regressors possess a non-negligible predictive power, and so we do not use more past prices.

Next up are the daily aggregates, i.e. we will use the minimum, maximum and average prices of the last three days, totalling nine variables. In other words, for each day ($d-1$, $d-2$, $d - 3$), we will have:

- Minimum price: the lowest price reached that day
- Maximum price: the highest price reached that day
- Average price: the average price over the twenty-four hours of that day

In addition to these variables, it is important to include exogenous variables in the model, such as the logarithm of load forecasts and factors that affect prices, such as wind energy forecasts. To summarise:

- $z_{d,h}$: Hourly load forecast for the same day and time
- $z_{d-1,h}$: Load forecast for the same hour as the previous day
- $z_{d-7,h}$: Load forecast for the same time last week
- $y_{d,h}$: Wind power forecast for the same day and time

It makes sense to use the system load and consumption forecast as exogenous variables, as these are directly related to consumption, as we have already concluded previously in this report.

To make the model even more interconnected, the authors believe it is important to use dummies to capture the seasonal patterns on the days of the week, as the energy markets often show specific behaviour depending on the day, such as lower electricity consumption in industries at weekend. In other words, we will have a dummy for each day of the week, where:

- For a specific day of the week, the dummy variable will take on the value for the corresponding day and zero otherwise

- For a public holiday, all dummies will take the value zero so that the model realises that it is a holiday

Finally, the authors consider that, as well as having dummies, it is crucial to capture the interaction between the days of the week and determining variables, such as the consumption and the price. Thus, the dummy of the day we want to predict it will interact with:

- Consumption forecast of the day we want to predict ($z_{d,h}$)

- Previous day hour price we want to predict ($p_{d-1,h}$)

5.2 Adapting to the MIBEL market

As explained above, the MIBEL market follows a marginalist model, in which prices are determined by the point of intersection between the supply and demand curve and is organised based on marginal production costs and the requirement is based on suppliers' purchase prices.

Based on the data analysed, it can be seen that hourly electricity prices in Portugal and Spain are, in most cases, identical, with discrepancies occurring in only six percent of cases. Given this degree of convergence, it has been decided to forecast hourly electricity prices only for the Spanish market.

The main factors influencing the price are the variables related to consumption in Portugal and Spain, the forecasts for renewable energies, solar and wind, in both countries.

Therefore, in order to adjust the fARX model to the MIBEL market, we must include:

Past prices

Similar to the original model, we will use the prices from Spain for the 24 hours of the previous days ($p_{d-1,h}$; $p_{d-2,h}$; $p_{d-3,h}$) and the price for the same hour a week ago ($p_{d-7,h}$). Another detail to bear in mind is that, in the Iberian market, there are cases where supply exceeds demand, resulting in negative prices. Since logarithmic transformation is only possible with positive values, these negative prices will be treated as missing values and assigned small values, in this case, one cent.

Exogenous variables

As explained above, since only the price in Spain will be forecast, and this is also influenced by consumption in Portugal, it was decided to aggregate the consumption of both countries into a single variable. Therefore, we will have three variables: consumption forecast, solar and wind consumption.

Seasonal effects

As far as dummies are concerned, we will maintain what the formula says, and we will have a dummy for each day of the week and have the interaction between the dummy and the consumption and the dummy and the price of the previous day.

5.3 Elastic Nets

After describing the reference model, fARX, which incorporates a vast set of explanatory variables, in this case 108, it is important to consider methods that reduce its complexity while increasing its predictive power. When using more robust models, such as the one under study, the probability of having uninformative explanatory variables is high, which can hinder interpretability and pose a risk of overfitting.

In order to combat these problems, we will use the method that obtained the best results in the article, which is Elastic Nets. This is a linear regression technique that combines the penalties of Lasso and Ridge Regression and acts as a shrinkage process, which reduces the magnitude of the coefficients, and a selection process, which eliminates variables that have little contribution to the prediction. This method was proposed in 2005 by Zou and Hastie (2005), which can be explained by the Formula 5.3, which uses a mixture of linear and quadratic penalty factors. α takes values between 0 and 1, and when α=1, elastic nets is reduced entirely to lasso, and when α=0, it becomes entirely ridge regression. The article believes that using ½ in the quadratic penalty part makes the optimisation process more efficient and intuitive. In addition to this parameter, it also uses a regularisation parameter, λ, which determines the regularisation strength of the coefficients. When it has a null value, no regularisation is applied, causing the coefficients to grow, leading to overfitting, while a high λ reduces the coefficients to practically zero. The methodology will explain in more detail how this process will be implemented.

$$\hat{\boldsymbol{\beta}}^{EN} = \underset{\beta_{h,i}}{\operatorname{argmin}} \left\{ \text{RSS} + \lambda \left(\frac{1-\alpha}{2} \sum_{i=1}^{n} \beta_{h,i}^2 + \alpha \sum_{i=1}^{n} |\beta_{h,i}| \right) \right\} \tag{5.3}$$

5.4 Methodology

This process can be divided into several stages.

The first process was creating the reference model, fARX, in which the consumption variables of each country were aggregated into a single variable, followed by the remaining variables proposed by the model, totalling 108 variables. There is a difference of one variable, as an exogenous variable, photovoltaic consumption, was added.

Once the database was built, the next step was to divide it into three sets: training, validation, and testing, as can be seen in the Figure 5.1.

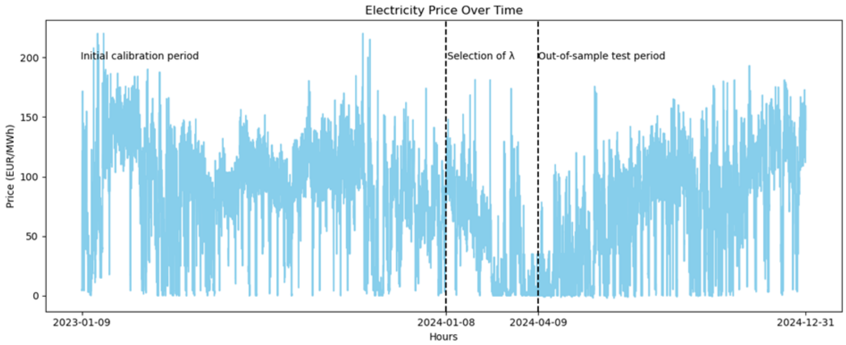

Figure 5.1: Electricity price over time

- The **training set**, covering a period of 365 days between 9 January 2023 and 8 January 2024, will be used to train the model and estimate the coefficients for each lambda. This training set will not be fixed, i.e. this initial date will be used to predict the first day of the test, and for the remaining days, the last 365 days of the day we want to predict will be used as training.

- The **validation set**, which corresponds to the 91 days following the training period, between 09/01/2024 and 08/04/2024, will be used to determine the optimal value of the regularisation parameter for each value of alpha. The choice will be made from a grid of exponentially-decreasing λ's, and the choice will be made based on the one that obtains the lowest WMAE value.

- After determining the regularisation parameter, in this last **test set**, which covers the remaining period of the database, between 09/04/2025 and 31/12/2024, the purpose will be to evaluate the final performance of the model. This process will be carried out using a sliding window with a period of 365 days, in which the window will be moved one day forward for each forecast for the following day, and repeated until the test day.

Regarding the model chosen for comparison, XGBoost, it will be implemented in the same way as Elastic Nets, in order to ensure a fair comparison. More specifically, this model will predict the daily market price, using a 365-day rolling window to train the model and predict the following day. The data will also be divided into training, validation and testing, based on WMAE. For each day we predict, the model will be trained, generate predictions and calculate the importance of the variables, after which the predictions will be converted to the original level for future analysis.

The main goal of this study will be to compare the performance of four models, three of which are Elastic Net with three different values of 0.25, 0.5, and 0.75, and one machine learning model, such as XGBoost. All forecasts and analyses will be based on the WMAE, weekly weighted average absolute error, error to identify the most effective model.

WMAE, according to Uniejewski et al. (2016), is a robust measure similar to MAPE, Mean Absolute Percentage Error, however the difference is that the normalised absolute error is based on the weekly average price in order to avoid the adverse effect of electricity prices close to zero. Next, the errors are aggregated into an average value over all weeks of the test period, as indicated by the Formulas 5.4 and 5.5.

$$\text{WMAE}_w = \frac{1}{\bar{P}_{168}}\text{MAE}_w = \frac{1}{168 \cdot \bar{P}_{168}} \sum_{d=\text{ Mon}}^{\text{Sun}} \sum_{h=1}^{24} \left| P_{d,h} - \hat{P}_{d,h} \right| \tag{5.4}$$

$$\overline{\text{WMAE}} = \frac{1}{w_{\max}} \sum_{w=1}^{w_{\max}} \text{WMAE}_w \tag{5.5}$$

It is crucial to note that the error is determined using the current price for time h, $P_{d,h}$ rather than the centered price, $p_{d,h}$. The expected price must therefore be reversed by adding the average and using the exponential to the price's logarithm before determining the error.

Chapter 6

Perfomance evaluation

This section aims to analyse the forecasts obtained by Elastic Nets, observing the model's performance in both the validation and test periods and checking for possible patterns. Subsequently, the performance of the XGBoost model will be analysed, making it possible to compare the performance of the two models.

6.1 Elastic Models

6.1.1 Variation in error between validation and testing

An initial analysis showed that the error obtained by the WMAE increased significantly from around 0.3 in the validation phase to around 1 in the test phase (6.1). This difference made it necessary to investigate whether the reason for this difference was related to the metric itself or to the nature of the data.

alpha	WMAE_val	MAPE_val	MAPE_val_excl<1	WMAE_test	MAPE_test	MAPE_test_excl<1
0.2500	0.3227	20.6212	0.3191	1.0110	27.2046	1.1886
0.5000	0.3266	20.5929	0.3203	1.0173	26.5423	1.1963
0.7500	0.3297	19.9919	0.3138	1.0364	24.9228	1.2002

Table 6.1: Performance metrics for different alpha values.

We therefore began by analysing the descriptive statistics for each set and found that in the validation phase, the average price was around 40€/MWh and the median 35€/MWH, while in the test phase the average and median values rose to 70€/MWh and 77€/MWh respectively (6.1). In addition, there was an increase in the dispersion of the results and a widening of the interquartile range, which could lead to the conclusion that there were relatively higher values, as the graphs show.

In order to complement this research, another error metric was used, namely MAPE. However, as this metric is sensitive to prices close to zero, as the values explode at low values, and due to the existence of more than 500 observations close to zero in both sets, it was decided to consider 2 versions. The first was to use all the values, including those close to zero, and the other without them.

Table 6.1 shows that the cause of the increase in error is not related to the metric itself, but to the nature of the data or other reasons, such as using a fix lambda instead of searching

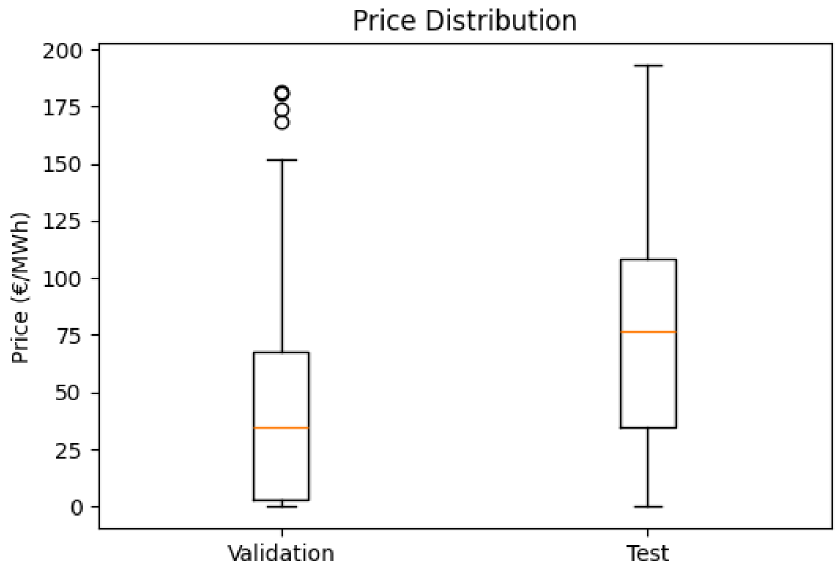

Figure 6.1: Price Distribution

the best lambda everytime we predict.

The model with the lowest error was the one with $\alpha = 0.25$, and as α increases, the error also increases slightly. The figure 6.2 compares the various WMAEs for each alpha value, and we can see that they all obtained very similar values.

This behaviour indicates that it is better to wager on reducing the coefficients rather than forcing them to zero for this type of data, and that lower α values generally yield slightly better predictions.

6.1.2 Error analysis during test

Focusing on the error during the test phase, we first chose to analyse the error by day and week, in order to understand how the error varies and to perceive possible general patterns.

The graph 6.3 shows the average absolute error per hour and it can be seen that during the morning there is a progressive drop in the error until it reaches its minimum between 1pm and 2pm, at around 35€/MWh. However, after this period, it quickly rises again, peaking between 6pm and 8pm at 120 €/MWh. These results indicate that at peak demand, where there is greater price volatility, the model has greater difficulty predicting the price.

A different pattern was obtained for the MAE per week 6.4. We can see that the values remain relatively constant during the working days, at around 80, 85€/MWh, but there can be a sharp drop, especially on Sunday, where the error is 53€/MWh. This behaviour can be explained by the fact that during the week there is more economic activity and consequently more fluctuations in prices, which can be a challenge for the model, while weekend patterns are easier to predict.

By comparing the correlation between the time of day and the day of the week between the validation and test periods, it was determined reasonable to identify patterns in errors across time and comprehend the challenges during particular times.

In the validation set, 6.5, the predominant colour is blue, which means that for most of the period, errors remain relatively low, with values between €2 and €10/MWh. However,

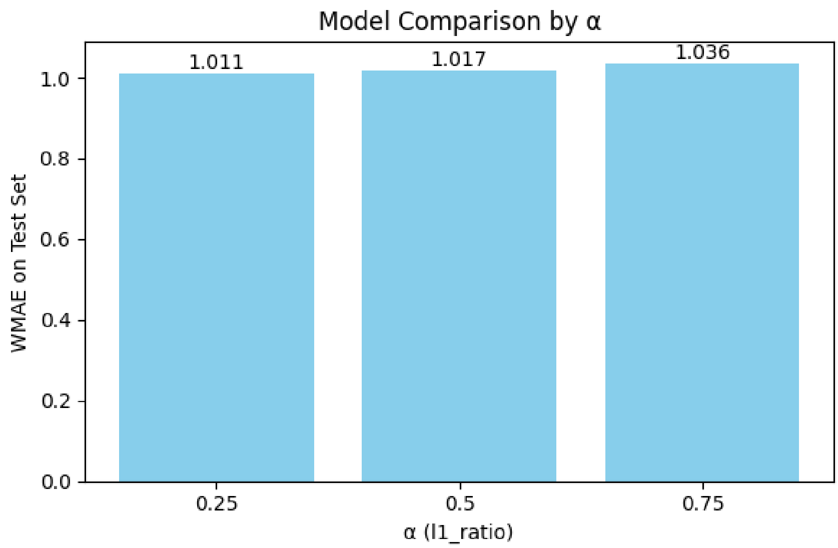

Figure 6.2: Model Comparison by α

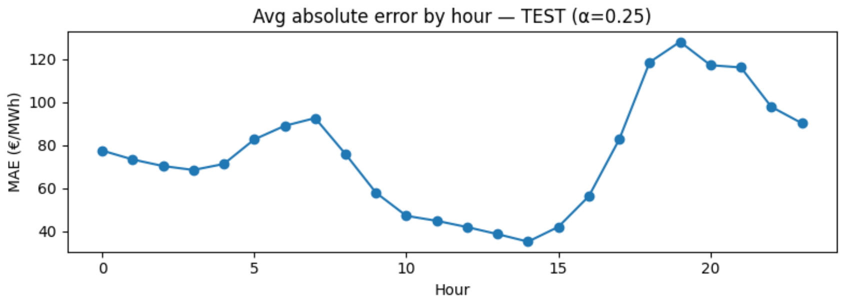

Figure 6.3: Average MAE by hour

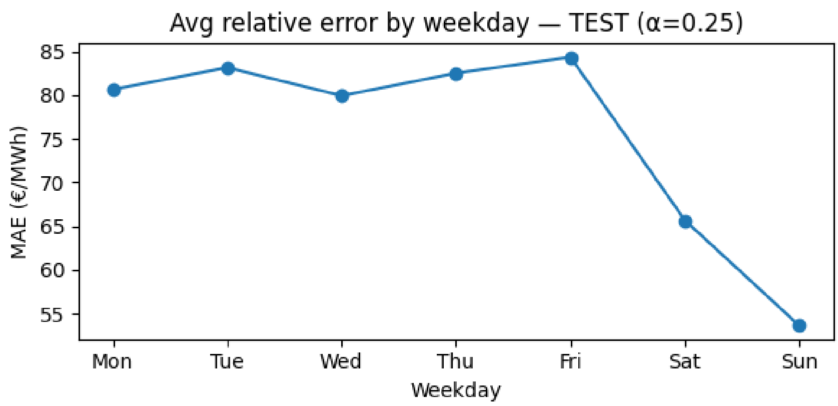

Figure 6.4: Average MAE by weekday

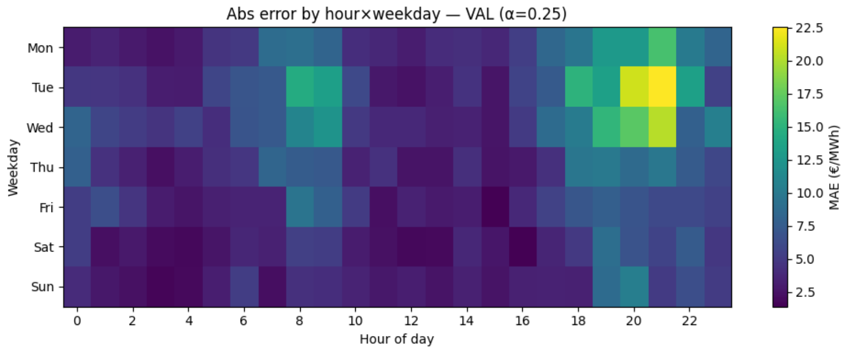

Figure 6.5: Absolute error between hour of day and weekday in Validation

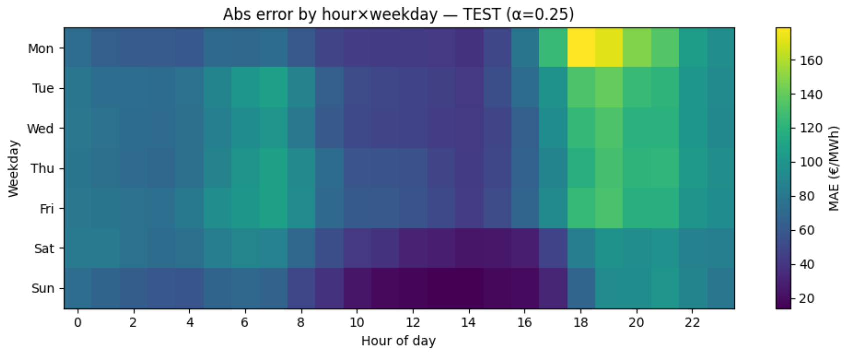

Figure 6.6: Absolute error between hour of day and weekday in Test

the error increases significantly in the late afternoon and early evening, between Tuesday and Wednesday, with the peak occurring on Tuesday between 7 p.m. and 9 p.m., when the MAE exceeds €20/MWh. This behaviour indicates that the model has greater difficulty capturing price peaks, especially at night on weekdays.

However, in the test set, 6.6, the 'scenario' becomes greener, as the average error is significantly higher, with values exceeding €160/MWh. In this case, the analysis reveals that from Monday to Friday, the error increases at the end of the day, between 6 p.m. and 8 p.m., especially on Mondays and Tuesdays, while in the morning, on weekends, the error is predominantly lower.

6.1.3 WMAE analysis

Proceeding on to the WMAE, we made the decision to observe the error's evolution throughout the testing phase.

The weekly WMAE graph 6.7first shows that the values were comparatively low in April and May, ranging between 0.6 and 0.8. Then, beginning at the end of June, there is a steady increasing trend, with the error surpassing the 1 threshold and remaining stable until the middle of September. From that month onward, the values begin to rise again, reaching a peak of approximately 1.3 at the end of October, but then begin to decline slightly until December.

The monthly behavior, illustrated in the figure 6.8, confirms the trend observed in the previous graph. We initially observed low values between May and June, followed by a sharp

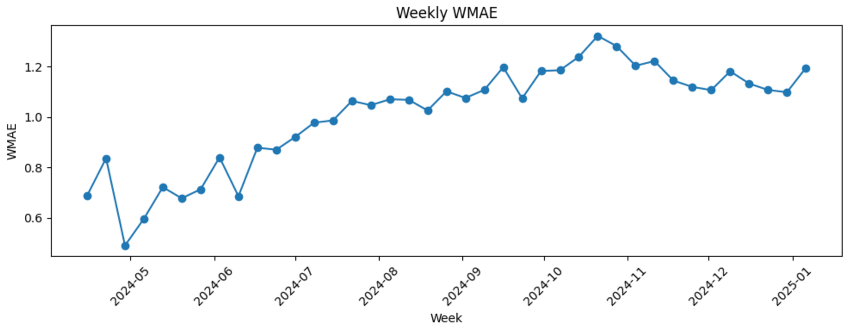

Figure 6.7: Weekly WMAE

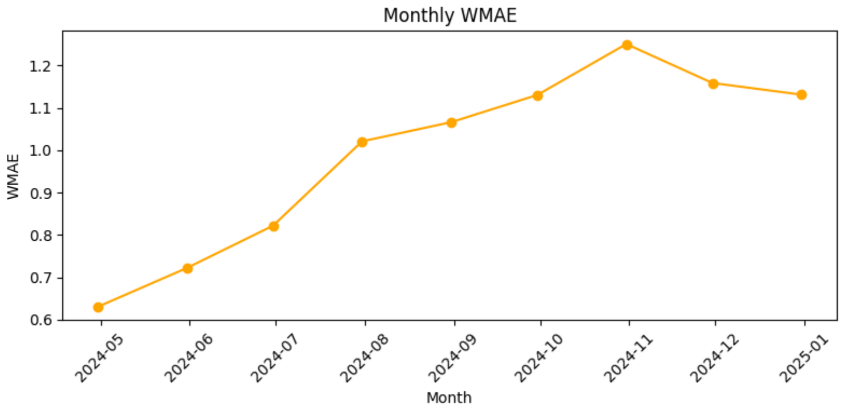

Figure 6.8: Monthly WMAE

rise between July and August, exceeding unity. On average, during the months of September and October, a steady increase is observed, around 1.1, reaching its peak in November (1.2). The two graphs show that the model degrades over the months, which may be reflected in price changes, with higher levels and, consequently, greater volatility. This may suggest that, despite the re-estimation of the coefficients, the model with fixed λ may not adapt to market fluctuations.

6.1.4 Importance of the variables in the test period

Analysing the coefficients of the model under study over the test period makes it possible to understand which variables are used the most and have the greatest weight in the forecast. As explained during the study, for each day of the forecast, the coefficients were re-estimated in a sliding window of 365 days, thus allowing us to calculate the frequency of selection, in percentage of days in which the coefficient was different from zero, and its magnitude, in which the average absolute value of the coefficient was calculated.

From the results of the x and y graphs, we can see that the model used a set of variables that were used consistently throughout the test period. In particular, the price of the previous hour (price lag 1) was selected on every day and had an average coefficient of more than 1.2, making it possible to conclude that this was the main variable for the forecast. Other variables, such as price lag 22, price lag 23, price lag 26 and price lag 27, were also selected almost every day, but with a lower magnitude.

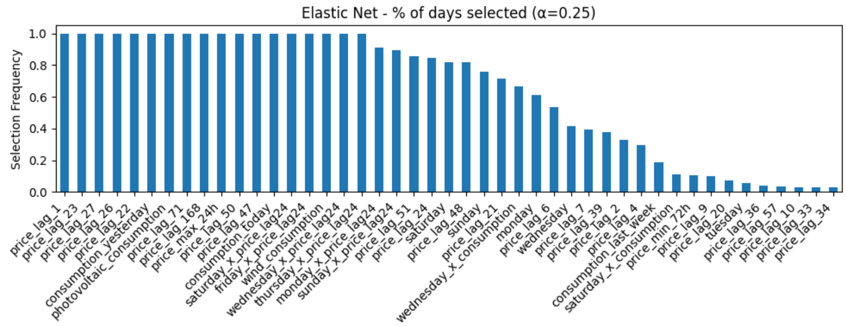

Figure 6.9: Percentage of days selected

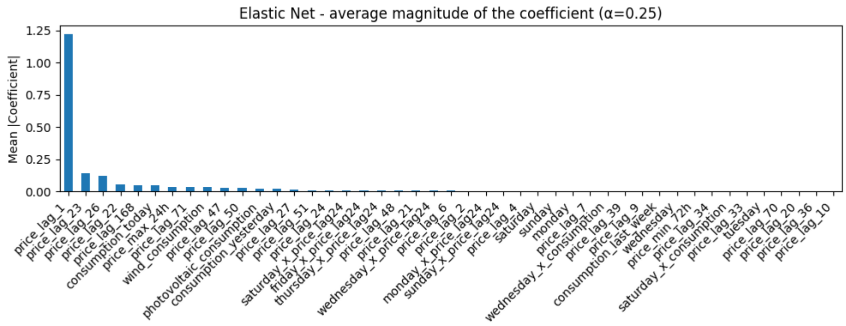

Figure 6.10: Average magnitude of the coefficient

Moving on to the exogenous variables, we can see that consumption and renewable production, photovoltaic and wind power, were among those selected most often, which means that the model also considers them relevant for price forecasting. However, from the graph of magnitudes, we can understand that they were used more as an adjustment factor, since the coefficients are practically zero.

6.2 XGBoost

This section will present the results obtained by the XGBoost model. The average weekly error will be analysed in comparison with the Elastic Nets model, as well as the importance of the variables for evaluating the model's performance.

Contrary to the values presented above, according to the table we can see that the WMAE value confirms the good performance of XGBoost, with 0.13, which is significantly lower than the value obtained by Elastic Nets of 1.01. This difference may show that XGBoost manages to considerably reduce the WMAE, bringing it closer to the real price values.

WMAE
Elastic Nets: 1.010950260085351
XGBoost: 0.12860669682437673

Table 6.2: WMAE values for Elastic Nets and XGBoost

The chart 6.11 shows that the XGBoost error remains low and stable throughout the test period, with few fluctuations, although the same cannot be said for the other model. This

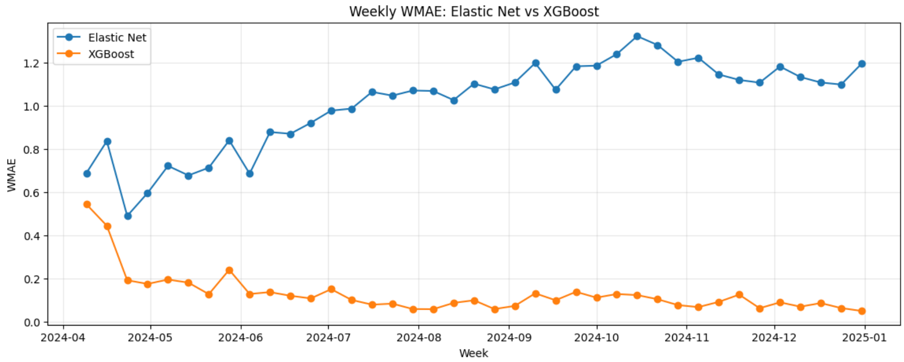

Figure 6.11: Weekly WMAE: Elastic Net vs XGBoost

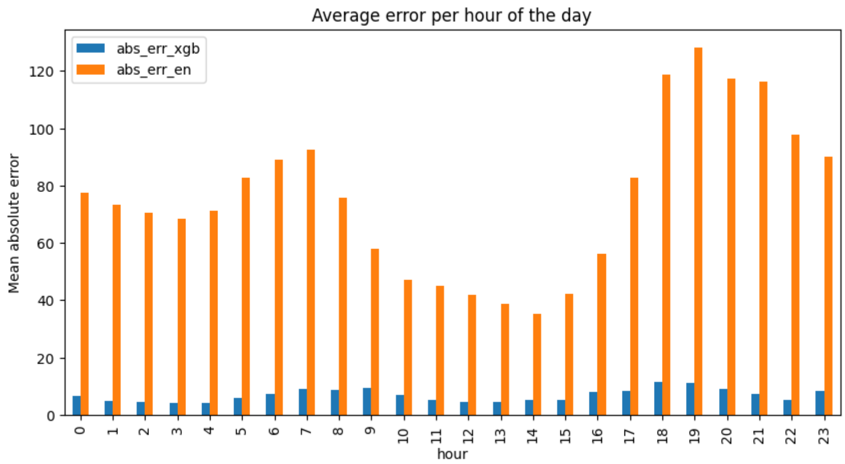

Figure 6.12: Average MAE per hour of the day EN vs XGBoost

low and stable value shows that the model is capable of continuing to predict well in the face of market variations, without losing accuracy. Elastic Nets, on the other hand, shows higher values and an upward trend, suggesting a loss of forecasting ability.

The graph below shows the average absolute error per hour of the day between the two models studied. Based on the previous analysis, we can see that the error obtained by XGBoost is substantially lower and more stable at all times of the day, demonstrating greater stability. Elastic Nets, on the other hand, has much higher and more unstable values, which are relatively lower between 10am and 3pm, but much higher in the late afternoon and early evening, when prices tend to be more volatile.

The following graph, 6.13, represents the average importance of the variables used for forecasting and, similar to Elastic Nets, the model gives significant weight to the previous hour's price variable, price lag 1, which represents more than 50 percent of the total importance.

Next comes price lag 2, which is much less important at around 0.1, followed by price mean 24h, which represents the daily average of the previous day's price. The remaining lags and exogenous variables are practically of residual importance, indicating that the model depends more on the information contained in the prices to generate its forecasts.

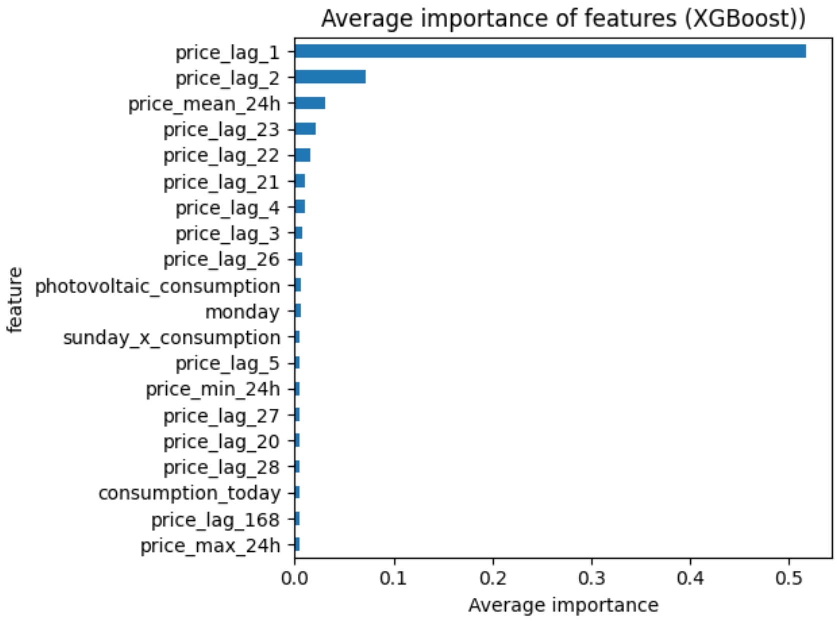

Figure 6.13: Average importance of features XGBoost

Chapter 7

Conclusion

The purpose of this study was to forecast the daily price of electricity on the MIBEL market, using an autoregressive model, fARX, as the base model, with regularisation methods such as Elastic Nets, and a model based on decision trees, XGBoost, as the comparison model.

In the validation phase, where the coefficients were re-estimated and the best lambda was chosen from a group of 60 values, the model managed to select the most relevant variables, presenting satisfactory values, with a WMAE close to 0.3. The model's performance, however, dropped over the test period, reaching an average WMAE value of 1.01. This suggests that when the λ parameter is fixed, the model tends to struggle more to adjust to market fluctuations, particularly during times of higher volatility.

On the other hand, XGBoost proved to be firmer in the face of seasonal price variations, with an average WMAE of 0.13 over the test period. According to the analysis, both models rely heavily on the previous hour, however the latter model is better at taking advantage of non-linear dependencies resulting in more accurate predictions.

Although the model under study has degraded over time, I believe it is still a good relevant model and future research could explore ways to improve its performance, such as a dynamic λ, where it is estimated every x period of time, rather than remaining fixed. With this implementation, I believe you can adapt more easily to changes in market behaviour. To improve price forecast, another implementation would be to incorporate forecasting variables for the consumption of renewable energy. In short, it is necessary to emphasise the importance of this type of flexible and adaptive predictive model in markets with high volatility, such as MIBEL. XGBoost proved to be the most accurate approach, however, with the necessary improvements, I believe that Elastic Nets could continue to be an efficient alternative.

Bibliography

Annual Report 2023. (2023). https://www.omie.es/sites/default/files/2024-02/Informe%20Anual%202023%20EN_0.pdf. ([Accessed 04-02-2025])

Bitirgen, K., & Filik, Ü. B. (2020). Electricity price forecasting based on xgboost and arima algorithms. *BSEU Journal of Engineering Research and Technology*, *1*(1), 7–13.

Cerasa, A., & Zani, A. (2025). Enhancing electricity price forecasting accuracy: A novel filtering strategy for improved out-of-sample predictions. *Applied Energy*, *383*, 125357. Retrieved from https://www.sciencedirect.com/science/article/pii/S030626192500087X doi: https://doi.org/10.1016/j.apenergy.2025.125357

Contreras, J., Espínola, R., Nogales, F. J., & Conejo, A. J. (n.d.). *ARIMA Models to Predict Next-Day Electricity Prices.* https://halweb.uc3m.es/esp/Personal/personas/fjnm/esp/papers/arimaprices.pdf. ([Accessed 04-02-2025])

Díaz, J., Álvaro Romero, & Dorronsoro, J. R. (2019, 03/2019). Day-ahead price forecasting for the spanish electricity market. *International Journal of Interactive Multimedia and Artificial Intelligence*, *5*(4), 42-50. Retrieved from http://www.ijimai.org/journal/sites/default/files/files/2018/04/ijimai_5_4_5_pdf_14997.pdf doi: 10.9781/ijimai.2018.04.008

Garcia, R., Contreras, J., van Akkeren, M., & Garcia, J. (2005). A garch forecasting model to predict day-ahead electricity prices. *IEEE Transactions on Power Systems*, *20*(2), 867-874. doi: 10.1109/TPWRS.2005.846044

Kath, C., & Ziel, F. (2018). The value of forecasts: Quantifying the economic gains of accurate quarter-hourly electricity price forecasts. *Energy Economics*, *76*, 411-423. Retrieved from https://www.sciencedirect.com/science/article/pii/S0140988318304018 doi: https://doi.org/10.1016/j.eneco.2018.10.005

Kristiansen, T. (2012). Forecasting nord pool day-ahead prices with an autoregressive model. *Energy Policy*, *49*, 328-332. Retrieved from https://www.sciencedirect.com/science/article/pii/S0301421512005381 (Special Section: Fuel Poverty Comes of Age: Commemorating 21 Years of Research and Policy) doi: https://doi.org/10.1016/j.enpol.2012.06.028

Mao, X., Chen, S., Yu, H., Duan, L., He, Y., & Chu, Y. (2025). Simplicity in dynamic and competitive electricity markets: A case study on enhanced linear models versus complex deep-learning models for day-ahead electricity price forecasting. *Applied Energy*, *383*, 125201. Retrieved from https://www.sciencedirect.com/science/article/pii/S0306261924025856 doi: https://doi.org/10.1016/j.apenergy.2024.125201

Mercado ibérico da energia elétrica – mibel. (n.d.). https://www.mibel.com/. ([Accessed 04-02-2025])

OMIE. (n.d.). *Omie.* https://www.omie.es/es/mercado-de-electricidad. ([Accessed 04-02-2025])

Uniejewski, B., Marcjasz, G., & Weron, R. (2019). Understanding intraday electricity markets:

Variable selection and very short-term price forecasting using lasso. *International Journal of Forecasting*, *35*(4), 1533-1547. Retrieved from https://www.sciencedirect.com/science/article/pii/S0169207019300123 doi: https://doi.org/10.1016/j.ijforecast.2019.02.001

Uniejewski, B., Nowotarski, J., & Weron, R. (2016). Automated variable selection and shrinkage for day-ahead electricity price forecasting. *Energies*, *9*(8). Retrieved from https://www.mdpi.com/1996-1073/9/8/621 doi: 10.3390/en9080621

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *67*(2), 301–320.

Appendix A

Appendix

A.1 Total Hierarchy

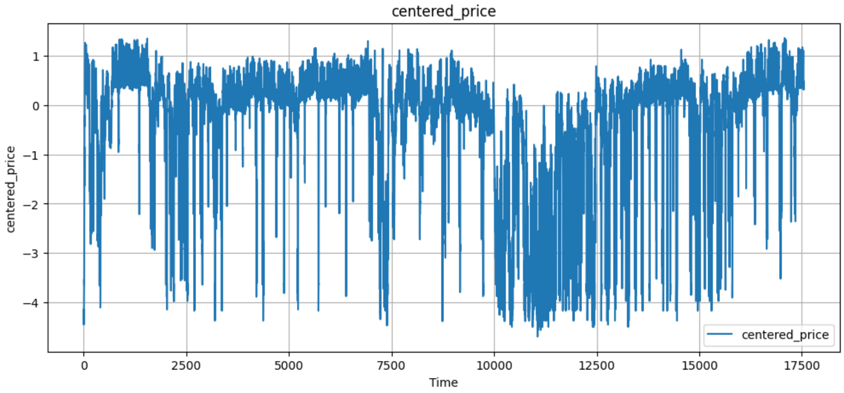

Figure A.1: Centered price

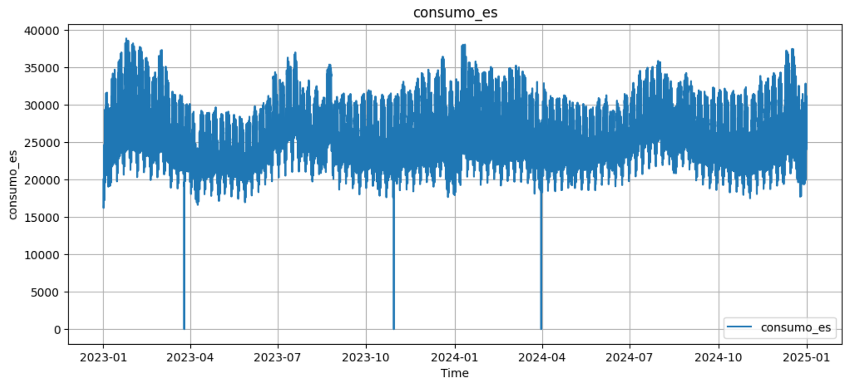

Figure A.2: consumo es

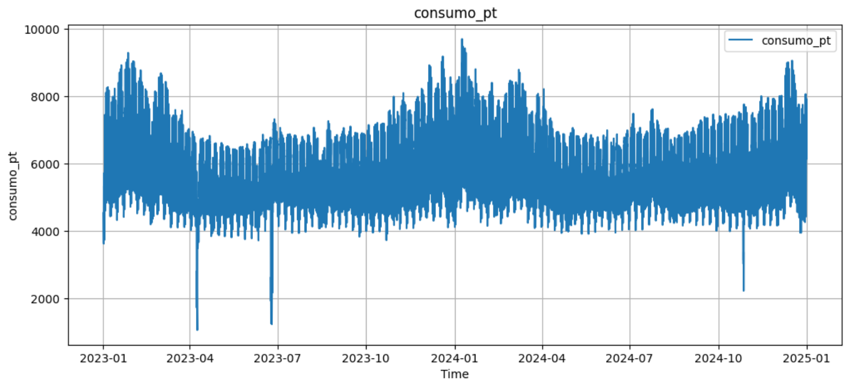

Figure A.3: consumo pt

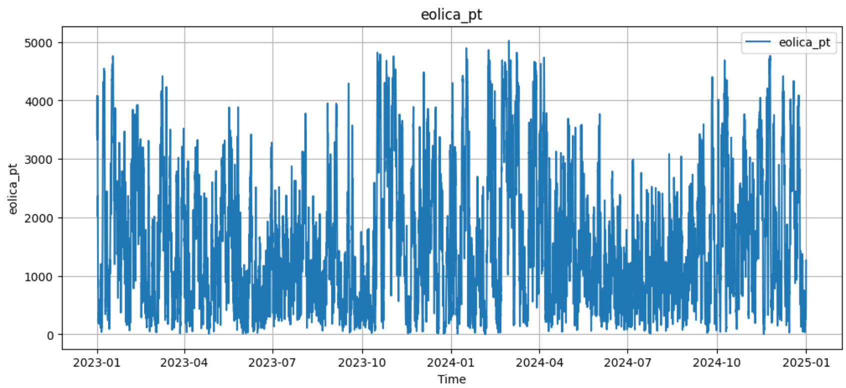

Figure A.4: eolica pt

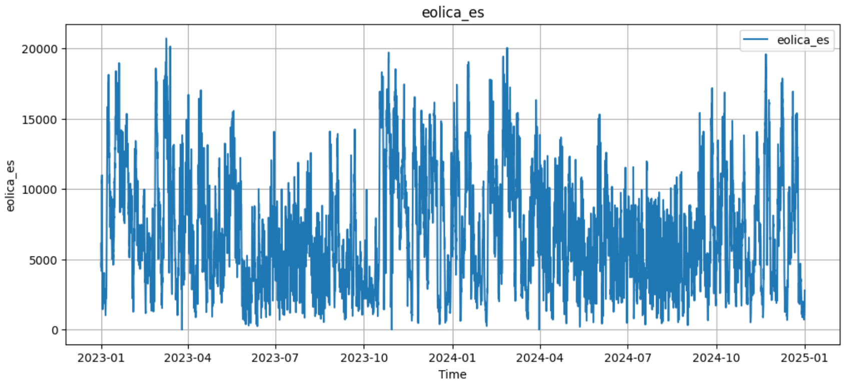

Figure A.5: eolica es

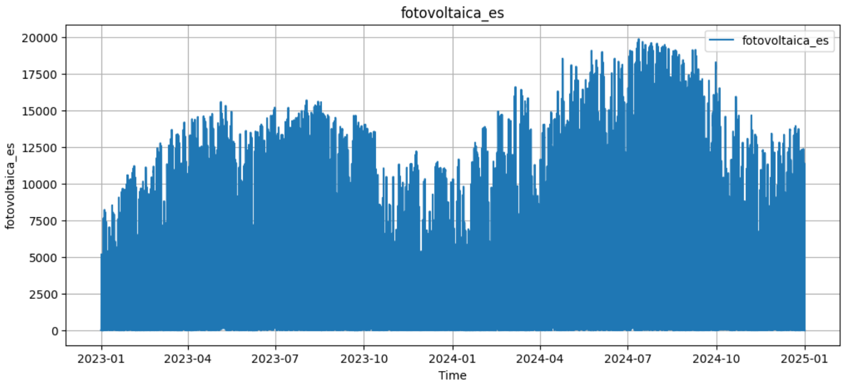

Figure A.6: fotovoltaica es

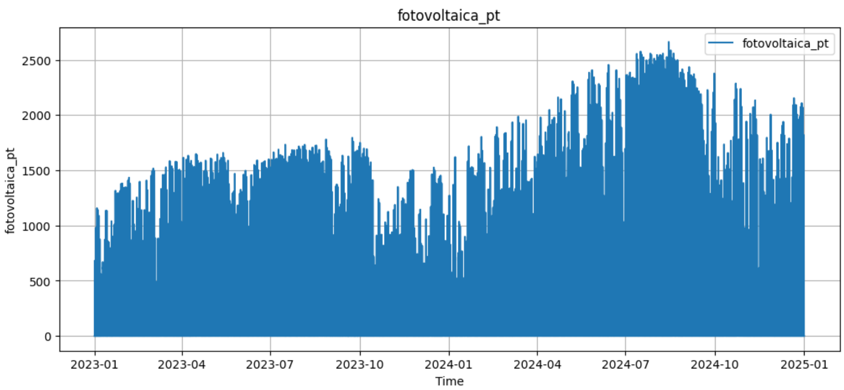

Figure A.7: fotovoltaica pt

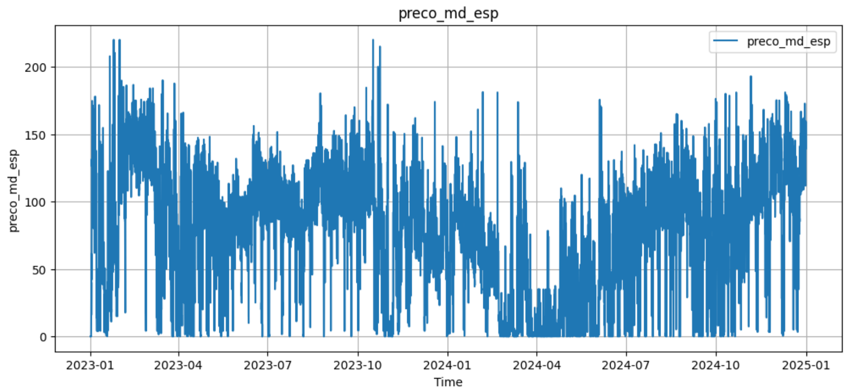

Figure A.8: preco md esp

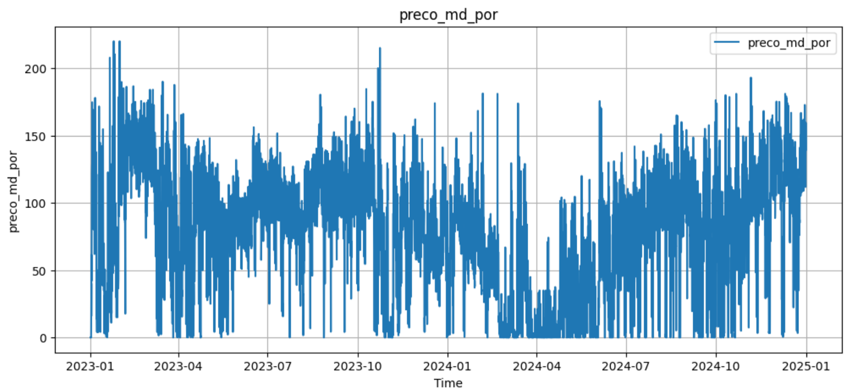

Figure A.9: preco md por

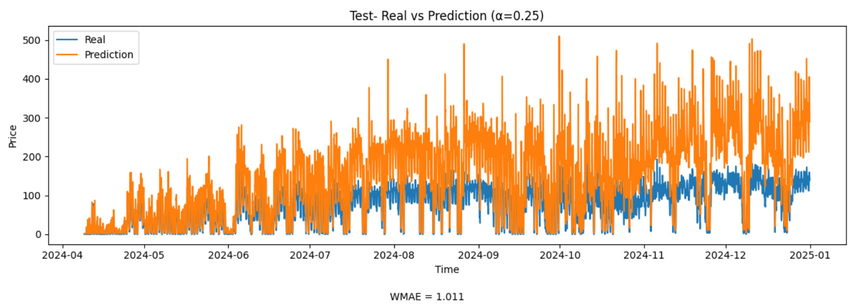

Figure A.10: previsao teste 0.25

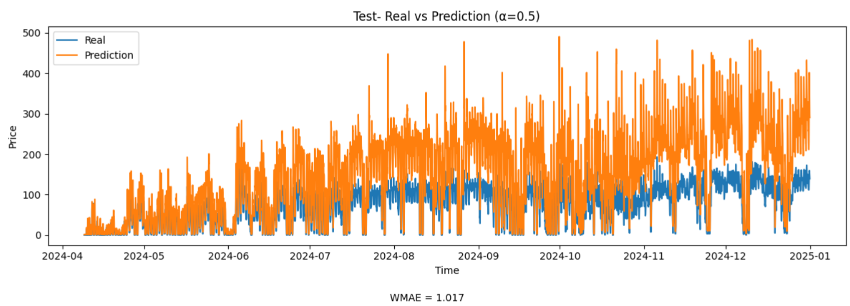

Figure A.11: previsao teste 0.5

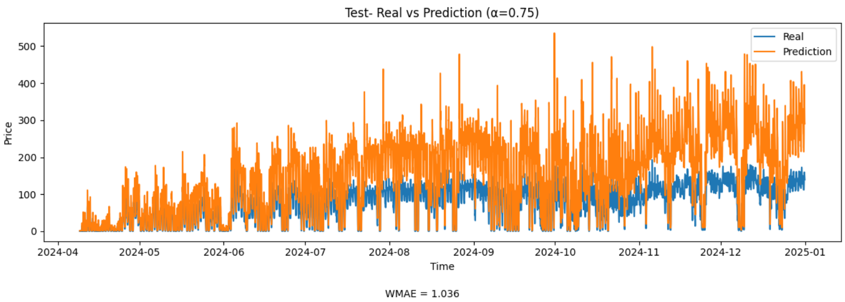

Figure A.12: previsao teste 0.75

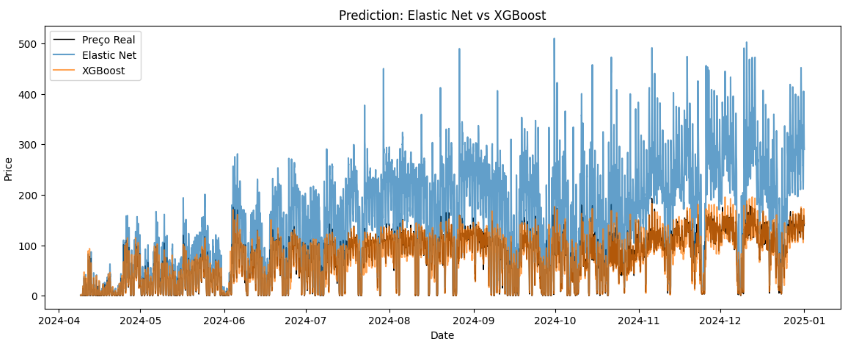

Figure A.13: ElasticNet vs XGBoost

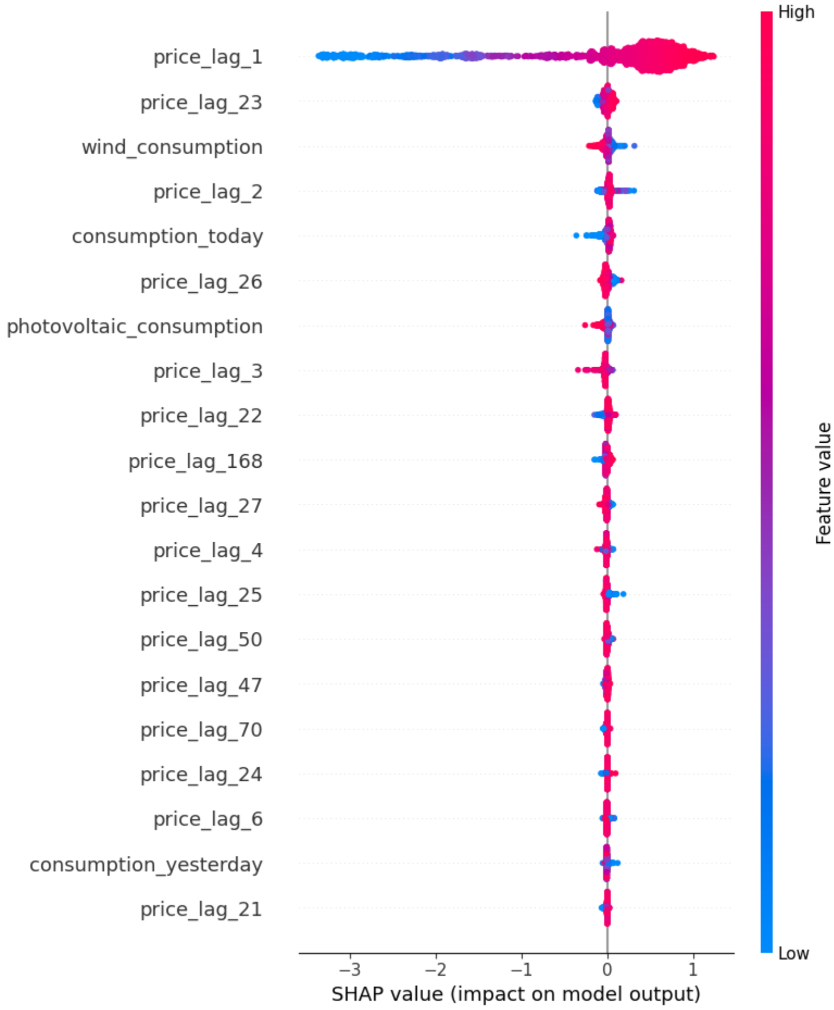

Figure A.14: SHAP value XGBoost