POLITECNICO MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

A Nonparametric Approach to Inference and Forecasting in the Italian Electricity Spot Market

Nonparametric Statistics Project
Mathematical Engineering - Ingegneria Matematica

Mattia Gastoldi, 10794648

Teachers:
Prof. Francesca Ieva
Prof. Simone Vantini
Dr. Alfredo G. Zapiola

Academic year:
2024-2025

Abstract[a] **:** In highly volatile markets such as the electricity sector, understanding structural dynamics is crucial for strategic planning and informed decision-making. This study focuses on the Italian electricity market, specifically on the Day-Ahead Market (MGP), where supply and demand offers are matched to determine hourly zonal prices. Unlike traditional time series approaches, we adopt a functional data analysis framework, treating entire supply and demand curves as single data objects. This allows for a more detailed understanding of market behavior. The first part of the study identifies recurrent patterns in the price evolution, revealing a characteristic double-peak structure within each day. These peaks are largely driven by demand surges, which shift the market equilibrium upward. Conversely, lower prices during weekends are explained by significant drops in demand, despite a generally increased production level. As a short-term prediction model, we implement a FAR(1), which leverages temporal correlations among the curves. Despite its simplicity and the absence of structural constraints, the model successfully captures the essential features of the data, proving preferable to more rigid semi-parametric alternatives in terms of predictive curve realism. Furthermore, conformal prediction intervals at 90% confidence are constructed using a modulated L^∞ non-conformity measure, accounting for local variability. While effective in many regions, these bands remain wide near areas of sharp functional jumps. Overall, this structural, curve-based approach offers both descriptive insight and predictive power.

[a]The source code is available at: **https://github.com/mattiagast/projectNPS**

Key-words: Nonparametric Inference, Functional Inference, Functional Data Analysis, Functional Autoregressive Model

1. Introduction

For a player operating in a highly volatile environment such as the energy market, understanding the underlying dynamics is essential to act efficiently and to plan appropriate strategies for production, purchasing, and selling. The Italian electricity market operates under a zonal pricing scheme and is organized into several trading platforms. The most relevant among them is the *Mercato del Giorno Prima* (MGP), or Day-Ahead Market, where the bulk of electricity is traded one day in advance. Every day, market participants submit hourly offers

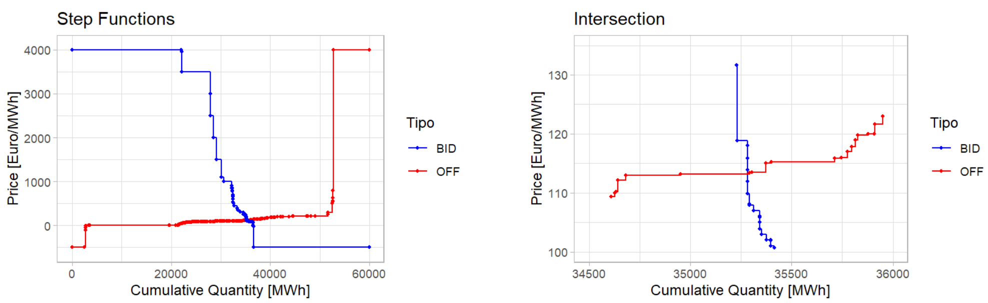
to buy or sell electricity for each hour of the next day. These offers include a quantity in megawatt-hours (MWh) and a corresponding price in euros per MWh.
The market mechanism is based on a uniform price auction within a specific geographic zones. The system operator aggregates all sell offers into an increasing supply curve and all purchase offers into a decreasing demand curve. The intersection of these curves determines the market-clearing price, known as the *Prezzo Zonale*, and the corresponding quantity to be exchanged. All transactions are settled at the clearing price, which ensures that all accepted sellers receive the same payment and all buyers pay the same price, regardless of their original bids. This mechanism ensures transparency, efficiency in dispatching generation units, and provides a reliable reference prices.
Instead of a common time series perspective, in this project, the whole supply and demand curves are considered. This study aims to analyze the structure and dynamics of the electricity market, with a particular focus on **identifying emerging trends** and **recurrent patterns** in the determination of the final price. This includes **understanding which factors may significantly influence** the final market price. Second, this project seeks to address the problem of **short-term forecasting**. In particular, to predict the supply and demand curves for the upcoming day based on historical market data.
After presenting the dataset, Section 2 focuses on the identification of relevant trends in the electricity market. Section 3 is devoted to the construction of functional representative curves for both supply and demand, using suitable nonparametric fitting techniques. Based on these representations, Section 4 performs a series of hypothesis tests aimed at investigating the structural properties of the market. Finally, Section 5 addresses the prediction task, with the objective of forecasting the future evolution of the supply and demand curves.

2. About the dataset

The data used in this analysis are provided by the Gestore dei Mercati Energetici (GME) [6], the Italian Energy Market Operator. All bids are collected in a daily `.xml` file, which contains detailed information about each offer, including the price, the quantity, the corresponding hour of the day, and the specific market zone. In addition, a categorical variable indicates whether the bid refers to the supply side (`OFF`) or the demand side (`BID`). The analysis focuses on the two-year period from 2023 to 2024, and it is limited to auctions that are accessible across all market zones.

2.1. Supply and Demand curve computation

Being the object of interest, the first task consists in processing properly the market curves. Specifically, for each daily file, all bids corresponding to a specific hour and market zone are extracted. These bids are then classified according to their type, supply (`OFF`) or demand (`BID`), and treated separately.
For supply offers, prices are sorted in increasing order, and the corresponding quantities are cumulatively summed. Conversely, for demand bids, prices are sorted in decreasing order, with quantities again cumulatively summed. Once the data are ordered in this way, the resulting stepwise functions can be fitted using appropriate tools (e.g., the `stepfun` function in `R`). An example of the resulting curves is shown in Figure 1.



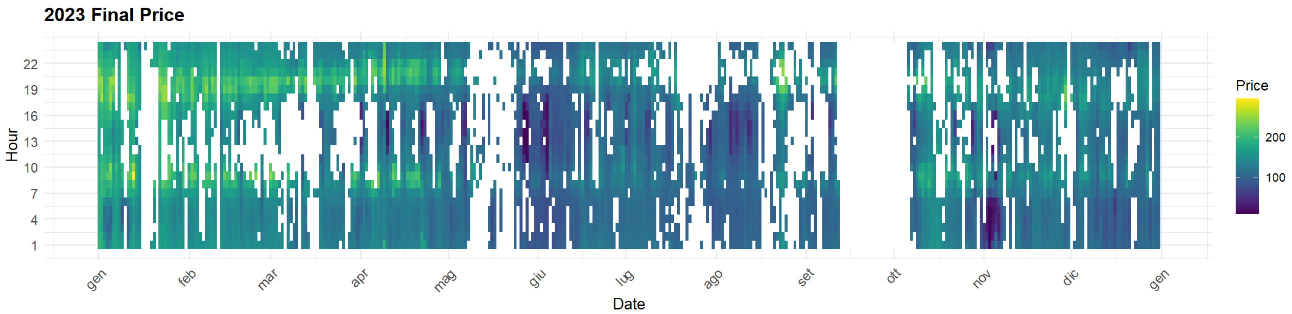
(a) Supply and Demand curves. (b) Close up to the intersection.

Figure 1: Resulting step-functions for Supply and Demand, along with the their intersection.

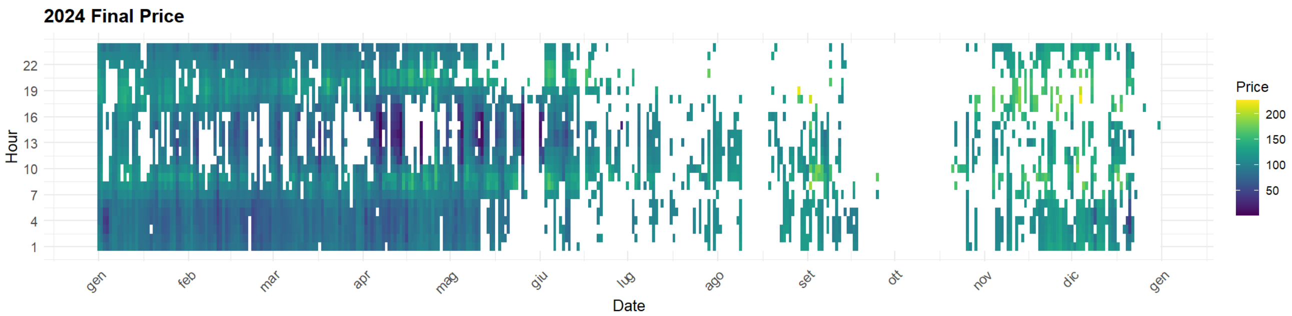
2.2. Trend identification

To reduce computational costs, this initial exploratory phase adopts a univariate approach, focusing directly on the variable of interest, the final market price, rather than directly modeling the entire supply and demand

curves.
Once the final prices have been extracted for both the years 2023 and 2024, for each day and each hour, and limited to the auctions accessible to all market zones, a heatmap representation provides a fast and intuitive overview of the data. In particular, it allows for a visual inspection of the temporal evolution of the price, and, most importantly, highlights the presence of missing values, that is, time slots in which the auction of interest did not take place. The corresponding heatmap is shown in Figure 2.



(a) Heatmap of the final price in 2023.



(b) Heatmap of the final price in 2024.

Figure 2: Heatmap of the final price in the period of interest.

From Figure 2, the following observations can be made:

- The price pattern appears to be regular and recurring throughout the two-year period: the data show that two peaks typically occur during the day, around 9:00 AM and 7:00 PM, while prices tend to drop in the middle of the day and overnight;
- The beginning of 2023 is characterized by generally higher prices, followed by a decreasing trend;
- In 2024, the absence of the relevant auction becomes more frequent, resulting in evident data sparsity.

To gain more insight from the final price, different levels of grouping are considered. Through the boxplots it is possible to observe the distribution of the data under a specific grouping factor. First, consider a grouping induced by the hour of the day, as in Figure 3.

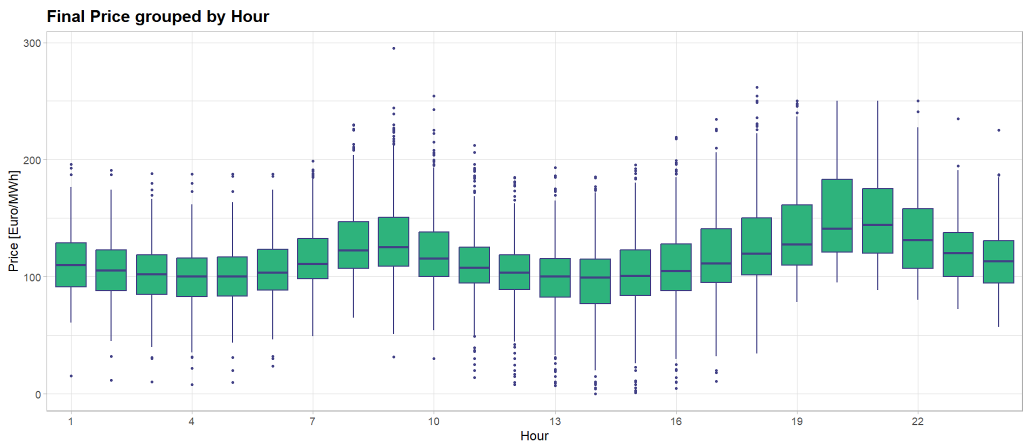
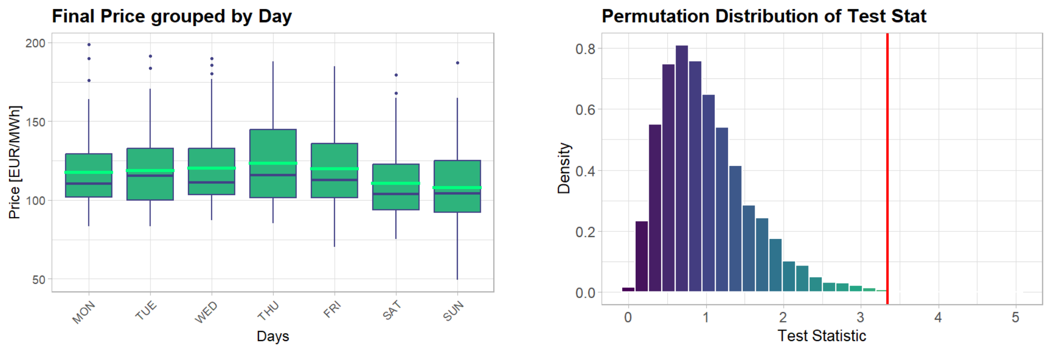


Figure 3: Boxplot of the final price considering as a grouping factor the hour of the day.

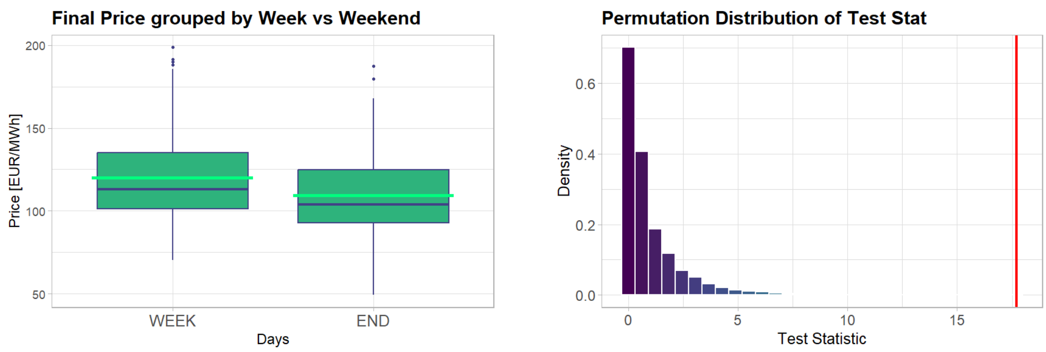
Figure 3 confirms the intuition regarding the daily behavior of the final price. The data show that two peaks occur during the day, around 9:00 AM and 7:00 PM, while prices tend to drop during midday and overnight. Typically, the medians are closer to the first quartile, indicating that the distribution of the final price is skewed toward lower values; however, large values are still present and frequently appear as outliers, i.e. with a really large value.
Moreover, consider the observations at 7:00 AM. This time slot is used as a baseline for the inference, as it provides a large number of observations, lower volatility, and fewer outliers. By fixing a specific hour, the collected observations can be reasonably treated as independent, since they are separated by 24-hour intervals. This is, however, a strong assumption, as in some periods temporal trends may induce dependence between observations. Nevertheless, this assumption is essential to perform many standard statistical inference procedures on the population.
Now consider, within the observations at 7:00 AM, a grouping factor based on the day of the week. To assess whether there are differences in the distribution of final prices across weekdays, a one-way ANOVA can be performed. However, since the assumption of normality is not fulfilled in any group, a permutation one-way ANOVA is used instead, as it does not rely on any parametric assumptions.
As shown in Figure 4, the null hypothesis of equal distributions is rejected, with an estimated p-value of 0.003, which is lower than any conventional significance level. The difference becomes even more evident when grouping prices based on whether they belong to a weekday or a weekend. In this case, the permutation p-value is essentially zero, leading to the conclusion that prices during weekends tend to be significantly lower. The result is displayed in Figure 5.



(a) Boxplot of the final price at 7:00 AM considering as a grouping factor the day of the week.

(b) ANOVA permutation distribution.

Figure 4: Permutation one-way ANOVA with respect to the day of the week.



(a) Boxplot of the final price at 7:00 AM, grouped by whether the day is a weekday or a weekend.

(b) ANOVA permutation distribution.

Figure 5: Permutation one-way ANOVA with respect to whether the day is a weekday or a weekend.

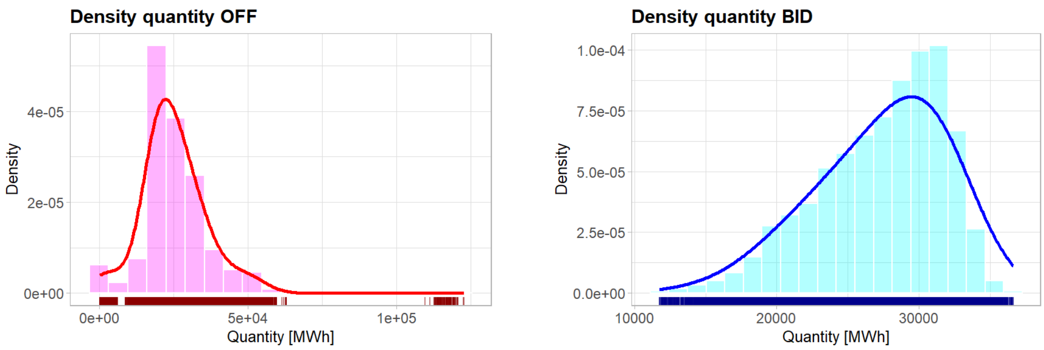
To summarize, this exploratory phase has highlighted several relevant properties regarding the behavior of the final price:

- Certain time slots during the day are typically associated with higher prices, while others tend to exhibit lower values;
- Prices during weekdays tend to be higher compared to those recorded over the weekend;

- In the early months of 2023, the final price shows a generally higher trend compared to the rest of the observed period.

2.3. Toward functional data analysis

Before applying functional data analysis to the Supply and Demand curves, it is necessary to first analyze the shape and structure of the curves themselves. Since these curves are step functions, they are characterized by relatively flat segments interrupted by a few sharp jumps, while other regions present many smaller steps. Through histograms, we can identify the areas of the domain that require more or less complexity when modeling the functions.
This step is essential not only to determine which portion of the domain should be considered for further analysis, but also to understand which zones are more stable (and can be modeled with lower complexity), and which ones, on the contrary, demand a more detailed representation due to their high variability.
The histograms are shown in Figure 6. Additionally, a Gaussian kernel smoothing has been applied using a bandwidth of 5000 for the Supply curves and 2500 for the Demand curves, purely for illustrative purposes. The smoothed density allows for a clearer visual interpretation of the underlying discrete histogram structure.



(a) Histogram of the bids for the Supply curves. (b) Histogram of the bids for the Demand curves.

Figure 6: Smoothed histograms of the bids for Supply and Demand curves.

From these plots, it can be concluded that the interval [0 MWh, 50000 MWh] is a sufficiently informative domain for smoothing the curves. This range entirely contains the bids associated with the Demand curves and excludes only a limited number of bids, definitely non-competitive, related to the Supply curves.
Regarding the modeling complexity, for the Demand curves, a more accurate representation of the central portion of the domain, while for the Supply curves, a more refined modeling is required over a broader range.

3. Smoothing

The first step in functional data analysis is the so-called smoothing procedure, which aims to retrieve proper functional representations of the statistical objects, in this case, the Supply and Demand curves. These functional representations are typically estimated via regression techniques based on observed data.
However, in the present context, the curves are already fully observed as step functions. Therefore, the goal shifts from estimating an unknown function to developing methods that can closely approximate, or almost overfit, the observed step functions. This approach allows us to retain the key structural features of the curves while obtaining smooth functional representations suitable for further analysis. The starting data are obtain by sampling from the real curves a large number of points.

3.1. Smoothing of the curves

In the literature a variety of methods for fitting step functions have been proposed. This section presents three techniques that are particularly suitable depending on the nature of the analysis to be performed.

Binning. Binning is a local regression method. In this context, the domain is divided into a finite number of windows, and within each window, the function is estimated by computing the mean. The resulting output is, by construction, a step function.

This approach is particularly useful, especially when the bins are defined in accordance with the insights from Section 2.3. Although this method is, by default, not very flexible or easily generalizable, it can provide highly accurate reconstructions when using a carefully chosen custom binning strategy. The specification for this technique are summarized in Table 1. With this choice, approximately 150 parameters need to be estimated for each curve.

Supply	**Values**
Sample from the curve	1000
Uneven breaks	seq(0,15000,by=500), seq(15200,35000,by=200), seq(35500, 50000, by=500)
Demand	**Values**
Sample from the curve	1000
Uneven breaks	seq(0,10000,by=2000), seq(10200,36000,by=200), seq(38000, 50000, by=2000)

Table 1: Summary of the hyperparameter for the binning.

Local averages. Local averaging is a local regression method. In this context, a moving window with a fixed bandwidth is constructed around each point of a grid, and the function is estimated by computing the mean within the window. The resulting output is, by construction, a step function.
This approach can be effective, but it requires a lower sampling density of the original data; otherwise, the steps tend to become overly smoothed. Moreover, since the method does not fit a function directly but rather estimates a set of discrete points, a sufficiently fine grid is needed to obtain an accurate approximation. The specification for this technique are summarized in Table 2. With this choice, the number of parameters that need to be estimated is equal to the cardinality of the evaluation grid for each curve.

Supply	**Values**
Sample from the curve	100
Bandwidth	500
Demand	**Values**
Sample from the curve	100
Bandwidth	500

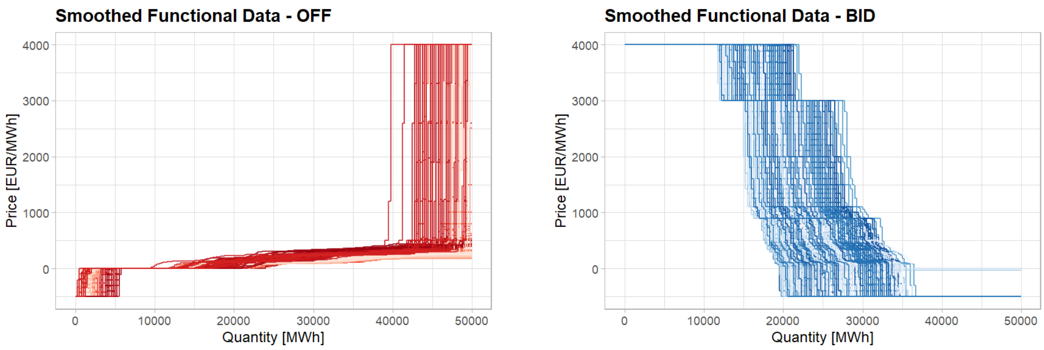
Table 2: Summary of the hyperparameter for the local averaging.

B-spline regression. B-spline is a global regression method. In this context, a set of basis functions (of order 0) with local effect, centered at specific knots, are combined. The problem is reformulated as the estimation of the weights that best fit the data. By construction, the resulting output is a step function.
By choosing an appropriate number of knots, it is possible to obtain a good global fit, achieving a trade-off between accuracy and computational efficiency. The specification for this technique are summarized in Table 3. With this choice, 200 parameters need to be estimated for each curve. This method is particularly useful for the functional hypothesis testing, where a proper regression is required.

Supply	Values
Sample from the curve	1000
Knots	200
Demand	**Values**
Sample from the curve	1000
Knots	200

Table 3: Summary of the hyperparameter for the B-spline regression.

An example of the smoothed functional data is shown in Figure 7



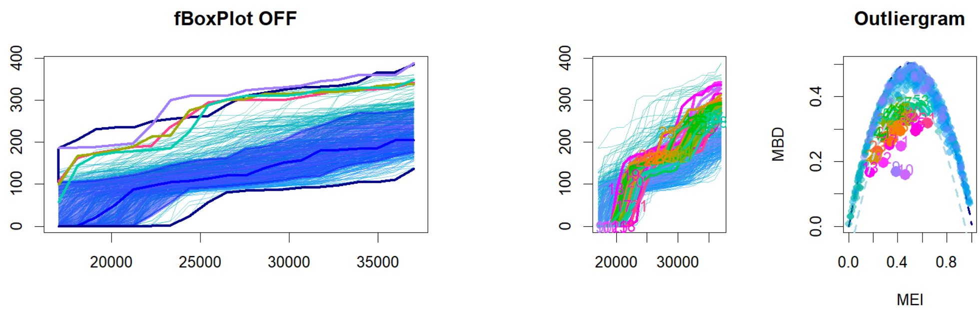
(a) Smoothed Supply curves.

(b) Smoothed Demand curves.

Figure 7: Smoothed functional data.

3.2. Outlier analysis

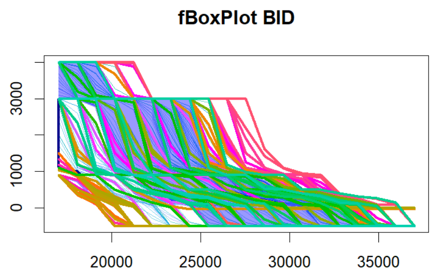
Outlier analysis is a crucial step before performing any type of inference on the data. Indeed, observations that behave very differently from the rest of the dataset can lead to misleading results or hide meaningful patterns. For this reason, it is essential to identify and handle each outlier and, when possible, understand its origin.
In the context of functional data, two main forms of outlyingness exist: amplitude outliers (analogous to magnitude outliers in multivariate statistics) and shape outliers (which are intrinsic to the functional nature of the data and do not have a counterpart in multivariate settings). These types are detected using two different tools: the functional boxplot for amplitude outliers and the outliergram for shape outliers. Given the nature of the functional data considered in this problem, which vary both vertically and horizontally, many observations may be flagged as outliers. For this reason, it is preferable to adopt a more conservative approach, removing observations only when they exhibit behavior that is clearly different from the rest of the dataset. The results of this analysis are shown in Figure 8 and Figure 9 for Supply and Demand curves, respectively.



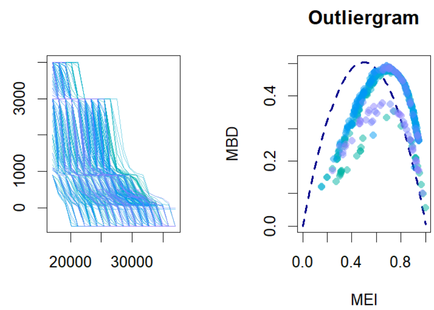
(a) Smoothed Supply curves.

(b) Smoothed Demand curves.

Figure 8: Outlier analysis for Supply curves.



(a) Functional boxplot.



(b) Outliergram.

Figure 9: Outlier analysis for Demand curves.

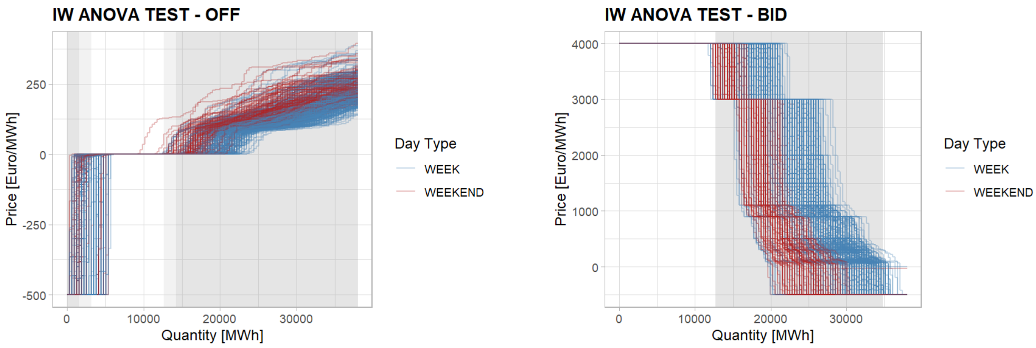
Overall, the Supply curves, which predominantly develop horizontally, at least in the central part of the domain, exhibit few amplitude outliers. In contrast, the Demand curves, which develop more vertically and are spread horizontally, display a large number of amplitude outliers. Conversely, the Supply curves appear to be more variable in terms of shape over the considered domain, while the Demand curves are noticeably more regular.
It is also worth noting that all shape outliers identified for the Supply curves belong to the first three months of 2023. Compared to the rest of the dataset, these curves exhibit a steeper slope in the central region, leading to intersections with the Demand curves at generally higher price levels. This behavior may explain why the final price tends to be higher during that period.

4. Inference

Hypothesis testing in the context of functional data can be naturally formulated as an extension of traditional multivariate tests, where the test statistic is defined as a distance in a functional norm (e.g., the L^2 norm). However, a major challenge in this setting is that parametric assumptions are rarely satisfied. For this reason, functional hypothesis tests are typically constructed in a permutation-based framework, which allows for the assessment of global significance without relying on parametric distributions.
However, the main interest often lies in identifying the specific regions of the domain that are responsible for the rejection of the null hypothesis. This leads to the definition of local tests, which can be interpreted as multiple hypothesis tests across the domain. To maintain control over the global significance level, a suitable correction strategy must be applied. In this context, an interval-wise correction procedure can be used, leading to the framework of interval-wise hypothesis testing (IWT).
In Section 2.2, several trends were identified. By applying interval-wise hypothesis testing to the functional curves, we can determine which structural parts are significantly responsible for the observed differences.

4.1. Weekdays test

Consider a grouping based on whether the price belongs to a weekday or a weekend. Assuming the functions belonging to different days to be independent, an interval-wise ANOVA test is performed on the curves. The result is reported in Figure 10. The dark-gray regions are the portion of the domain where the null hypothesis is rejected.



(a) IW ANOVA Test on Supply curves.

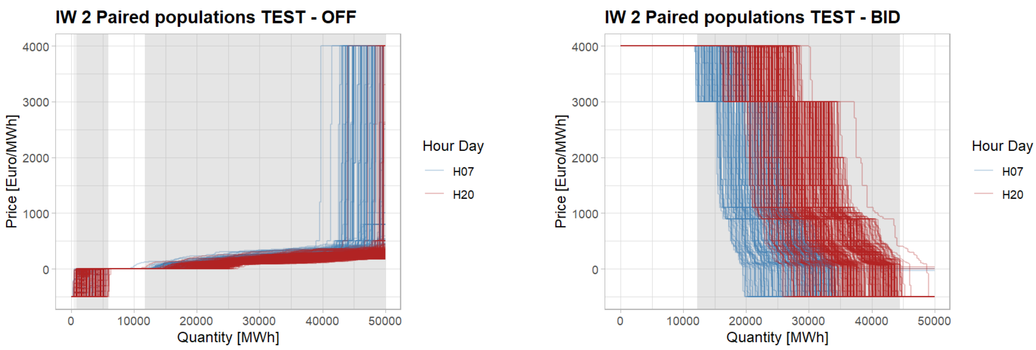
(b) IW ANOVA Test on Demand curves.

Figure 10: IW ANOVA Test with respect to whether the day is a weekday or a weekend.

From this test, it is clear that the curves associated with the two groups differ. In particular, for the Supply curves of the weekend, we observe a clear upward shift that originates from the very beginning of the domain. However, the most relevant factor affecting the price is that the Demand curve of the weekend is significantly shifted to the left, resulting in an intersection at a lower ordinate.

4.2. Hours test

Now consider the grouping based on the hour of the day. In this context, we are interested in comparing the location of two paired samples of functional data, corresponding to 7:00 AM and 8:00 PM. The result is reported in Figure 11. The dark gray regions represent the portions of the domain where the null hypothesis is rejected.



(a) IW two-paired sample location Test on Supply curves.

(b) IW two-paired sample location Test on Demand curves.

Figure 11: IW two-paired sample location Test considering as a grouping factor the hour of the day.

From this test, it is clear that the curves associated with the two groups differ. In particular, for the Supply curves of 8:00 PM, we observe a clear delay which also characterizes the very beginning of the domain. However, the most relevant factor affecting the price is that the Demand curve is significantly shifted to the right, resulting in an intersection at a higher ordinate.

5. Prediction

In the context of time series, the most well-known predictive models are certainly autoregressive (AR) models. Autoregressive models are typically built under stationarity assumptions and predict a new observation as a function of the previous p observations.
In the context of functional data, these models are referred to as functional autoregressive models (FAR). In general, a FAR(p) model can be expressed as:

$$X_n = \alpha + \sum_{i=1}^{p} \Psi_i(X_{n-i}) + \epsilon_n$$

where Ψ_i is a linear operator. In the Hilbertian setting, Ψ_i is a Hilbert–Schmidt operator, which can be represented as:

$$\Psi(X)(t) = \int \psi(t,s)X(s)\,ds$$

where $\psi(t,s)$ is the kernel function. Moreover, $X(t)$ is assumed to be a L^2 function, such that $\mathbb{E}[X] = 0$, and the errors ϵ_i are be independent and centered at zero.
The theory of autoregressive and more general linear processes in Banach spaces is developed by Bosq [2], while several prediction methods are presented by Besse [1]. In recent years, a variety of advanced methods have been proposed. A thorough review and comparison is provided in Chen [4].

5.1. FAR(1)

Prediction equations may be derived explicitly for general stationary functional time series, but they seem difficult to solve and implement. As a consequence, much of the research in the area has focused on the first-order functional autoregressive model, shortly FAR(1). This is the simplest functional autoregressive model, and is used to describe the one-step-ahead prediction of a functional observation given the current one, and it is structured as follows:

$$X_n = \alpha + \Psi(X_{n-1}) + \epsilon_n.$$

To estimate the coefficients of this model, several techniques have been proposed. One such approach is the standard first-order predictor proposed by Bosq [2]. This method represents the functional counterpart of the Yule-Walker equations, which aims to estimate the kernel of the autoregressive operator using the 1-step lag autocovariance structure. Using Functional Principal Component Analysis (FPCA), the estimation can be performed in a lower-dimensional subspace. In this framework, the kernel is then represented through a basis expansion, typically using the empirical eigenfunctions of the covariance operator, always obtained with the FPCA, as result of the spectral theory. The estimation technique can be summarized as follow:

$$\hat{\psi}_{ji} = \hat{\lambda}_i^{-1}\frac{1}{N-1}\sum_{n=1}^{N-1}\langle X_n, \hat{v}_i\rangle\langle X_{n+1}, \hat{v}_j\rangle$$

which is an empirical analog of the relation $\psi_{ji} = \lambda_i^{-1}\mathbb{E}[\langle X_{n-1}, v_i\rangle\langle X_n, v_j\rangle]$ and $\hat{\lambda}_i$ the eigenvalue corresponding to $\hat{v}_i$. Taking d the low dimension approximation, whose choice depends on the decay rate of the eigenvalues, we calculate the predictions as

$$\hat{X}_{n+1}(t) = \int \hat{\psi}_d(t,s)X_n(s)ds = \sum_{k=1}^{d}\left(\sum_{\ell=1}^{d}\hat{\psi}_{k\ell}\langle X_n, \hat{v}_\ell\rangle\right)\hat{v}_k(t).$$

Finally, in the prediction, the mean is added. Didericksen [5] have evaluated several competing prediction models in a comparative simulation study, finding that this method have the best overall performance.

5.2. Prediction and conformal prediction intervals

To assess the performance of the aforementioned model, we consider the period January–February 2024. As shown in Figure 2b, the assumption of stationarity in the price dynamics appears plausible during this time window. Moreover, there are no missing values in this period, which significantly simplifies the preprocessing phase. To incorporate an appropriate quantification of the uncertainty associated with the predicted curve, we also construct conformal prediction intervals.
Consider the 45 observed curves from January 11th, 2024 to February 24th, 2024, and the corresponding one-lagged 45 curves from January 12th, 2024 to February 25th, 2024. These are paired and then split into two sets: the first 15 paired curves are used as the *calibration set*, while the remaining 30 are used as the *proper training set*, preserving the temporal ordering. The goal is to predict the curve of February 27th using the observation from the previous day.
With this setup, the model described in Section 5.1 is trained on 30 consecutive pair of curves that are temporally close to the target prediction. After training, the 15 paired curves from the calibration set are used to compute non-conformity scores by comparing the true curve with the one predicted by the FAR(1) model. Specifically, the following non-conformity score is considered:

$$A(X_{i-1}, X_i) = \sup_{t\in I}\frac{|X_i(t) - \widehat{X}_i(t)|}{\sigma(t)},$$

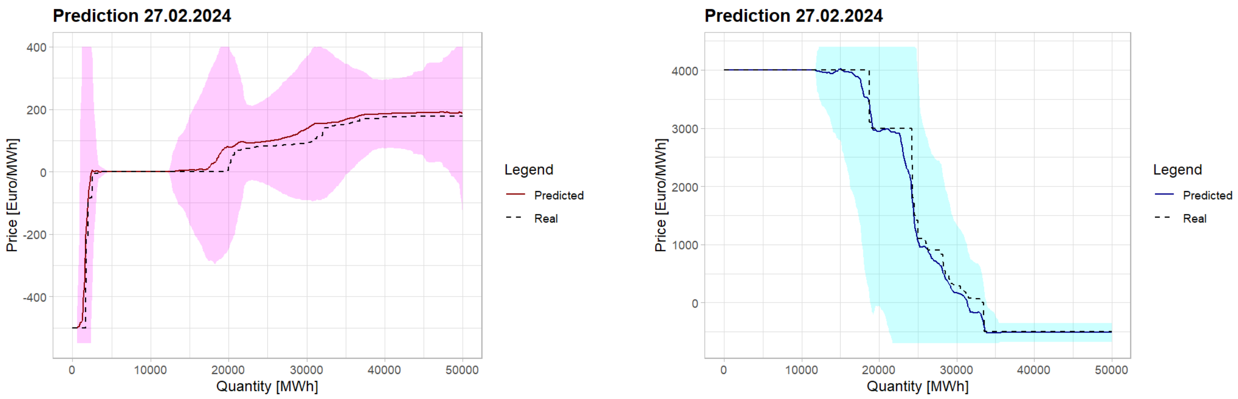
where $\widehat{X}_i(t) = \text{FAR}(X_{i-1})(t)$ is the predicted curve obtained from the FAR(1) model.
This choice allows the scores to be modulated by the local variability across the domain, thus accounting for heteroskedasticity in the functional observations. Being $\hat{X}_{n+1}$ the predicted curve, the $(1-\alpha)$ conformal prediction bounds are then computed as:

$$CP_{1-\alpha}(t) = \left[\hat{X}_{n+1}(t) - \epsilon_{(1-\alpha)} \cdot \sigma(t),\ \hat{X}_{n+1}(t) + \epsilon_{(1-\alpha)} \cdot \sigma(t)\right],$$

where

$$\epsilon_{(1-\alpha)} = \text{quantile}_{\lceil (1-\alpha)(n-m) \rceil} \left(\{A(X_i, X_{i+1})\}_{i=m}^{n} \right)$$

The results of the prediction are shown in Figure 12 with $(1-\alpha) = 90\%$.



(a) Predicted Supply curve. (b) Predicted Demand curve.

Figure 12: Predicted curves and their conformal prediction intervals.

The results demonstrate that the prediction method, despite not being constrained by any structural assumptions, is able to capture the overall behavior of the real curves. In particular, it reproduces reasonably well the levels at which the jumps occur. The predicted curve appears to be slightly left-shifted with respect to the observed one and the Supply curve seems to be higher in the zone of the intersection. Indeed, as shown in Figure 13, the predicted value for the final price is 146.7 Euro/MWh, compared to the observed value of 111.1 Euro/MWh.
Although the model is relatively simple, its overall performance is satisfactory.
Regarding the conformal prediction bounds, the modulation by local variability is crucial to mitigate the extremely large residuals caused by the steep jumps in the functions. In particular, using the L^{∞} norm without such modulation would result in overly wide prediction bands across the entire domain. However in correspondence of the wide jumps, being the curves extremely variable the prediction bounds are still excessively high, leading to a not very efficient coverage of the prediction set.

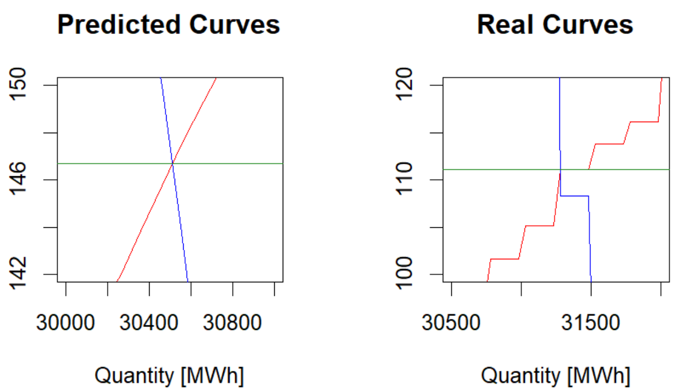


Figure 13: Comparison between predicted and real curves at the intersection neighborhood.

6. Conclusions

This study reveals several interesting trends and patterns in the energy market.
First, a certain regularity emerges in the daily behavior of the final price. Typically, within a day, the price reaches two peaks, approximately around 9:00 AM and 8:00 PM. These increases are primarily driven by a significant rise in demand, which shifts the demand curve to the right, thereby increasing the market-clearing price. Although a general increase in production is also observed in those time-spots, which flattens the supply curve and should, in principle, lower prices, this effect is not sufficient to counterbalance the impact of the increased demand. As a result, the final price remains elevated.
Similar reasoning explains why prices tend to be lower during weekends. In those cases, although the production curve generally increases, the demand curve shifts significantly to the left due to reduced consumption, causing the market-clearing point to fall at a lower ordinate. This results in a lower average final price.
For the prediction phase, a FAR(1) model was implemented. This model is among the most widely used in practice for exploiting the temporal correlation between functional observations. The estimation of the operator coefficients, despite being data-driven and relatively simple, successfully captures the general behavior of the curves. Therefore, even in the absence of structural constraints, the model is capable of generating realistic predictions. This contrasts with more rigid semi-parametric estimation techniques (e.g., Ridge regression), which typically require larger datasets and often produce overly regularized predictions.
However, the point predictions of the model tend to be concentrated around the empirical mean, due to the stationarity assumptions inherent in the model—assumptions that are not always realistic in real-world data.
To incorporate an appropriate quantification of the uncertainty associated with the predicted curves, conformal prediction intervals were constructed at a confidence level of $\alpha = 90\%$. The non-conformity measure adopted was L^∞ norm of the residual, scaled by the local variability of the domain. While this modulation is crucial to mitigate the global impact of the jumps and abrupt changes, in some regions the prediction bands remain excessively wide. This is partly due to the limited efficiency of the L^∞ norm as a non-conformity measure, accentuated in the presence of sharp discontinuities.
Overall, the results obtained are satisfactory and provide valuable insight into the functional dynamics of the energy market. Nonetheless, we highlight that the analysis is entirely structural and based exclusively on curve shapes, without incorporating any external covariates that are likely to play a significant role in price determination. For instance, important variables such as natural gas prices, renewable energy production, or macroeconomic indicators like inflation have not been included.
Future improvements may include the analysis of correlations with such external covariates. From a predictive standpoint, more advanced methods such as FARX [4] or transformations into well-suited functional spaces approaches that have shown great effectiveness in similar contexts [3] could be explored.

References

[1] Philippe C Besse, Hervé Cardot, and David B Stephenson. Autoregressive forecasting of some functional climatic variations. *Scandinavian Journal of Statistics*, 27(4):673–687, 2000.

[2] Denis Bosq. *Linear processes in function spaces: theory and applications*, volume 149. Springer Science & Business Media, 2000.

[3] Antonio Canale and Simone Vantini. Constrained functional time series: Applications to the italian gas market. *International Journal of Forecasting*, 32(4):1340–1351, 2016.

[4] Ying Chen, Thorsten Koch, Kian Guan Lim, Xiaofei Xu, and Nazgul Zakiyeva. A review study of functional autoregressive models with application to energy forecasting. *Wiley Interdisciplinary Reviews: Computational Statistics*, 13(3):e1525, 2021.

[5] Devin Didericksen, Piotr Kokoszka, and Xi Zhang. Empirical properties of forecasts with the functional autoregressive model. *Computational statistics*, 27(2):285–298, 2012.

[6] Gestore Mercati Energetici. Mercato del giorno prima, http://www.mercatoelettrico.org, 2025.