Politecnico di Torino

ScuDo
Scuola di Dottorato ~ Doctoral School
WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation
Doctoral Program in EECE (37^{th}cycle)

Key Concepts and Methods in Point and Probabilistic Forecasting of Electricity Prices

By

Francesco Moraglio

******

Supervisor(s):
Prof. Carlo S. Ragusa, Supervisor

Doctoral Examination Committee:
Prof. Andrea Cristofolini, Referee, Università di Bologna, Bologna.
Dr. Domenico Giordano, INRiM, Torino.
Prof. Antonino Laudani, Referee, Università di Catania, Catania.
Prof. Andrea Mazza, Politecnico di Torino, Torino.
Prof. Igor S. Stievano, Politecnico di Torino, Torino.

Politecnico di Torino
2025

Declaration

I hereby declare that, the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Francesco Moraglio
2025

* This dissertation is presented in partial fulfillment of the requirements for **Ph.D. degree** in the Graduate School of Politecnico di Torino (ScuDo).

This work represents a form of free expression of thought and the personal opinions of the author. It does not reflect institutional positions or those attributable to other industrial or scientific organizations.
It is recommended for an adult and mature audience capable of understanding the proposed topics without distorting them, interpreting them partially or biasedly, thereby compromising their proper expression and original meaning.

Acknowledgements

Italian

La presente Tesi è stata redatta al termine del percorso di dottorato, finanziato nell'ambito del PON REACT-EU 2014-2020 – Azione IV.5 "Dottorati su tematiche green", in attuazione del Decreto Ministeriale del 10 agosto 2021, n. 1061.

English

This thesis was written at the conclusion of the PhD program, funded under the PON REACT-EU 2014–2020 – Action IV.5 "PhDs on green topics," pursuant to Ministerial Decree No. 1061 of August 10, 2021.

Abstract

Italy, as a developed country and a member of the European Union, is undertaking the process commonly referred to as the energy transition. This involves the rapid integration of high percentages of renewable energy generation into the electric system. Unlike traditional thermoelectric generation from fossil fuels, renewables lack flexibility and thus require greater care and optimization to be made both energy-efficient and financially viable. Unfortunately, this evolution is taking place during a particularly challenging period: Europe has to deal with an unstable geopolitical framework. After the Russian aggression to Ukraine, a tariff war on gas prices together with and intense wave of speculation have resulted in historically unprecedented commodity prices. Certain segments of industry have faced serious difficulties, particularly electric utility companies, which had to incur over-indebtedness to ensure supplies. In order to control financial risks, forecasting the movement of future electricity prices becomes critical for electric operators. While long-term market outcomes remain difficult to predict, short-term electricity price trends can be analyzed and understood. This understanding means—if not the generation of precise numerical predictions (point forecasting)— at least the modelling of stochastic price behavior (probabilistic forecasting). This thesis consists of three self-contained chapters that explore the problem of Electricity Price Forecasting (EPF), progressing from introductory aspects to the latest trends.

In Chapter 1, we present a comprehensive introduction. We start with an outline of the energy sector on European and Italian scales. We place particular emphasis on the recent crises affecting the EU, which have become a serious concern and have dramatically conditioned electricity markets in recent years. After that, we present some intuitions about the fundamentals of EPF and an anticipation of the main results.

In Chapter 2, we present a case study on short-term EPF in the Italian electricity. This part has two main objectives: exposing which factors a forecaster should take into

account when developing a statistical model, and clarifying the difference between point and probabilistic forecasting. We present a series of univariate regression analyses that illustrate the differences between models belonging to different classes (linear/nonlinear, parametric/ nonparametric) and across different inputs. After that, we conduct both point and probabilistic multivariate regression analyses, with detailed evaluation of the results. The key to success lies in choosing the right predictors, which proves more critical than the choice of statistical forecasting algorithm.

The final research Chapter 3 addresses one of the most current areas of research in EPF: Model Averaging (MA). This involves combining forecasts from multiple, diverse models to enhance prediction accuracy and robustness. The concept is introduced with point forecasts and extended to the contemporary topic of MA for probabilistic models. Various aggregation schemes are evaluated for different classes of models, and the advantages of this approach are demonstrated: a reduction of roughly 10% in the error metrics can be achieved. Specifically, while point forecasting can benefit from a nearly indiscriminate aggregation of models, probabilistic forecasting requires a more cautious and deliberate analysis.

The concluding Chapter offers broader reflections and summarizes the lessons learned from the EPF problem. Finally, we outline the potential directions for future development of this work.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The present Chapter will summarize the outcomes of this PhD project, starting with the initial premises and motivations and concluding with an anticipation of the main results. It is important to stress that we have made several efforts in order to make each Chapter self-contained. Hence, each Reader can potentially avail herself of this research product with a partial consultation. This first part is structured in two main Section. The former clarifies the difficulties that modern European energy markets present to researchers and industry professionals. The latter presents a summary of the methods used and results achieved in the problem of Electricity Price Forecasting (EPF), keeping the treatment intuitive even for the uninitiated.

The approach is similar in the following Chapter 2, where particular emphasis is placed on explaining the differences between linear and nonlinear statistical forecasting models. The process is firmly based on describing how algorithms and the underlying data relate: this facilitates the Reader in becoming familiar with both the methodologies and the problem-solving perspective that must be adopted in EPF applications. In particular, this includes the development of a critical view with respect to the results obtained, thus avoiding a mere finding of numerical outcomes.

Chapter 3 has a more technical style, and it is devoted to analyzing the results that can be achieved by employed the so-called ensemble approaches, which are the state-of-the art in contemporary forecasting applications.

1.1 Motivation

Predicting future electricity prices in Italy is the central problem in this Thesis. Let us begin by understanding why this is a topic of interest. Indeed, it is, and it has become particularly hot in recent years, especially 2022-2024 (those covering my PhD period), due to certain crises which drove energy prices, particularly electricity prices, to concerning levels for both industries and governments. Such crises do not only concern Italy, but are rooted in wider, worldwide and European processes. Specifically, it is worth reflecting on the European context, as most of the Italian energy policies are based upon transposing directives from the European Union (EU), which has achieved many milestones in creating an energy infrastructure consistent across its Member States.
Unluckily and for nearly two decades, Europe has been struggling with a series of financial, economic, political, and social crises. Among these, the 2008 Global Financial Crisis (GFC) stands out as the most severe economic downturn since 1929. Although the EU experienced notable difficulties during this time, the core of the crisis was in the United States, with its impact on the Old Continent being relatively indirect, particularly in Mediterranean countries such as Italy, where financial markets played a smaller role in the economy. However, the GFC set off a domino effect that culminated in Europe's catastrophic Sovereign Debt Crisis (SDC) in the early 2010s. Those were the years of austerity policies for Europe[1], where measures aimed at reducing government deficits and debt accumulation were implemented, often involving spending cuts and tax increases. When, a few years later, timid signs of recovery began to appear, the public opinion converged on two fundamental points:

1. the need for new investments to revitalize economies and societies;
2. the necessity of doing so *in a sustainable manner*.

Indeed, sustainability has become the central concept of modern human and technological development, and it is worth stepping back to reflect on its meaning. The early 2000s, immediately preceding the GFC, witnessed an intensification of certain socio-economic

[1]This period should not be confused with the first era of austerity in the West, which occurred in the years following 1973. The austerity of the 1970s was largely a response to the oil crises, which caused a sharp rise in energy prices and inflation, leading many Western nations to implement strict economic measures to stabilize their economies.

dynamics from the previous two decades. On one hand, the 1980s and 1990s were periods of great prosperity and development for the Western world. During the former decade, neoliberal policies were implemented, characterized by a general disregard for social issues in favor of economic and technological development. Governments worldwide were not aware that society, the economy, and the environment had become increasingly interconnected, and that neglecting this relationships would have proven harmful in the near future. In other words, the approach was not sustainable in the long term and led to significant imbalances: income inequality widened, social safety nets weakened, and, above all, environmental degradation accelerated. This logic, based upon rampant consumerism, was fundamentally flawed. The pursuit of endless production is inherently unsustainable because it fails to account for the finite nature of natural resources and the environmental and human costs of unchecked industrial activity.

This issue was already known to experts, but a shared vision between industry and politics was missing. Indeed, as early as in the late 1980s, new regulations were introduced in developed countries to address labor standards, environmental concerns, and market practices. In response, many companies began offshoring production to countries with lower labor costs and less stringent regulatory frameworks, a transformation that contributed to job losses and widened social disparities in their home economies.

This shift away from domestic production marked the beginning of a broader economic transformation. Over time, developed countries moved from being manufacturing powerhouses to service-oriented societies. While this transition was often framed as a natural evolution toward a more advanced economy, it came with significant trade-offs. Unlike the industrial economy, which provided stable jobs and produced tangible goods, the service-based economy proved to be less resilient and more susceptible to volatility. Within this new landscape, the financial sector emerged as a dominant force, driving what some have called a "virtual economy." Fueled by unchecked debt and speculative practices, this system created an illusion of prosperity. Growth figures and stock market performance painted a picture of economic health, but beneath the surface, the foundations were increasingly fragile [1]. The focus on short-term profits and financial engineering overshadowed long-term investments in productive capacity and social well-being. The consequences of this imbalance became glaringly apparent during the years following 2008. In the aftermath of the GFC, a global consensus began to emerge: sustainable prosperity requires a more holistic approach, one that balances economic, social, and environmental priorities.

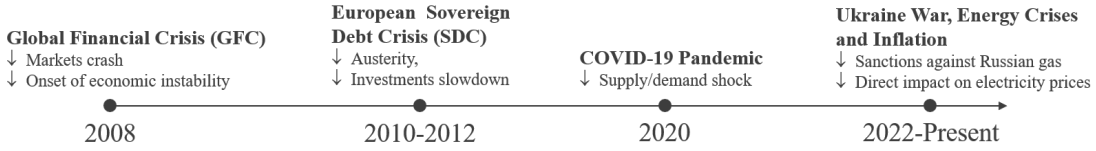

Fig. 1.1 **Timeline of major economic crises and their implications for the European energy landscape.** *The present European situation arises from the four pivotal phenomena represented in this chart. Key events include the Global Financial Crisis (2008), the European Sovereign Debt Crisis (2010–2012), the COVID-19 pandemic (2020), and the Energy Crisis (2022–2024). Each event reflects a critical turning point influencing both macroeconomic stability and electricity price dynamics across Europe.*

Still, despite the growing global awareness of sustainability and the urgent implementations of a more balanced approach to economic development, why is the situation continuing to worsen? Moreover, why is Europe losing ground to the two current superpowers, the United States and China?

Several factors contribute to this troubling trend. First, additional crises impacted our society in recent years. Second, while the EU has been a leader in setting ambitious sustainability goals, its implementation has often been slow and fragmented. Third, the USA and China have leveraged their economic and technological ability to dominate key industries of the future. The USA benefit from a dynamic private sector, massive investment in research and development, and a culture of entrepreneurship. China, on the other hand, combines state-driven strategic planning with aggressive investment. While the EU emphasizes regulation and sustainability, it has struggled to match the scale and speed of its competitors. Additionally, Europe's reliance on outdated industrial models and its slower adoption of digital transformation have further widened the gap. Finally, geopolitical dynamics play a significant role: the USA and China have been able to leverage their global influence to secure resources, markets, and strategic partnerships, often at the expense of Europe.

The European Commission, deeply concerned about the future competitiveness of the EU in the global arena, appointed a special figure in 2023 to address the issue: Mario Draghi. The former president of the European Central Bank subsequently drafted a detailed and ambitious plan aimed at revitalizing the economic system and restoring the EU to a leading global role. Published in September 2024 and divided into two sections, the report outlines both the existing critical issues and programmatic guidelines for potential solutions. Below, we review some highlights on topics relevant to this Thesis.

1.1.1 Understanding Critical Issues Within the EU

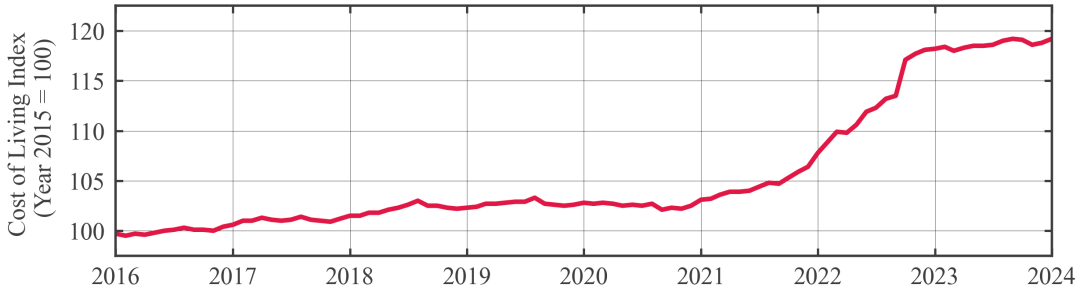

Fig. 1.2 **Cost of Living Index in Italy (2016-2024).**
The figure illustrates the sharp increase in the cost of living that Italians have experienced starting from the latter months of 2021. Specifically, the index is computed by Istat, the National Italian Institute for Statistics, and represents the cost of living for families of employees and workers (Famiglie di Operai e Impiegati, FOI).The computation is based upon a balanced basket of goods and services, including food, rent, transportation, instruction, and financial services. Alcoholic beverages are included, but tobacco products are excluded. See [2].

Throughout contemporary history, the most severe damages to civilization often result not from a single crisis, but from the convergence of multiple crises within a short time frame. These overlappings tend to create vicious socio-economic cycles that are particularly hard to manage in the presence of weak governance. According to the Draghi report [3], in recent years, two major phenomena have significantly worsened the European scenario:

1. The COVID-19 pandemic;
2. The Russian-Ukranian conflict.

First and foremost was the coronavirus crisis, which disrupted economies and exposed vulnerabilities in both national and international markets. Importantly, the economic decline was further intensified by mismanagement in the post-crisis period. Indeed, under the leadership of European institutions, Member State governments implemented recovery measures aimed at rebooting critical sectors such as energy or infrastructures, and modernizing production processes, with a clear emphasis on sustainable investments. This same PhD program was partially funded under the REACT-EU initiative, which includes €47.5 billion funds [4]. Despite the ambitious scope of these recovery plans,

many fell short of their objectives. Some measures proved ineffective, while others unintentionally caused additional harm. Financial injections, though substantial, often failed to deliver the desired outcomes due to structural inefficiencies, poor implementation, and a lack of coordination among Member States. The rushed nature of certain initiatives led to wasteful spending and misallocated resources. In Italy, a striking example is the Bonus Edilizi (i.e. "Building Incentives") programs, particularly the widely known Superbonus 110%. This measure was criticized as inflationary and market-distorting by the Corte dei Conti, the Italian Court of Auditors, as well as by the aforementioned Mario Draghi, who, during his tenure as Prime Minister, imposed limits on its application.
As a result, the anticipated recovery did not materialize, leaving the EU with persistent economic instability and uncertainty. This has translated into rising inflation and an increasing cost of living, driven by a demand inflated through excessive financial availability. The expression "too much, too fast" has become a fitting summary of the financial investment strategies of this period. In our case, the present Italian situation can be understood by looking at the Famiglie di Operai e Impiegati (FOI) cost of living index, as computed by Istat, the Italian national institute of statistics. Figure 1.2 shows that, during 2022, the cost of living in Italy increased by around 20%.
The second major phenomenon impacting European economies is the ongoing[2] war in Ukraine, which undermines the European competitiveness by causing excessively high energy prices [3]. The EU, underestimating the extent of its Member States' dependence on Russian fossil fuels and failing to anticipate a significant wave of speculation, imposed heavy sanctions on Russia's exports immediately after the invasion of Donbass. This decision, while politically and morally justified, had far-reaching consequences. Indeed, the sudden disruption of energy supplies triggered a sharp increase in prices, exacerbated by speculative trading in global markets.This resulted to previously unseen increases in prices of Natural Gas (NG) and, consequently, electricity. To address this shortfall, recent years have seen a reliance on importing Liquefied NG (LNG), primarily from America but also from other sources. Since the liquefaction process is energy-intensive, and shipping requires subsequent regasification, LNG transportation significantly increases costs. Consequently, in the EU, gas prices have averaged 3x higher than those in the USA [3]. In Italy, the overall effect was an increase to 5x in wholesale electricity prices, with peaks up to 10x [5], and a series of detrimental consequences on Italian businesses and households, too.

[2] as of April 2025

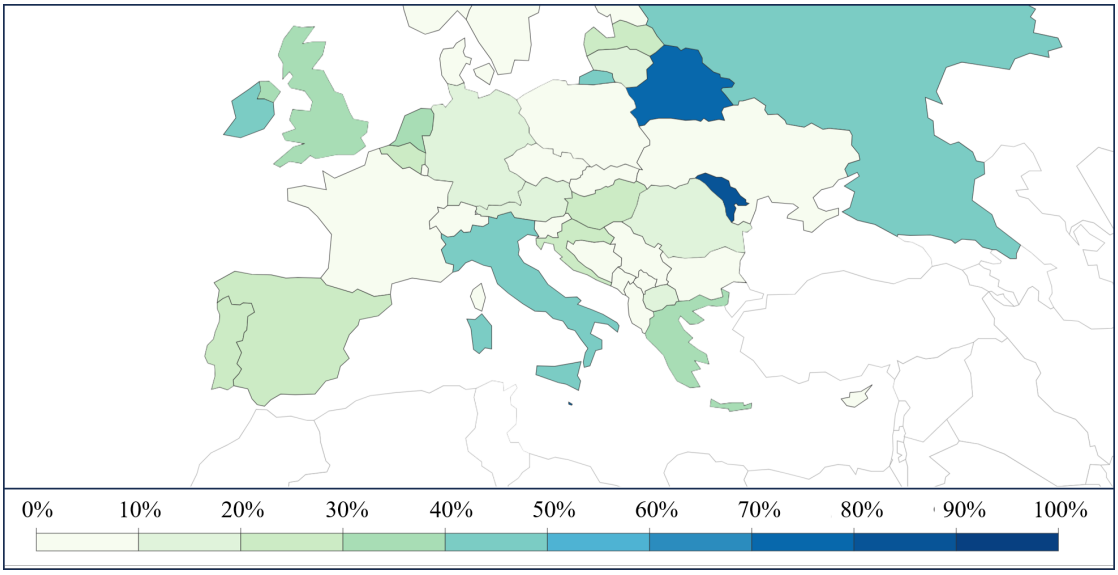

Fig. 1.3 **Share of Electricity Generation from Gas in Europe (2023).**
During the 1990s, many European countries, particularly those without a strong nuclear infrastructure, started invensting in the transition from coal-fired generation to gas-fired generation. This was driven by two motivations. First, Russia had recently abandoned communism and it was regarded as a reliable supplier for cheap natural gas. Second, natural gas is cleaner than coal, emitting nearly 50% less carbon per unit of electricity produced.

Europe's shift towards a more sustainable electricity industry began over 25 years ago, with a significant transition from coal and oil to NG. This decision was driven by two main factors: NG is cleaner than coal, and there was a readily available and affordable supply of natural gas from Russia. At the time, Russia had recently transitioned from a communist regime to a capitalist model, making it seem like a reliable commercial ally. This reliance on NG became particularly crucial for countries that did not have access to nuclear power, the only viable alternative to thermal generation in terms of stability and continuity of production. Figure 1.3 represents the share of electricity generation from gas across Europe, as of 2023. With more than 40% of NG-fired generation, Italy is a prime example of this dependency. The country's nuclear phase-out began in 1987, driven by strong anti-nuclear protests following the Chernobyl disaster. This led to the deactivation of Italy's three operating nuclear power plants by 1990. A similar issue arose in Germany, whose denuclearization started in the early 2000s, driven by political pressure from green movements. By 2021, six of the nine operational nuclear plants in Germany had been shut down, before their end-of-life. However, with the onset of the Russo-Ukrainian crisis and the loss of Russian gas supplies, Germany was forced to

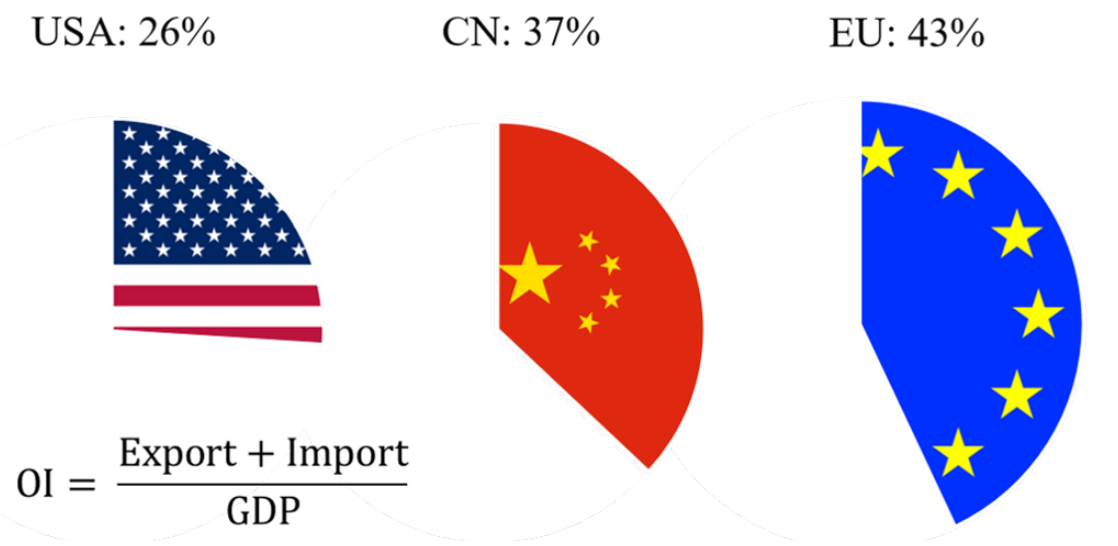

Fig. 1.4 **Trade Openness Index (OI) in 2019: EU vs China vs USA**
The Openness Index (OI) measures the extent to which an economy is dependent onto the global trade system, calculated as the ratio of the sum of Exports and Imports to Gross Domestic Product (GDP). In 2019, the European Union recorded the highest OI at over 43%. China followed with 37%, and the USA trailed at 26% . These figures underscore the EU's heightened vulnerability to geopolitical tensions and global market fluctuations, compared to China and the United States.

reactivate coal plants, a move that not only inflicted financial damage but also sharply increased carbon emissions. Italy, though to a lesser extent, also had to resort to coal-fired power plants.

From the Berlin Wall fall until a few years ago, there was a vision of a globalized and highly integrated economy centered around the United States. Draghi explains that this project has, unfortunately, failed, and we now face a global economy that is increasingly fragmented. The world is polarized between the USA and China, two major competitors with whom Europe must contend. The EU's weakness lies in its insufficiently integrated economy compared to its competitors. Member States each maintain distinct industrial regulations, making it difficult for businesses—especially innovative ones—to access a large, unified market. This fragmentation also exacerbates energy cost issues, as EU countries often compete against one another to secure supplies, driving prices higher [3]. This lack of integration proves even more detrimental to competitiveness against the USA and China, as Europe is more dependent on foreign economies. To illustrate this

dependency, Draghi uses the trade Openness Index (OI), defined as

$$\text{OI} = \frac{\text{Export} + \text{Import}}{\text{GDP}}. \tag{1.1}$$

In 2019, the EU had an OI greater than 43% [3], compared to China's 37% [6] and the USA's 26% [3]. These data clarify that the EU is much more exposed to geopolitical tensions, global market fluctuations, and disruptions in trade flows.

1.1.2 Strategic Responses and the Role of Forecasting

According to Draghi, with Europe's current economic management, the continent could remain in a phase of stagnation until 2050, relegating it to a marginal role on the global stage. To break free from this trajectory, he argues for a massive investment plan that would enable Europe to rise to the level of an economic superpower. Specifically, three sectors are outlined as priorities for intensive financing:

1. technological innovation;
2. decarbonization;
3. defense.

Regarding the present work, it is particularly relevant to focus on the first two points. Innovative companies need support because, in Europe, the previously mentioned poor integration among Member States hinder their growth. Furthermore, at the European level, there is no clear regulatory framework that, for instance, defines what constitutes an innovative enterprise. As a result, providing financial assistance through incentives and similar measures becomes more challenging, as there is no enabling legislation to foster technological innovation. We can gain some clarity with a few key data points. Of the 50 most influential tech companies in the world, only 4 are based in Europe. Additionally, between 2006 and 2021, 147 unicorns—innovative startups with private valuations exceeding €1 billion—were founded in the EU [3]. However, about one-third of these relocated to the USA, where they have access to a larger market and to simplified regulations. In essence, Europe is a significant incubator of ideas but struggles to convert these ideas into successful businesses. All of this is particularly true for companies in the

sector that, according to Draghi, represents a potential driver of transformation for entire economies: Artificial Intelligence (AI).
We will return to AI shortly, but first, we must address decarbonization, as the electricity sector is most closely tied to this effort: power generation from fossil fuels produces roughly 40% of global carbon emissions [7]. In light of this issue, green generation technologies have become the cornerstone of international strategies aimed at cutting emissions and achieving climate neutrality. Italy is also moving along this path. Figure 1.5 presents the gradual replacement of thermal generation systems based on fossil fuels with renewable energy sources. Specifically, starting in the early 2010s, a massive deployment of photovoltaic systems took place, largely driven by substantial national incentive programs. Wind power also experienced significant growth, albeit on a smaller scale compared to solar photovoltaics.
Europe is a global leader in renewable energy technologies, accounting for roughly one-fifth of the world's research in this field [3]. The issue, as with innovative companies, is that production does not take place in Europe—this time because it is too expensive. Industrial-scale manufacturing of renewable technologies in the EU costs between 70% and 130% more than in China. This is largely due to the limited availability of raw materials, which, like NG, Europe is forced to import from abroad. To put this into perspective, from 2017 to 2022, global demand for lithium grew by 300%, cobalt by 70%, and nickel by 40%. When considering all these materials together, Europe controls only 2% of the world's mines, compared to China's 43%. In his conclusions, the former prime minister is unequivocal: Europe will not achieve autonomy in the decarbonization process.
The energy sector is precisely one of those in which, according to Draghi, advanced computational tools, including AI, could make a decisive contribution. At this point, it is important to clarify that the term AI is used here in a broader sense than its more recent, mainstream interpretation—largely associated with language models such as GPT. In its original and more comprehensive meaning, AI refers to the entire class of quantitative and algorithmic methods that enable machines to perform tasks typically requiring human or super-human intelligence. These include, but are not limited to, pattern recognition, optimization, decision-making under uncertainty, and predictive modeling, and many of them have long been applied in energy forecasting, including electricity price prediction. Importantly, nearly all efforts in the ecological transition have focused on the so-called Non-Programmable Renewable Energy Sources (NPRES),

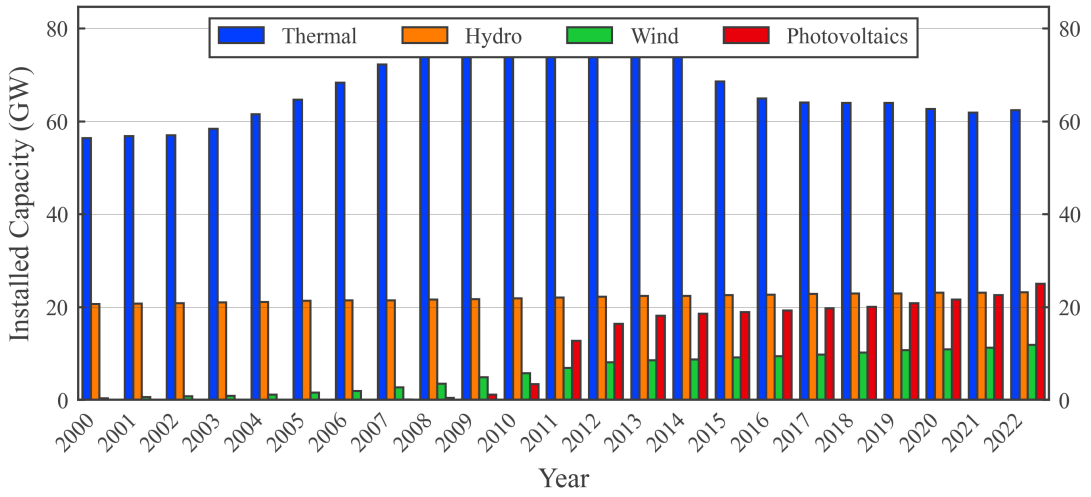

Fig. 1.5 **Evolution of Installed Generation Capacity in Italy (2000-2022).**
The figure depicts the gradual replacement of thermal generation systems from fossil fuels (Thermal) with renewable generation systems in Italy. Specifically, from the early 2010s, ha massive installiation of photovoltaic generation systems has been conducted, incentivized by massive national funding programs.

such as photovoltaic and wind energy. These sources are much easier to implement compared to programmable ones like geothermal and hydroelectric energy, which are relatively scarce across Europe[3], especially in Italy. The energy production from NPRES cannot be controlled or scheduled on demand. Unlike conventional (i.e., thermal or nuclear) power plants or even some renewable sources like hydroelectric or geothermal, which can adjust output based on need, NPRES depend on natural conditions—sunlight and wind. This unpredictability poses a challenge for maintaining a stable energy supply, especially when these sources make up a significant portion of the energy mix. The inherent unpredictability of generation from NPRES leads to electricity markets exhibiting highly volatile and difficult-to-predict prices, and it is precisely within this context that our research project takes place. This work focuses on traditional statistical learning algorithms, which provide a robust and interpretable framework for addressing energy forecasting challenges. A practical and well-defined example is EPF, where forecasts enable market operators to anticipate price fluctuations, optimizing buying and selling strategies and improving the efficiency of the energy system.

[3]Hydroelectric power is relatively abundant in Northern Europe, though not to the extent seen in countries like Brazil, where the abundance of large rivers allows for roughly 80% of the electricity demand to be met through hydroelectric power.

1.2 Electricity Price Forecasting

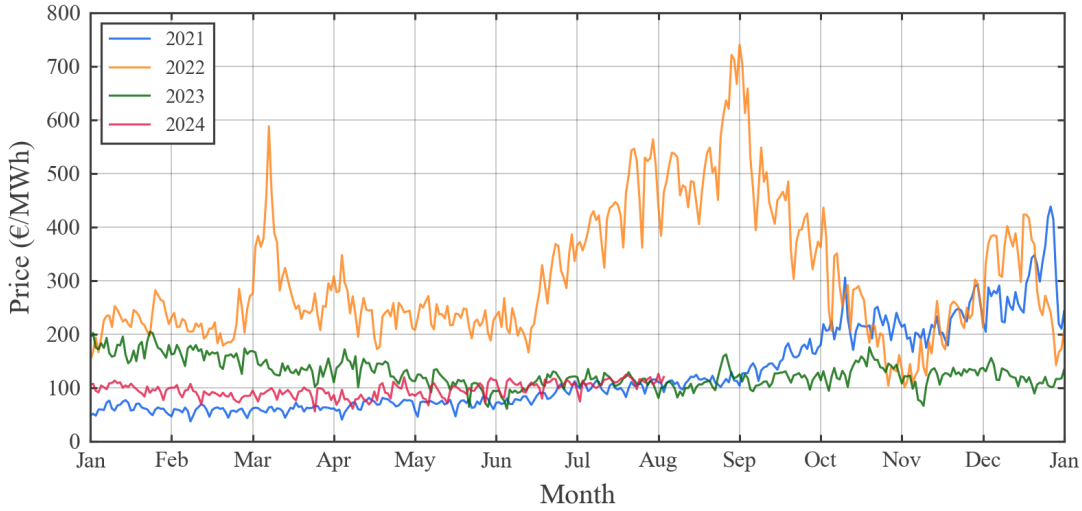

Fig. 1.6 **Electricity price evolution in Italy, daily averages (2021-August 2024).**
The chart highlights unusual patterns in the Italian day-ahead electricity price over recent years. During 2022 (orange), the escalation of the Russia-Ukraine conflict triggered a severe energy crisis, boosting an upward trend that had already begun in late 2021 (blue). This earlier increase in prices was attributed to a mix of contributing factors, such as water shortages and simultaneous maintenance shutdowns of several power plants. War further intensified the situation, with price peaks coinciding with periods of increased geopolitical tension, e.g. when the EU sanctioned energy imports from Russia. By contrast, the 2023 trend (green) appears to represent a middle ground between the anomalous spikes of the preceding years and the more stable pricing observed prior to 2021. In the first eight months of 2024 (red), the data follow a trend comparable to the latter months of 2023.

A small portion of my research activities has been carried out at an Italian utility company. Utility companies are organizations that provide essential public services, such as electricity, gas or water. Starting from the final months of 2021 (Figure 1.6, highlighted in green) and dramatically throughout 2022 (highlighted in orange), Italian electricity prices reached unprecedented levels: around twelve times the average recorded over the previous decade. The situation, driven by multiple factors that will be discussed later in this thesis, triggered a sort of panic in the industrial world due to a price trend that was decidedly unpredictable in the long term. Electric utilities are particularly vulnerable to such price fluctuations, as they operate with thin margins and are exposed to significant financial risks. In extreme cases, this can lead to severe consequences, such as the 2001 energy crisis in California, where unpredictable price spikes caused the state's primary utility operator, to declare bankruptcy [8].

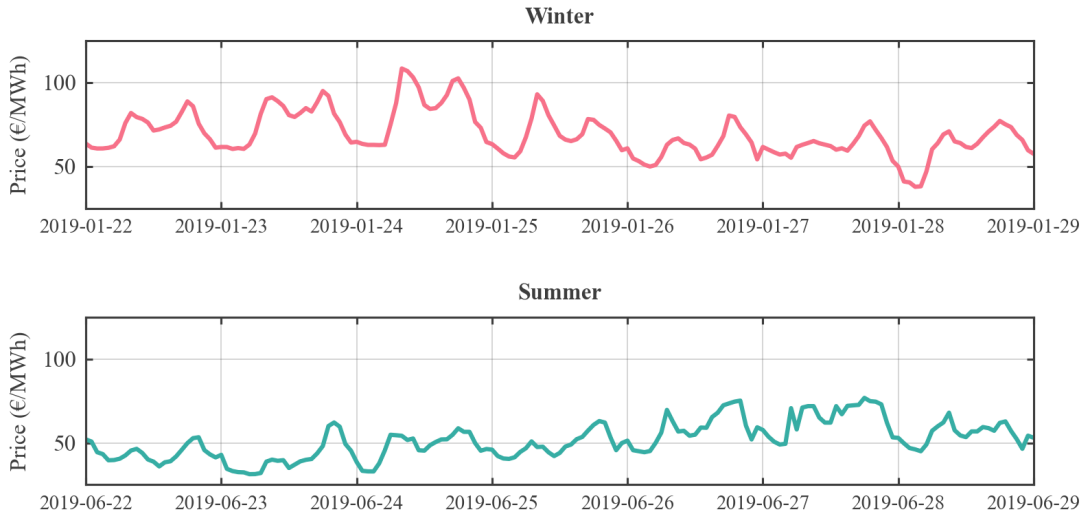

Fig. 1.7 **Seasonal Variations in Day-Ahead Electricity Prices (2019)**
The figure showcases two example weeks of Italian electricity prices in the day-ahead market for 2019. The top section, in red, depicts a winter week, while the bottom section, in light blue, represents a summer week. Two main differences stand out: during winter, prices are generally higher and follow a more regular daily pattern. This increase in prices is driven by colder temperatures, which inflate demand for heating systems. In contrast, summer prices are lower on average due to abundant photovoltaic generation. However, this also introduces some irregularity in supply, which is reflected in the price fluctuations.

Although the situation had not escalated to a crisis comparable to the one in California, it was essential to mitigate financial risks. Accurate forecasts of the day-ahead market were crucial, as it typically represents the primary short-term source of energy for businesses in the sector[4]. Reliable predictions enable companies to optimize their strategies, reduce exposure to price volatility, and ensure more stable operations.

In the day-ahead market, every day trading occurs separately for the 24 hourly intervals of the following day. Before market closing time, each market participant can submit bids specifying the purchase and/or sale of variable energy quantities for specific hourly periods. At the end of the bidding period, an algorithm determines the equilibrium price of the market. Since offers are kept secret before closing, market participants cannot compute prices in advance. Estimating these prices on a day-ahead basis undergoes the research field of short-term Electricity Price Forecasting (EPF) and is the central focus of this thesis.

[4]For energy companies, the procurement of electricity takes place in multiple stages and through various markets. As the delivery time approaches, the volumes traded typically decrease, with the day-ahead market being one of the most significant for securing short-term adjustments.

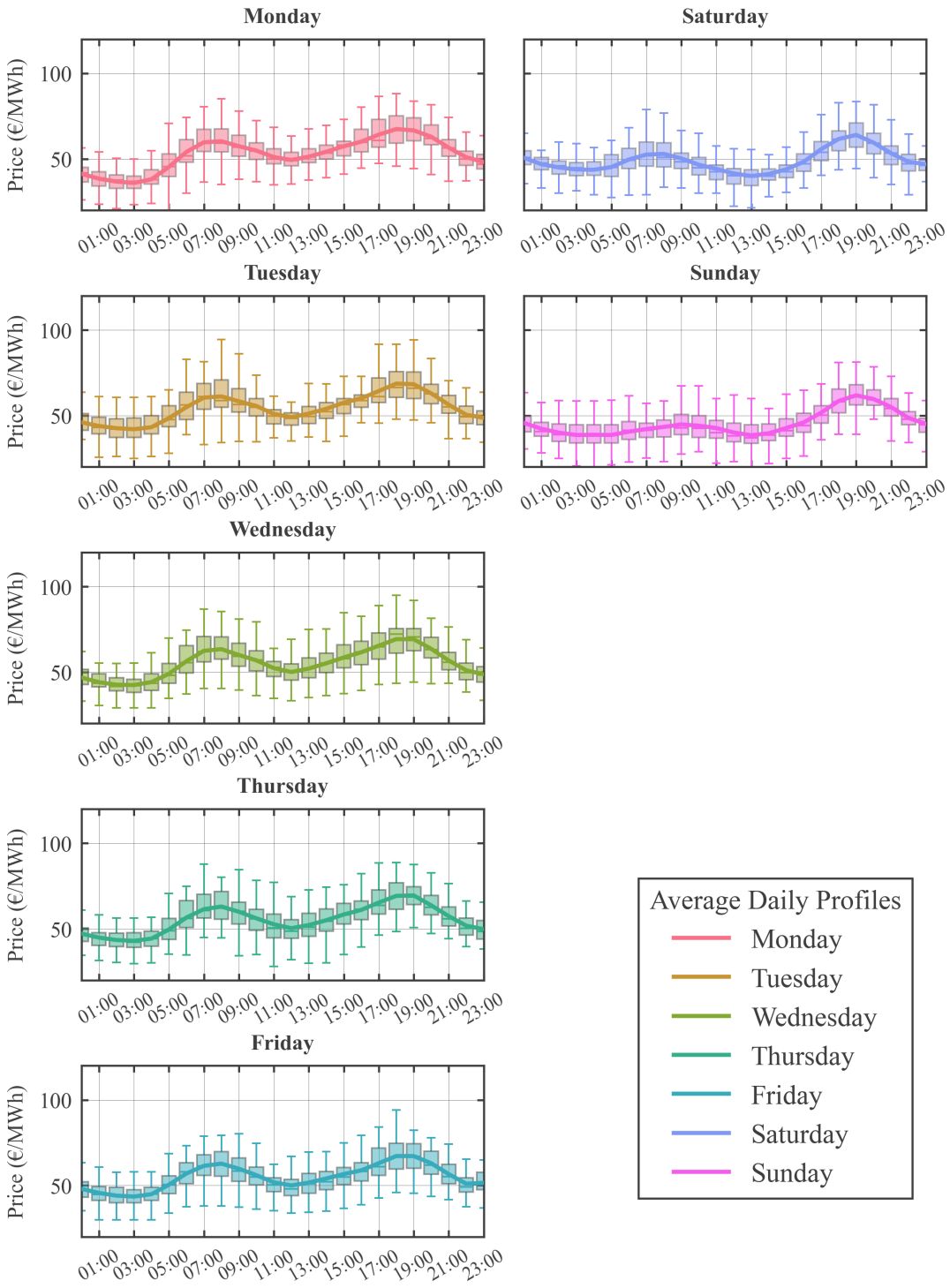

Fig. 1.8 **Hourly Electricity Price Profiles: Weekdays and Weekends (2019)**
Each subplot shows the daily average price profile (colored lines) alongside the full distribution (boxplots in the background). The x-axis represents hours of the day (00:00–23:00), and the y-axis shows the price in euros. A shared legend in the lower-right corner identifies the day of the week for each average profile.

For those in electrical engineering, it might initially seem intuitive to treat the time series of electricity prices as a signal, given the fact that it possess certain periodic patterns. For example, prices tend to be lower during the central hours of the night, when the demand for electricity is lower. Conversely, during the daytime, prices typically rise as electricity demand increases, driven by higher consumption from households, businesses, and industrial activities. However, as illustrated in Figure 1.7, a closer examination of the data reveals that this signal is highly noisy. While seasonal and daily trends are observable—such as the regular daily periodicity in winter and the influence of photovoltaic generation in summer—the presence of irregularities, sudden spikes, and unpredictable fluctuations complicates the analysis. As we will see shortly, the inherent irregularity in the price series makes it nearly impossible to generate a reliable point forecast. A point forecast refers to a single numerical prediction, such as, *"tomorrow's price at 9 AM will be €100/MWh."* For a forecast to be considered reliable, it must consistently maintain an error margin within acceptable limits, ensuring its practical utility for decision-making processes. However, the noisy and volatile nature of electricity prices, influenced by factors like weather, demand fluctuations, and supply variability, often results in predictions that fall outside these acceptable thresholds. This limitation underscores the need for alternative forecasting approaches that account for uncertainty and variability, rather than relying solely on deterministic predictions. This is further illustrated in Figure 1.8, where each subplot shows the daily average price profile (colored lines) alongside boxplots for each hour of the day. While the average profiles exhibit some regularity, the underlying distributions display substantial variability. The x-axis represents hours of the day (00:00–23:00), and the y-axis shows the price in euros. The stark contrast between the average trends and real-time instances highlights the difficulty of creating accurate statistical predictions.

1.2.1 Point Forecasting of Electricity Prices

As an example, let us consider the problem of forecasting the electricity price for Wednesday at 9 AM, one day in advance. Let Y be the random variable representing this price. A prediction can be made by constructing a model that uses a vector of p input variables $\mathbf{X} = (X_1, \ldots, X_p)$, termed predictors. The forecast is obtained as the estimate of the expectation of the next day's price Y, conditioned on the realization $\mathbf{x} = (x_1, \ldots, x_p)$ of the predictors $\mathbf{X}$:

$$\hat{y}(\mathbf{x}) \simeq \mathbb{E}[Y|\mathbf{X} = \mathbf{x}]. \tag{1.2}$$

To build intuition, it is useful to start from the simplest case: the univariate setting, where only a single predictor ($p = 1$ and $\mathbf{x} = x$ is a scalar) is considered. To this end, we select the electricity price values from the previous day, Tuesday, at the same hour of the day. Given access to a training dataset $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$—that is, a historical record of observations—we can construct statistical models that capture and reconstruct the relationship between variables. The essence of forecasting using Machine Learning (ML) lies in employing such a model—trained on historical data—to predict the values of the response variable Y based on arbitrary input values of X. A large portion of contemporary research focuses on developing, improving, and comparing algorithms—often of very different nature—that produce statistical models. We believe that this line of work is primarily technical. In this thesis, by contrast, we aim to focus on understanding the fundamental aspects of predictive modeling. For this reason, we consider as illustrative examples two methods that reflect the main distinction within ML approaches: linear models and nonlinear models. Specifically, we consider Linear Regression (LR) and Kernel Regression (KR).

- LR assumes that the expectation of the response variable Y depends linearly on $\mathbf{x}$:

$$\hat{y}_{\mathrm{LR}}(\mathbf{x}) = \theta_1 x + \theta_0, \tag{1.3}$$

  where $\theta_k \in \mathbb{R}$. The parameter vector θ is learned from data by minimizing the Mean Squared Error (MSE):

$$\min_{\theta} \frac{1}{n} \sum_{i=1}^{n} \left(\hat{y}_{\mathrm{LR}}(\mathbf{x}_i) - y_i\right)^2. \tag{1.4}$$

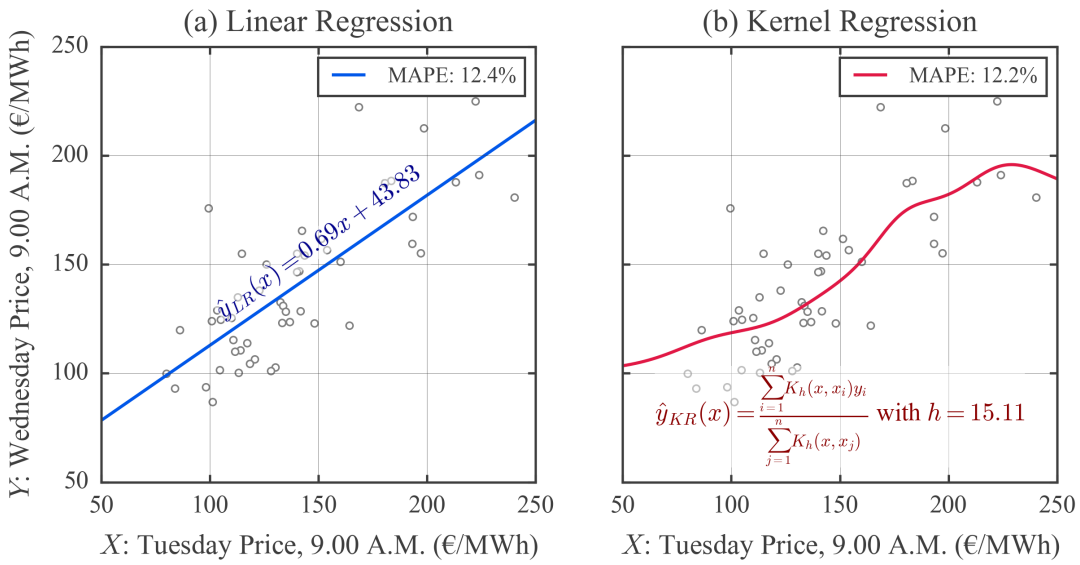

Fig. 1.9 **Comparison of Linear and Kernel Univariate Regressions**.
This chart presents a comparison between Linear Regression (LR, left, blue) and Kernel Regression (KR, right, red) in modeling the relationship between Tuesdays and Wednsedays prices over the course of 2023. The figure also reports the analytical expressions of both models along with the corresponding estimated parameters.

- In contrast, the KR is a method that estimates Y by locally averaging nearby values, weighted by a kernel function:

$$\hat{y}_{\mathrm{KR}}(\mathbf{x}) = \frac{\sum_{i=1}^{n} K_h(\mathbf{x}, \mathbf{x}_i) y_i}{\sum_{j=1}^{n} K_h(\mathbf{x}, \mathbf{x}_j)}, \tag{1.5}$$

  where K_h is a kernel function with bandwidth $h > 0$. In this study, we use the Radial Basis Function (RBF) kernel:

$$K_h(x_1, x_2) = \exp\left[-\frac{1}{2}\left(\frac{x_1 - x_2}{h}\right)^2\right], \tag{1.6}$$

  which measures the similarity between the input $\mathbf{x}$ and the training samples. The bandwidth value can be obtained by a generic unconstrained optimization method.

In Figure 1.9, we illustrate how these two models capture the relationship between input and output variables over the course of 2023. On the left panel (blue), the LR model is shown; on the right panel (red), the KR estimator is displayed. The analytical expressions of the models and the values of their respective parameters are provided. Predictive accuracy must be evaluated using an error measure. Here, for example, we consider the Mean Absolute Percentage Error (MAPE). While the MAPE has its limitations—which we will discuss elsewhere—it is particularly appealing in this context due to its ease of interpretation. For the training dataset, it is defined as

$$\text{MAPE} = 100 \times \frac{1}{n} \sum_{i=1}^{n} \frac{|\hat{y}_{\text{LR}}(\mathbf{x}_i) - y_i|}{|y_i|} \%. \tag{1.7}$$

As shown in Figure 1.9, although the two models differ both mathematically and qualitatively—producing noticeably different regression lines—they achieve very similar MAPEs, both relatively high. This leads us to the two key message of this thesis: first, the "noisy" nature of the electricity price series—characterized by significant variability—prevents the construction of an accurate regression, thus implying the need of a probabilistic approach. The second message is that, in EPF, the choice of the statistical algorithm is secondary to a thoughtful analysis of the data.

1.2.2 Probabilistic Forecasting of Electricity Prices

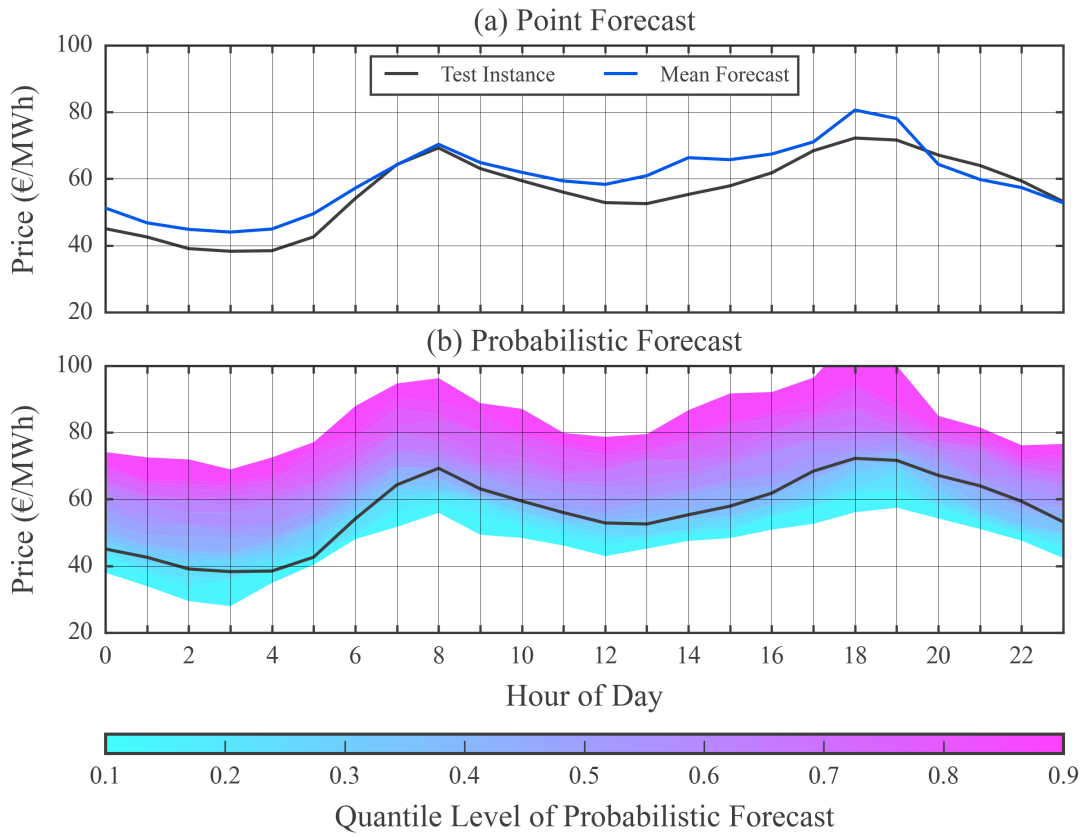

Fig. 1.10 **Comparison of Point Forecast and Probabilistic Forecast Using Quantile Regression.** *The figure compares two approaches: point forecasting (a) and probabilistic forecasting (b). In the top panel, point forecasts represent single-value predictions for each hour (e.g., the mean forecast). In contrast, the bottom panel showcases probabilistic forecasts derived from quantile regression, where shaded regions represent different quantiles, offering a range of possible outcomes with associated probabilities. The entire band defines an 80% prediction interval, as its lower and upper bounds correspond to the 0.1 and 0.9 quantiles, respectively.*

A *probabilistic forecast* provides information not only into the expected price but also into the uncertainty and likelihood of various future scenarios. A typical example of probabilistic forecasting is the following: *"Tomorrow at 9 A.M., the electricity price will range between €90 and €110 with a 95% probability."* The 95% value is often referred to as the confidence level of the forecast. Figure 1.10 provides a visual example. The top panel displays point forecasts, which provide single-value predictions for each

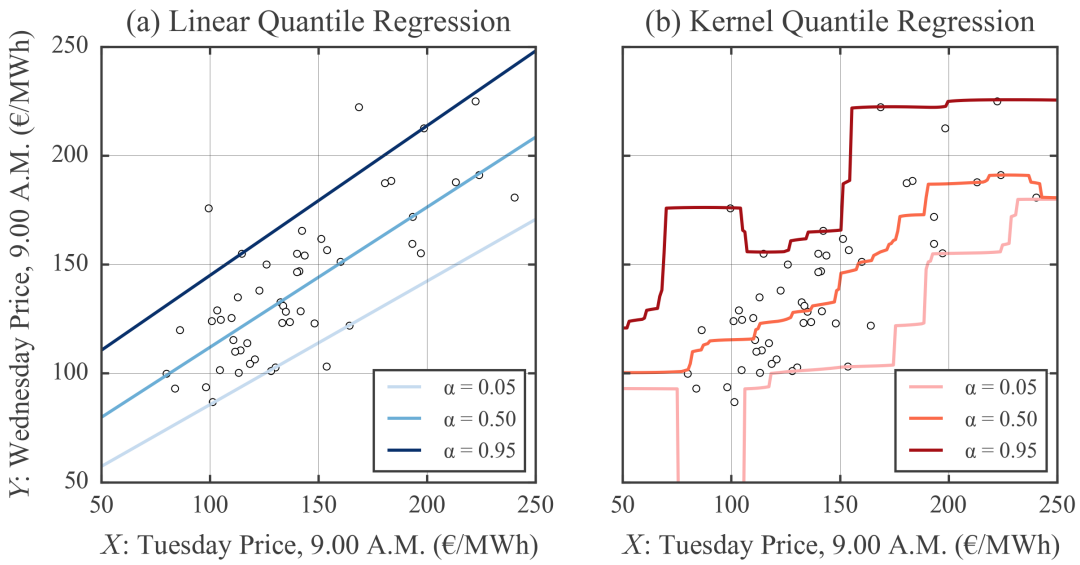

Fig. 1.11 **Comparison of Linear and Kernel Univariate Quantile Regressions**.
The chart presents a visual comparison of Linear Quantile Regression, in blue shades on the left, and Kernel Quantile Regression, in red scale on the right. The estimators model the relationship between Tuesdays and Wednsedays prices at 9 A.M. over 2023. Different shades of color indicate different quantile levels α.

hour of the day. In comparison, the bottom panel presents probabilistic forecasts, with shaded regions indicating various quantiles that capture a range of potential outcomes and their corresponding probabilities. In particular, the lower and upper limits of the colored interval are the quantiles of level 0.1 and 0.9, respectively, and thus define a prediction interval with 80% confidence.
Such estimates can be obtained using techniques from the Quantile Regression (QR) [9] family of methods. Let us recall that, for $\alpha \in (0,1)$, the α-quantile of the random variable Y, denoted as $Q_Y(\alpha)$, is the value $y^{(\alpha)}$ such that:

$$\mathbb{P}(Y \leq y^{(\alpha)}) = \alpha. \tag{1.8}$$

This quantile represents a point in the distribution of Y below which a proportion α of the observations are expected to fall. Extending this concept to conditional quantiles, for any realization x of the predictor variable X, the conditional quantile of Y given $X = x$ is defined as:

$$Q_{Y|X}(\alpha|x) = \inf\{y \in \mathbb{R} : \mathbb{P}(Y \leq y|X = x) \geq \alpha\}. \tag{1.9}$$

Consider again the problem of predicting electricity prices for Wednesday at 9 A.M., given the prices of Tuesday at the same hour. Following the same structure of the previous example, we consider here Linear Quantile Regression (LQR) and Kernel Quantile Regression (KQR), which extend LR and KR, respectively, to the estimation of the future quantile $y^{(\alpha)}$ for a given level $\alpha \in (0,1)$.

- Fix $\alpha \in (0,1)$.The mathematical representation of LQR closely resembles the standard LR form (3.2):

$$\hat{y}^{(\alpha)}_{\text{LQR}}(\mathbf{x}) = \theta_0^{(\alpha)} + \theta_1^{(\alpha)} x. \tag{1.10}$$

  Unlike LR, which minimizes the MSE, LQR is designed to minimize the average Pinball Loss (PL), defined as:

$$\text{PL}_\alpha\left(\hat{y}^{(\alpha)}, y\right) = \begin{cases} \alpha \cdot |y - \hat{y}^{(\alpha)}|, & \text{if } y \geq \hat{y}^{(\alpha)}, \\ (1-\alpha) \cdot |y - \hat{y}^{(\alpha)}|, & \text{if } y < \hat{y}^{(\alpha)}, \end{cases} \tag{1.11}$$

  where y is the observed value and $\hat{y}^{(\alpha)}$ is the predicted quantile. The $\theta_0^{(\alpha)}$'s are estimated by minimizing the Mean PL (MPL):

$$\min_{\theta_0^{(\alpha)}, \theta_1^{(\alpha)}} \frac{1}{n} \sum_{i=1}^{n} \text{PL}_\alpha\left(\hat{y}^{(\alpha)}_{\text{LQR}}(\mathbf{x}_i), y_i\right). \tag{1.12}$$

- KQR also has a functional form very similar to that of the KR. It is based upon estimating the conditional cumulative distribution function (CDF), defined as:

$$F(y|x) = \mathbb{P}(Y \leq y | X = x), \quad y \in \mathbb{R}. \tag{1.13}$$

  The KQR, in its simplest form, can estimates this CDF as

$$\hat{F}(y|x) = \frac{\sum_{i=1}^{n} K_h(x, x_i) I[y_i \leq y]}{\sum_{j=1}^{n} K_h(x, x_j)}. \tag{1.14}$$

Using the same data as in the example concerning point forecasts, a comparison between LQR and KQR is presented in Figure 1.11. For each method, the regression line is depicted in correspondence of quantile levels $\alpha = 0.05$, $\alpha = 0.50$ and $\alpha = 0.95$.

1.2.3 Data Understanding

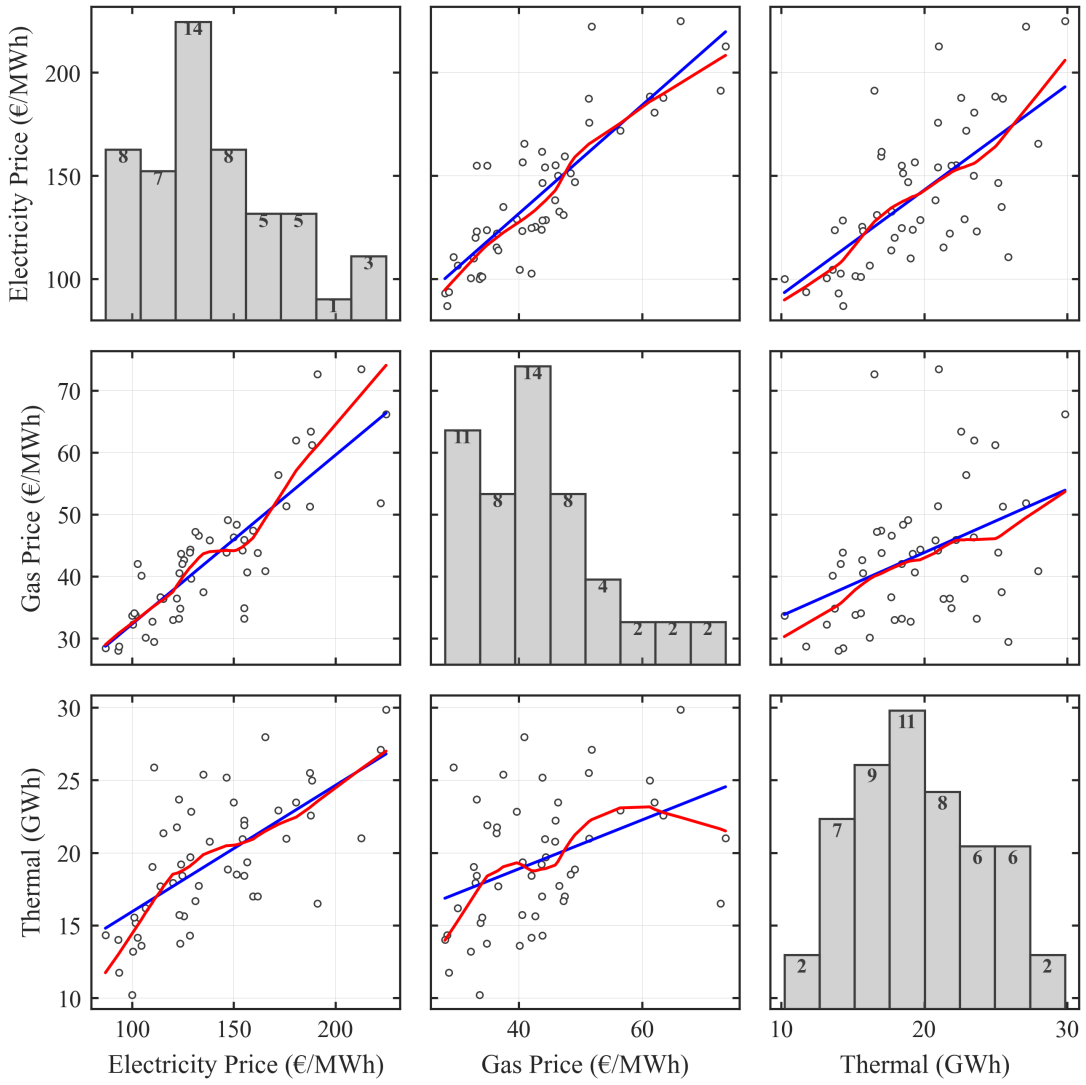

Fig. 1.12 **Matrix of Variable Relationships and Distributions with Regression Analysis**.
The figure displays a matrix of graphs where the variables are arranged identically along both the vertical and horizontal axes. Scatterplots populate the off-diagonal entries, illustrating the pairwise relationships between variables. Each scatterplot includes a blue line representing the linear regression fit, and a red line for the kernel regression fit, to emphasize correlations between the paired variables. Along the main diagonal, frequency histograms are provided, showing the distribution of each individual variable.

After reviewing these initial conceptual examples, we must move on to consider additional variables that are correlated with electricity prices. While the past values of the target variable are often the most intuitive choice in any forecasting application, in domains like EPF we are fortunate to have access to a broader set of features which—at least in principle—exhibit meaningful correlations with the target. Here, $p = 3$ input variables are considered:

1. as before, the electricity price at the same hour on the previous day;
2. the average NG price from the previous day (Gas Price);
3. the forecast value of thermal generation (Thermal).

A visualization of the variables is provided, for the year 2023, in Figure 2.3. The chart presents a matrix of graphs with identical variables on both the vertical and horizontal axis. In the off-diagonal entries, scatterplots display the relationships between each pair of variables, while along the main diagonal frequency histograms are shown per each individual variable. For each off-diagonal plot in Figure 2.3, the LR line is drawn in blue, and the KR line is drawn in red. This representation warrants at least a couple of remarks. First and foremost, it clearly reveals a strong correlation between the variables—one that is not only statistical but also rooted in fundamentals. As previously mentioned in this Chapter, thermoelectric generation accounts for approximately 40% of electricity production in Italy, and the vast majority of it is gas-fired. Coal-fired generation, once more prevalent, has now almost entirely disappeared from the Italian mix, with the notable exception of Sardinia. In this context, thermoelectric generation can be interpreted as a proxy for the supply side of the electricity market, while gas prices represent the primary financial driver of electricity prices. The second observation arises from the fact that, even when these input variables are used—each plot corresponding to a univariate regression model—the resulting mean forecast lines differ only marginally. This limited variation suggests that, at least in this context, the added complexity may not yield significant benefits. For this reason, in the final considerations of this chapter, we will focus exclusively on linear models. A similar argument applies to probabilistic regression, although it is not shown here. A more comprehensive treatment of this topic is provided in Chapter 2.

1.2.4 Multivariate Regression and Model Averaging

To integrate the predictive power of multiple variables into a statistical model, we have several options. Here, we specifically consider two key approaches that conceptually encompass all others.

1. **Multivariate Regression**. In this case—the most classic and intuitive approach—a univariate regression model is extended to accommodate multiple inputs. This is a common practice, given that most statistical literature over the past 50 years has focused on multivariate cases. For example, the LR model for mean forecasting (3.2)—when considering the three input variables previously referred to as Price, Gas, and Thermal—takes the following form:

$$\hat{y}_{\text{LR}}(\mathbf{x}) = \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_0, \tag{1.15}$$

   where x_1,x_2 and x_3 are the input values corresponding to Price, Gas and Thermal, respectively. Coefficients θ_i, for $i = 0,\ldots,3$, are determined analytically by minimizing the MSE, exactly as in the univariate setting.

2. **Model Averaging (MA)**. This approach, more modern, consists in combining the outputs of multiple individual models rather than combining their input features. In its simplest form, MA can be applied to univariate models, each trained on a single predictor. These models can be represented as:

$$\begin{cases} \hat{y}_{\text{Price}}(x_1) = \theta_1 x_1 + \theta_0^{(1)} \\ \hat{y}_{\text{Gas}}(x_2) = \theta_2 x_2 + \theta_0^{(2)} \\ \hat{y}_{\text{Thermal}}(x_3) = \theta_3 x_3 + \theta_0^{(3)} \end{cases} \tag{1.16}$$

   Here, x_1,x_2,x_3 represent the input variables corresponding to Price, Gas, and Thermal, respectively. The coefficients θ_i and intercepts $\theta_0^{(i)}$, $i = 1,\ldots,3$, are estimated by minimizing the MSE separately for each regression. The set

$$\{\hat{y}_{\text{Price}}, \hat{y}_{\text{Gas}}, \hat{y}_{\text{Thermal}}\} \tag{1.17}$$

is termed an ensemble. The arithmetic mean of the individual forecasts yields the MA prediction:

$$\hat{y}_{\text{MA}}(\mathbf{x}) = \frac{1}{3}\left(\hat{y}_{\text{Price}}(x_1) + \hat{y}_{\text{Gas}}(x_2) + \hat{y}_{\text{Thermal}}(x_3)\right) \tag{1.18}$$

This approach is appealing for its simplicity and robustness, and it has been shown to perform surprisingly well in practice—even without any model-specific tuning. More generally, one could apply MA to multivariate models as well, rather than limiting the ensemble to univariate regressions. This broader formulation will be discussed in detail in Chapter 3.

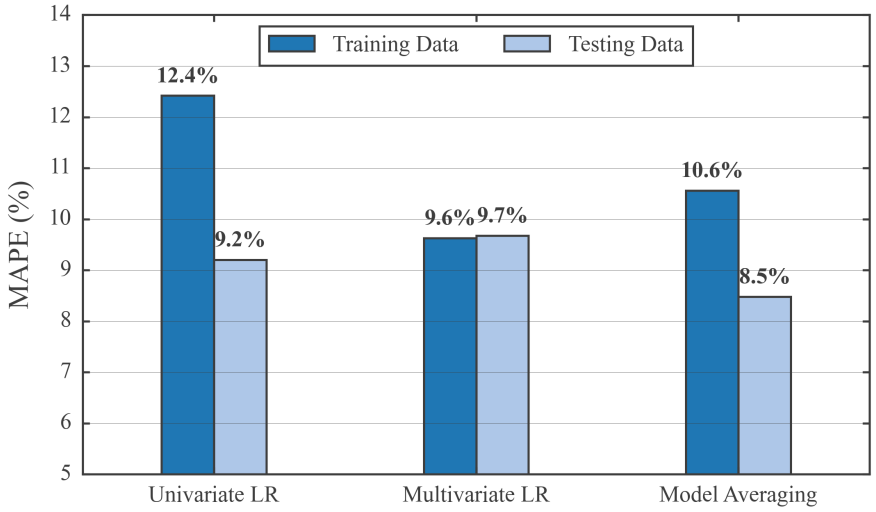

Fig. 1.13 **Comparison of Multivariate Regression and Model Averaging, Mean Forecasting.** *Test performance (light blue) is worse for the multivariate model than for the univariate one, despite lower training error (dark blue)—a sign of overfitting. In contrast, Model Averaging improves both metrics, indicating better generalization.*

At this point, an important clarification is needed. The available data are typically divided into two subsets: the training dataset $\{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$, used to fit and optimize the model, and the test dataset $\{(\mathbf{x}'_k, y'_k)\}_{k=1}^{m}$, used to assess the model's performance on previously unseen data [10, 11]. The test set plays a crucial role in evaluating the model's ability to generalize to new observations and, in time-series applications, it must chronologically follow the training data. In our case, the test dataset spans the first six months of 2024, with the intention of capturing recent market conditions. The training data, on the other

hand, covers 2023 (see Figure 1.6). Furthermore, data points corresponding to holidays have been excluded from both datasets.

While preliminary visual analysis—aimed at identifying correlations among variables—can be effectively conducted on the training data alone, any meaningful evaluation of a predictive model must ultimately rely on the test data, which become the most relevant component. In much of the forecasting literature, as well as in the subsequent Chapters of this Thesis, the focus is indeed placed on test data. However, in what follows, it is worth making a few joint considerations based on both the training and test sets—an approach that is generally advisable when initially tackling a forecasting problem.

Figure 1.13 presents a comparison of training and test MAPE for (from left to right) the univariate model, the multivariate model, and the MA model, which computes the arithmetic average of the forecasts produced by the three univariate models corresponding to the Price, Gas, and Thermal input variables. The first and most striking observation concerns the performance on the test set (light blue bars), which is our primary focus. Surprisingly, the multivariate model performs worse than the univariate one. This highlights an important fact in regression modeling: even if input variables are correlated with the target, adding more variables does not necessarily lead to better predictions. In this case, the additional variables may have introduced noise or redundancy, ultimately hurting performance. This is further supported by the training error (dark blue bars), which is significantly lower for the multivariate model. This drop suggests a classic case of overfitting—a situation in which the model becomes too closely tailored to the training data, capturing its specific patterns and fluctuations, but failing to generalize well to unseen data. In contrast, the MA approach shows improvements on both the training and test sets. The reduction in MAPE across both datasets suggests that MA achieves better generalization.

The same concepts can be extended to quantile forecasting using LQR (1.10). We consider three representative quantile levels: a low level ($\alpha = 0.05$), a median level ($\alpha = 0.5$), and a high level ($\alpha = 0.95$), applying the same training and test datasets as before. Results are shown in next page in Figure 1.14. In all three cases (displayed from left to right in increasing quantile order), the multivariate model underperforms the univariate one, while the MA demonstrates better generalization and consistently outperforms both. Interestingly, the median proves to be the most difficult to predict—its behavior closely resembling that of the mean forecast.

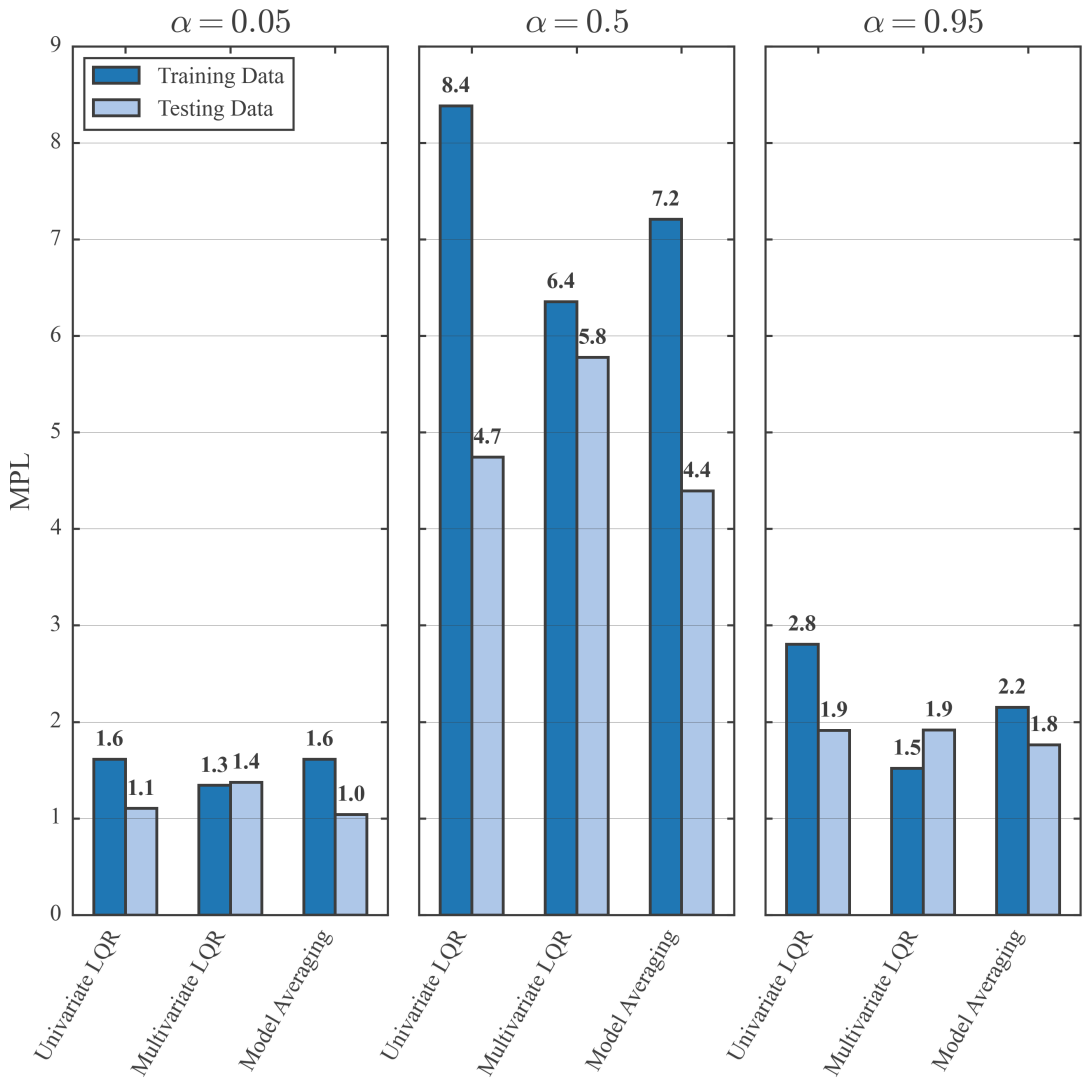

Fig. 1.14 **Comparison of Multivariate Regression and Model Averaging, Quantile Forecasting**
For quantile forecasting, the multivariate model underperforms the univariate one at all three quantile levels (0.05, 0.5, 0.95), while Model Averaging consistently outperforms both. Notably, the median (0.5 quantile) proves to be the most challenging to predict, as it behaves similarly to the mean.

Chapter 2

Linear and Nonlinear Methods for Short-Term Electricity Price Forecasting

In competitive electricity markets, prices are determined by the collective behavior of suppliers and consumers. These systems depend on the balance between supply and demand: sudden changes in underlying conditions can result in significant price fluctuations. This is the case of the Italian electricity market, which has recently undergone substantial transformation, mainly due to the growing integration of supply from Non-Programmable Renewable Energy Sources (NPRES), and the contemporary energy crises. This Chapter assesses the capabilities of some traditional forecasting methodologies in this scenario. Specifically, we compare Linear Regression (LR) and Kernel Regression (KR) for the prediction of future electricity prices, both in point and probabilistic terms. The analysis highlights the potential of both linear and non-linear models when the forecasting problem is carefully structured, including the appropriate selection of data and variables. The findings indicate that, despite their fundamental differences, linear and non-linear models are comparable in terms of performance.

The focus of the Chapter is to explain which factors a forecaster should take in account when dealing with an industrial context traumatized by sudden market transformations. Indeed, the authoritarian-led Russia launched an aggression against democratic Ukraine at an extremely delicate moment for Europe. In 2022, Italy was in the midst of a virtuous

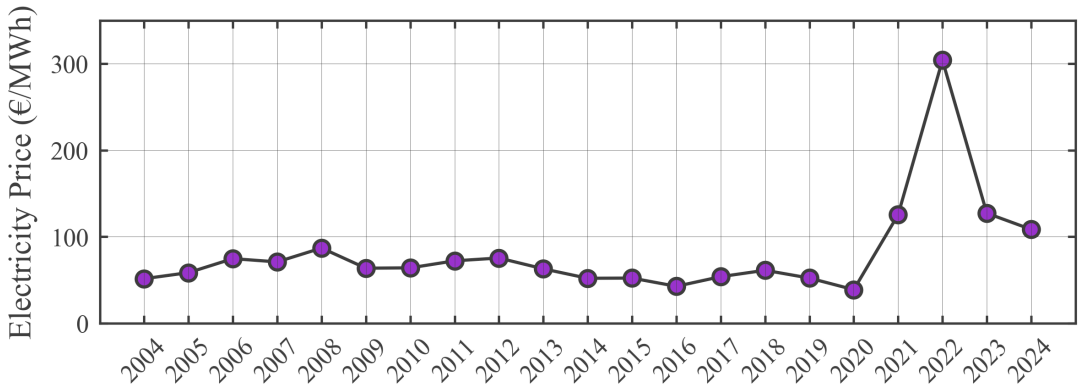

Fig. 2.1 **Evolution of Yearly Electricity Prices in Italy (2004-2024).**
The chart shows the annual averages of Italian wholesale electricity prices. After remaining below €100 for over 15 years, prices surged due to geopolitical tensions arising between Russia and Ukraine, and driven largely by market speculation.

transition toward renewable energy, a process that had been underway for about a decade and had positioned the country as a European reference point, with approximately 40% of its electricity generation coming from non-fossil sources. Furthermore, like the rest of the West, Italy was facing the challenge of economic recovery following the COVID-19 pandemic, which includes the problem of dealing with an energy demand that had surged rapidly after a period of stagnation. In essence, although Italy had made significant progress in adopting renewable energy generation, the support of gas-fired thermal generation remained indispensable. After the beginning of the Russian-Ukrainian conflict, the sanctions on Russian gas were seen as a golden opportunity by energy market speculators, who, through their operations, drove prices to unprecedented levels. These, in turn, had a direct impact on the Italian electricity market, whose prices reached peaks over ten times the average consistently recorded in the previous decade. In addition to the unpredictability of long-term price trends, another challenge emerged: volatility, referring to price fluctuations that occur over short time intervals. This volatility is largely driven by the availability of non-dispatchable renewable electricity generation. Specifically, in Italy, the main source of such generation is solar power, which produces abundantly during sunny daylight hours, less when it is cloudy, and none at night. Electricity prices reflect this pattern: for example, they are low during midday hours, and can be up to 100% higher when, between late winter and early spring, the sun sets while electricity demand remains high.

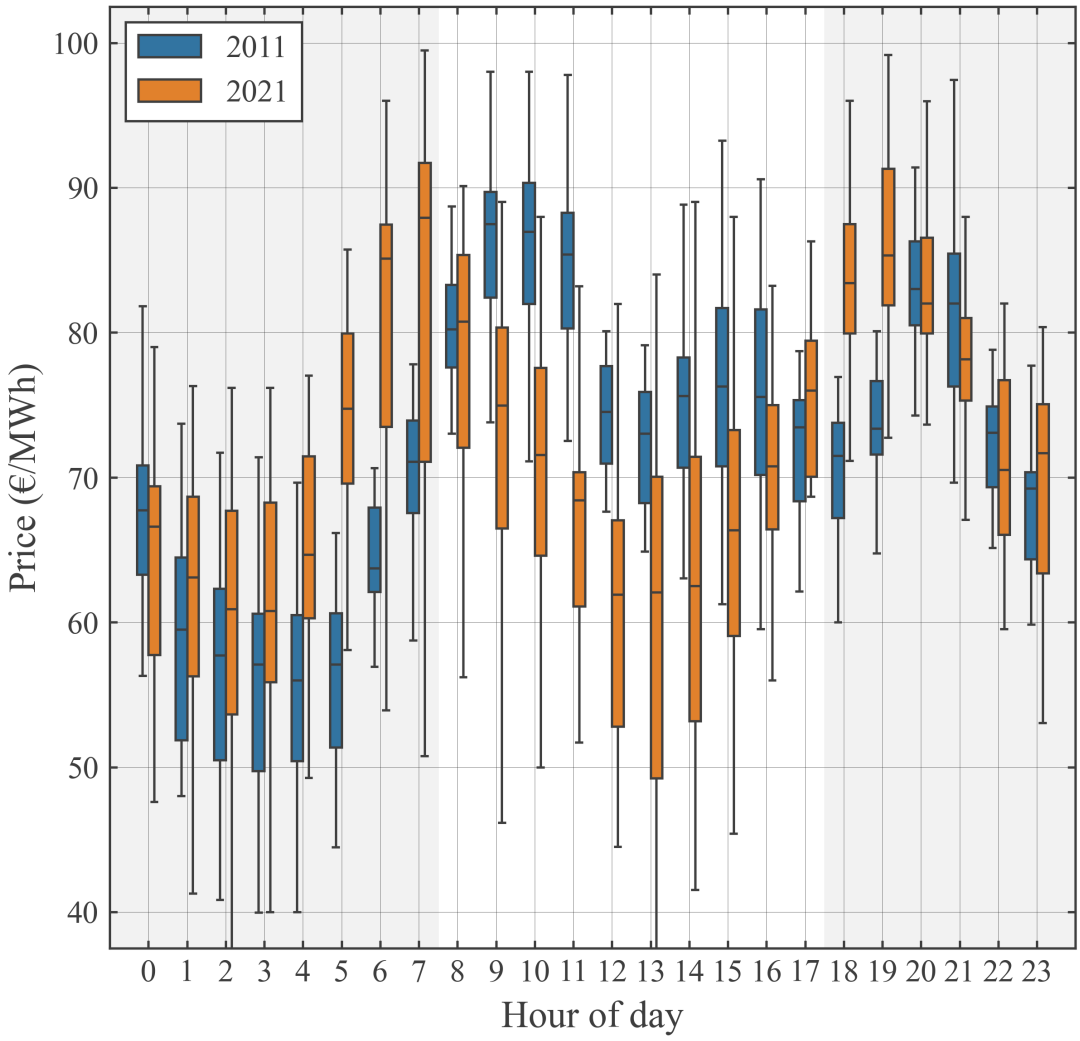

Fig. 2.2 **Evolution of Hourly Electricity Prices in May (2011 vs 2021)**.
The figures depicts the evolution of daily price profiles on the MGP, the Italian day-ahead electricity market, from 2011 (in blue) to 2021 (in orange). It is interesting to note that the percentage of power generated from photovoltaic sources, relative to the total installed generation capacity in Italy, has significantly increased over the years. In 2011, it accounted for less than 2%, while in 2021, it has surpassed 10% [5]. Each hourly period in the y-axis corresponds to two boxplots, each representing the distribution of prices for a specific year. The box represents the interquartile range (IQR), with the median marked by a line within the box. Whiskers extend from the box to the minimum and maximum values within 1.5 times the IQR. Outliers, namely values falling outside the whiskers, are here omitted for clarity.
The significant increase in photovoltaic generation capacity has caused a decrease in price during the central hours of the day, when irradiation is usually abundant, as is evidenced by the light area in figure. Also, an higher level of volatility can be observed, with significantly larger IQRs for 2021. By contrast, hours with few or no solar generation available exhibit higher average prices. Considering EU's contemporary energy policies, these tendencies are expected to increase in the near future.

2.1 The Evolution of Electricity Price Forecasting. A Brief Literature Review.

The field of Electricity Price Forecasting (EPF) is a subset of the broader and long-established domain of energy forecasting. For a comprehensive review of this whole field, we refer the Reader to [12]. This discipline includes a wide range of methods designed to predict future values of key variables in the energy sector, including commodity prices and the technical parameters of distribution networks, such as electric load or available generation. Energy commodities, in particular, present several difficulties from a predictive standpoint, making them a focal problem for both industry practitioners and academic researchers. Consequently, numerous experts in statistics, econometrics, and engineering have devoted significant attention to addressing these issues. The massive financial interests within the energy sector have driven advancements in EPF, a trend further amplified by the rapid expansion of NPRES, which now represent a global industry. Consider for example the Italian electric systems, which can be regarded as a good representative of the European scenario [13]. The large proportion of NPRES that has been recently integrated into the Italian generation system has added complexity to EPF. At the end of 2021, Italy's gross final electricity consumption was covered by 36% from renewable sources, of which roughly one third of NPRES, and two thirds of programmable sources, mainly geothermal and hydroelectric [5]. The percentage is set to increase, and precisely to double in next few years. Indeed, investments in NPRES projects, which involve European funding, are regulated in EU member States by a government document known as National Energy and Climate Plan (NECP). According to the 2023 version of the Italian NECP, approved by the EU Commission [14], the percentage of consumption to be covered by renewables by 2030 has been raised to 65%. The presence of NPRES like photovoltaic and wind, introduces significant fluctuations in electricity generation, resulting in increased price volatility. This can be understood with the help of Figure 2.2, which illustrates the difference between daily price profiles in May for year 2011, in blue, and 2021, in orange. The substantial rise in photovoltaic generation capacity, from less than 2% to almost 11% in ten years [5], has led to a reduction in prices during the midday hours when sunlight is typically abundant, as indicated by the light grey area in the figure. Additionally, there is an increased level of volatility, evident in significantly larger InterQuartile Ranges (IQRs) for the year 2021.

Conversely, hours with limited or no solar generation show higher average prices.
Going back to the domain of commodities, it is crucial to recognize that electricity possesses the characteristics of an unconventional commodity. Unlike resources such as oil or natural gas, electricity cannot be easily stored or preserved. Once generated, it must be transmitted, distributed, and consumed immediately, as any imbalance between supply and demand can create significant network issues and, consequently, financial anomalies. This makes the management and handling of electricity considerably more complex, translating to time series data that are inherently more difficult to predict. Although electricity storage technologies exist, as of nowadays, they are still financially inefficient for large-scale applications [15].
A solid contribution to electricity forecasting is presented in [16], which also explains the unique features of electricity as a commodity. The study emphasizes that forecasters primarily focus on two key variables: electricity price and electric load. Among these, load forecasting is a more mature discipline, as the time series representing electric loads, typically on national scales, exhibit greater stability and clear regularity patterns. These characteristics enable the development of highly accurate point forecasting models using traditional statistical approaches. Above all, the well-established LR family of methods, which includes ordinary and penalized least-squares methods, allows the achievement of solutions that satisfy industrial requirements. It is also important to note that load forecasting has a longer history than EPF, both as an industrial necessity and as an academic field. Accurate load predictions are critical not only for market operations but also for the effective management of electric grids. In fact, many of the initial techniques applied in EPF were adaptations of statistical methods originally developed for load forecasting.
Interest in EPF emerged with the advent of deregulated energy markets. Notably, the first comprehensive review of this field, focusing on point forecasting and day-ahead markets, appeared only relatively recently, ten years ago [17]. Before delving into further classifications, it is essential to clarify that day-ahead EPF, which is the focus of this work, falls within the domain of short-term forecasting. This category typically covers forecasting horizons ranging from a few hours to a few days. In an industrial context, short-term forecasting is particularly significant for daily supply management and speculative trading on a day-to-day basis. In contrast, medium- and long-term forecasting serve different purposes, such as supporting risk management and informing investment decisions. However, these applications extend beyond the scope of this Thesis.

According to the author of [17], the decade leading up to his work saw a significant expansion of research in EPF, with a wide array of solutions being proposed. However, he pointed out lack of coherence within the field. This was largely attributed to variations in the markets studied, which can be substantially different depending on the examined country, and the absence of universally accepted criteria for modeling and evaluating forecasting approaches. Despite progress in addressing these issues over the past ten years, the field has yet to establish a standardized framework [18]. Apart from the differences across national markets [19], it is important to note that the field remains primarily divided between engineers and statisticians, each group adhering to its own methodologies and conventions [20].

In this Chapter, we analyze and evaluate KR methods, which fall under the category of Computational Intelligence (CI) or nonlinear statistical algorithms. As previously mentioned, the most established category of methods is linear statistical models, specifically LR, which we use as a reference point. These models are considered reasonable benchmarks because the question of whether CI methods outperform linear approaches remains a subject of debate [13]. As we will see in Section 2.3, while CI methods have the ability to capture the complexities intrinsic to nonlinear relationships, their practical advantage over LR, in terms of accuracy, is not guaranteed. Among the various CI methods, Neural Networks (NN) are particularly popular [17, 20–23], encompassing architectures such as multi-layer perceptrons, recurrent NNs, and convolutional NNs. However, our decision to focus on kernel methods instead of NNs depends on several factors. KR methods are conceptually simpler, easy to implement, and generally less data-intensive. Furthermore, recent literature shows that, in the Italian context, nonparametric kernel methods perform better than NNs [14], while the superiority of kernel methods to autoregressive linear models is still discussed [24, 13]. Importantly, the issue of sample efficiency in NNs is well-documented and remains an active area of research [21–23]. Deep learning models, in particular, often require carefully designed data preprocessing techniques to achieve optimal performance. For example, [22] highlights that price spikes in electricity price series are highly problematic for model training, proposing an oversampling system to correct these anomalies. Similarly, [21] introduces a generative model for data augmentation. Another recent study, [23], specifically examines the problem of sample efficiency in deep NNs, noting that data insufficiency is particularly acute in the context of EPF. This is largely due to the low-to-moderate sampling frequency (e.g., hourly data) and the dynamic nature of electricity markets, where market rules and operator preferences

evolve rapidly, often rendering older data unsuitable for predictions. In [23], too, the suggested approach involves data augmentation through the generation of synthetic data. Without delving into this topic, as it lies beyond the scope of this work, we limit to expressing some skepticism regarding the applicability of data augmentation techniques for EPF.

Other EPF methodologies involve simulation models—both statistical and multi-agent-based—fundamental models that explicitly represent supply and demand dynamics, and similar-day techniques [17, 20]. For instance, [25] presents a simulation-based approach where probabilistic forecasts are derived by aggregating simulations performed with established econometric models. This study focuses on the German and Austrian markets, though the data analyzed predates 2018. Conversely, the fundamental modeling approach is explored in [26], where bid data from the Spanish day-ahead electricity market are interpolated using basic statistical models to reconstruct the aggregated supply and demand curves. The predicted price is then calculated as the intersection of these curves and assessed with various evaluation methods. In contrast, multi-agent models simulate the collective behavior of numerous market participants on a large scale. Also in this case, price forecasts are determined from the intersection of the simulated supply and demand curves. Lastly, similar-day methods involve identifying the most comparable day (or days) from a historical database based on factors such as the day of the week, load, weather, and other relevant indicators, then using these days as forecasts for the current day. The kernel-based method used in this study can also be viewed as a type of similar-day approach, as some kernel functions can be interpreted as affinity measures. In a broader sense, it is crucial to note that the boundaries between different forecasting methods are not rigid, and many approaches are hybrid [17, 20].

Another degree of freedom in EPF research involves the selection of predictor variables, for which there is no universally agreed-upon approach. While the debate over whether multivariate models are superior to univariate models remains open [27], the inclusion of exogenous variables[1] is widely adopted. Commonly incorporated variables include those related to load and supply, as well as weather factors such as temperature, wind speed, and irradiation. Additionally, the prices of other commodities, such as Natural Gas (NG), are often included as predictors. It is important to emphasize that the choice of variables is highly dependent on the specific market being analyzed, meaning that the most significant exogenous variables can vary from State to State. For instance, different

[1] Exogenous variables are those that are not part of the price series itself (the endogenous variable).

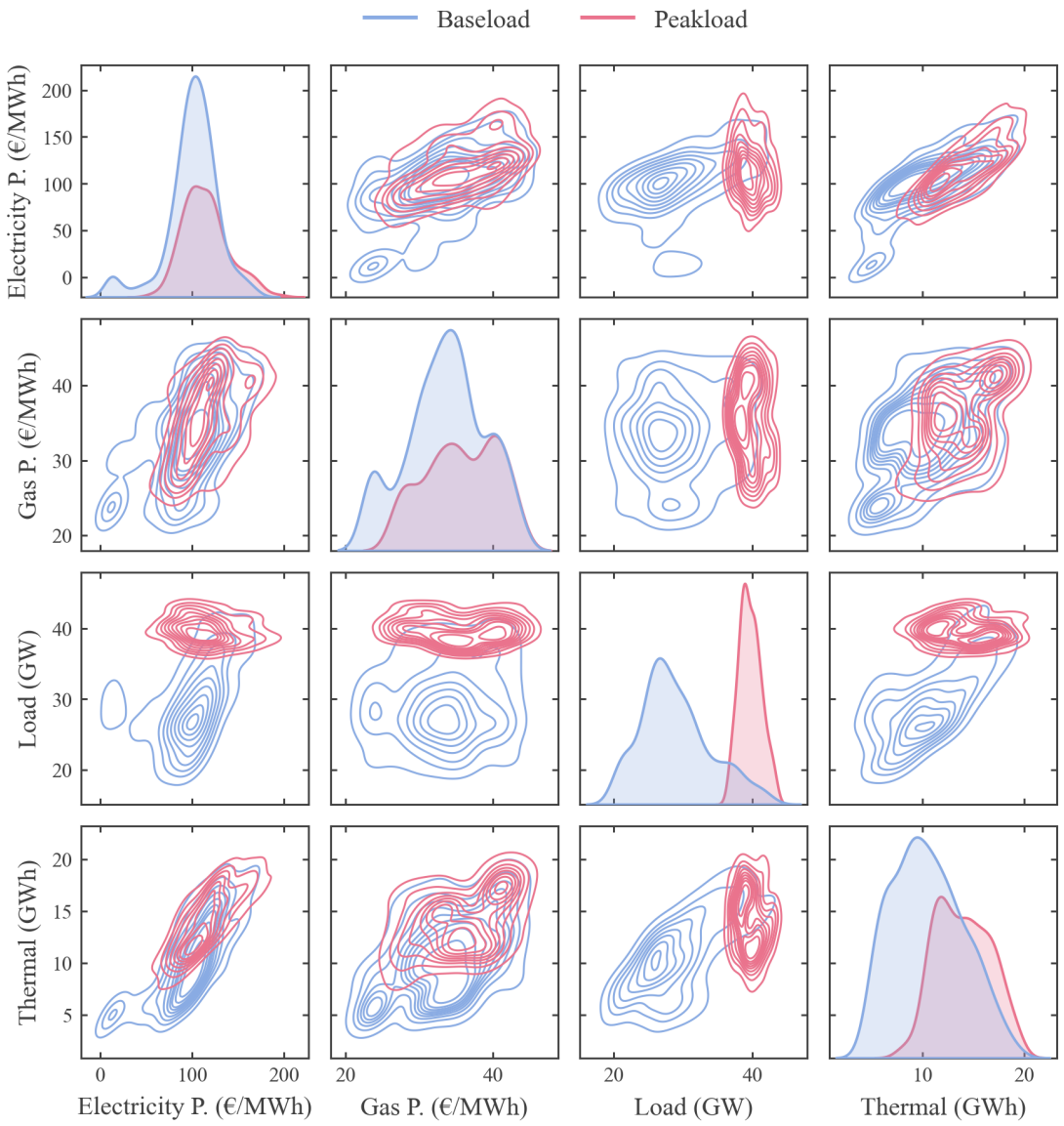

Fig. 2.3 **Disparities Between Peakload and Baseload (May 2023).**
The figure highlights the distinction between periods categorized as "peakload" and "baseload", using May 2023 data as a reference. In Italy, the peak period spans from 8:00 AM to 8:00 PM, Monday to Friday, while the off-peak period includes the remaining hours and the entire weekend.
The chart presents a matrix of graphs with identical variables on both the vertical and horizontal axes, namely electric energy price (Electricity P.), natural gas price (Gas P.), national electric load (Load), and thermal generation (Thermal). Baseload data is represented in blue, while peakload data is depicted in pink. Along the main diagonal, Kernel Density Estimation (KDE) plots for each considered variable are showcased, reconstructed using a Gaussian kernel. In the off-diagonal entries (i.e. entries corresponding to different variables), contour plots illustrate the two-dimensional KDE applied to the respective variable pairs, providing a visualization of the joint probability distribution.
The graph explains the rationale behind the terminology "peakload" versus "baseload": focusing on the 'Load' variable, the data for the baseload period (in blue) exhibit a lower mean compared to the data for the peakload period (in pink). This distinction is highly significant in the industrial world, and many wholesale contracts are based on it.

European countries have vastly different generation mixes, which directly impact price formation. There are also exceptional cases, such as a study on the UK market [28], which highlights the limited usefulness of exogenous variables in that particular market. Another relevant study [19] investigates an additional factor thought to influence market price formation: the effect of integrating European electricity markets. While the authors focus on the Dutch market, their findings more broadly demonstrate how the performance of various models, both statistical and CI, can differ significantly across markets. The only widely used variable that we did not consider in this thesis is the weather variable, primarily because weather forecasts are typically available to companies under subscription and are not easily accessible to researchers. Further, prominent research products, such as [17] and [20], clarify that short-term EPF models often incorporate weather variables indirectly. For example, instead of directly using solar irradiance, models might use photovoltaic generation, or replace wind speed with wind generation, and so on. The same study also explains that pure numerical weather predictions are useful for generating few-days-ahead forecasts, as estimates for NPRES generation are often only available for the next day.

Taking a step forward, although the key variables influencing electricity price formation are well-known, integrating their predictive power into a single, standalone forecasting model proves difficult in practice. In fact, the accuracy of predictors such as electric load or hourly generation heavily depends on the specific conditions at the time of the forecast or on factors that are difficult to identify and isolate in advance. In other words, the same predictor can perform well or poorly depending on circumstances that are often unpredictable beforehand. In the next Chapter 3, we will address a promising solution to this problem, which is a hot topic in contemporary EPF research: Model Averaging (MA) [20]. Rather than selecting the single model that performs best, one can consider a set of diverse models, in terms of both input variables, training data and statistical algorithms, whose individual forecasts can be subsequently aggregated using various techniques.

Regarding the selection of temporal lags for price data, we opted for models of order one, which are effective in capturing the inherent daily seasonality of the dataset [13]. Although the time series was not explicitly decomposed into trend, seasonal, and residual components, as this technique is predominantly associated with classical statistical methodologies, exceptions can be found in the literature. For instance, [21] explores a hybrid approach combining a generative NN with seasonal decomposition, augmented by frequency analysis using Variational Mode Decomposition (VMD) [29], applied to

each component.

The growing interest in probabilistic forecasting methods reflects the challenges posed by markets increasingly influenced by less predictable generation sources, as discussed in the previous Section. Despite its importance, probabilistic EPF remains rather underexplored. To date, the most comprehensive review in this domain is presented by [20], which offers an overview of existing techniques. These methods, similar to point forecasting, exhibit significant variability in their nature and conventions. Probabilistic forecasts can be generated using several approaches, the two primary ones being Prediction Intervals (PIs) and the modeling of the entire distribution of the variable of interest. While the latter encompasses PIs as a subset, PIs remain the most commonly utilized form of probabilistic forecasting due to their interpretability and ease of evaluation. For further insights into the evaluation of probabilistic models, the Reader is encouraged to consult [20] and [30]. Interestingly, [20] also notes that the majority of probabilistic EPF studies are based on NNs, which aligns with the approach introduced in the seminal work on this topic [31].

2.2 Methodology

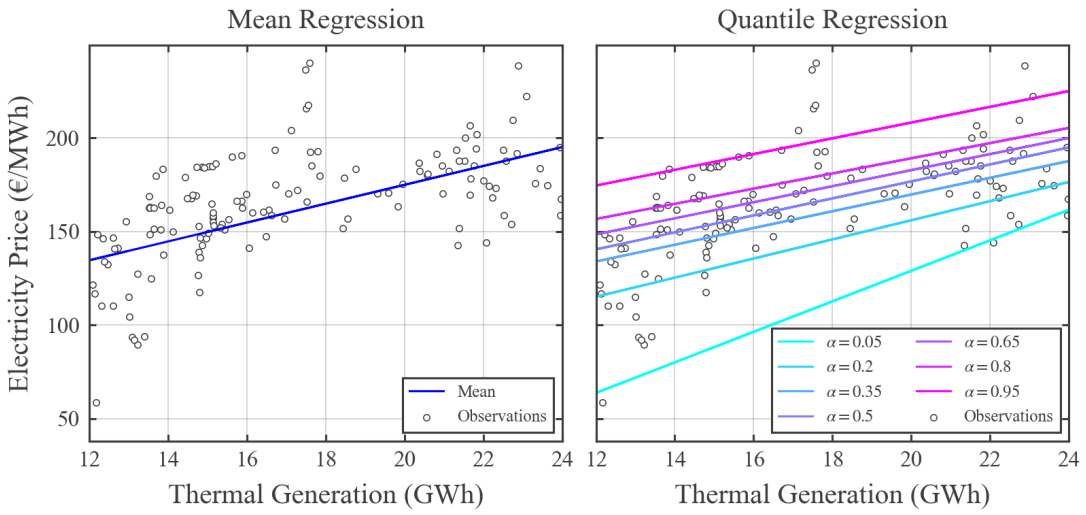

Fig. 2.4 **Comparison of Linear Mean Regression and Quantile Regression**.
The figure provides a visual comparison of linear mean regression and quantile regression for electricity price (€/MWh), as a function of thermal generation (GWh). On the left, the mean linear regression model is shown. On the right, the quantile regression model is displayed for different quantile levels, capturing the conditional distribution of electricity prices across different levels. Observations cover the third week of 2023.

The development of a comprehensive EPF system entails two regression tasks: mean regression, aimed at producing point forecasts, and Quantile Regression (QR), for constructing distribution estimates without reliance on Gaussianity assumptions. As a baseline, we consider linear prediction methods. Conversely, the decision to take kernel methods as representative for nonlinear models is motivated, on the one hand, by the properties of CI methods as outlined earlier, and, on the other hand, by the intent to avoid more complex techniques, such as those involving NNs, which may introduce unnecessary complications without offering clear advantages. Further, as will be discussed below, KR enables the formulation of models that are conceptually analogous for both mean and quantile forecasting. It is also relevant to note that the a similar choice is taken in one of Koenker's most important works [9], where the primary theme is extending linear models to to quantile prediction tasks, and Kernel Quantile Regression (KQR) is later presented as an alternative.
Additionally, some theoretical insights will be provided on the method used for ex-

ploratory data analysis, which is an essential step when implementing a predictive model. Without visualizing the behavior of the variables under consideration, there is a risk of wasting time by potentially including variables that are not correlated with the target variable.

2.2.1 Linear Regression and Linear Quantile Regression

LR is an essential approach to modeling the correlation between a dependent variable Y and a single explanatory variable X. It assumes a relationship of the form:

$$Y = \beta_0 + \beta_1 X + \varepsilon, \tag{2.1}$$

where β_0 is the intercept, β_1 is the slope coefficient, and ε is an independent error term, typically assumed to follow a Gaussian distribution with zero mean and constant variance [32]. The parameters β_0 and β_1 are estimated by minimizing the Mean Squared Error (MSE):

$$\text{MSE} = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2, \tag{2.2}$$

where $\{(x_i, y_i)\}_{i=1}^{n}$ are the observed data points. In this context, the MSE is oftern termed the Residual Sum of Squares (RSS). Indeed, for each $i = 1 \ldots, n$, the quantity

$$e_i = y_i - \beta_0 - \beta_1 x_i \tag{2.3}$$

is termed i-th residual for the LR model. The analysis of these quantities is essential for assessing the model's goodness of fit. Since the approach relies on the assumption of normality, the residual distribution must be (approximately) Gaussian. In any case, the frequency of values that deviate significantly from normality should be either zero or very low.
Once the regression coefficients have been determined, we can view a generic input x as a realization of the random variable X. The expected value of Y, conditioned on this realization, can then be estimated linearly as:

$$\mathbb{E}[Y \mid X = x] \approx \hat{y}(x) = \beta_0 + \beta_1 x, \tag{2.4}$$

where $\hat{y}(x)$ is the prediction associated to x. We now need to choose an error metrics. Keeping in mind that the objective is to forecast electricity prices, expressed in euro, we require an interpretable metric. The Mean Absolute Error (MAE) is an excellent choice, since it is also expressed in euros and is more robust to anomalies than the MSE. Let us denote predictions as $\hat{y}_i = \beta_0 + \beta_1 x_i$ for each $i = 1 \ldots, n$; the MAE is defined as

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \quad (€). \tag{2.5}$$

At this point, it is worth noting that in forecasting applications—of any kind and particularly in EPF—the Mean Absolute Percentage Error (MAPE) is widely used. It is defined as

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{|y_i|} \quad (\%). \tag{2.6}$$

This is an unfortunate choice, as MAPE tends to be high for low prices and vice versa. In this context, it is not optimal because, from an industrial perspective, accurately predicting high prices is more important than low prices, as they are associated with greater financial risk.

Moving on to probabilistic methods, Linear QR (LQR), generalizes the LR framework by modeling the conditional quantiles of Y given X. For a given quantile level $\alpha \in (0,1)$, LQR models the α-th quantile of Y, conditional on $X = x$, as:

$$Q_Y(\alpha | X = x) = \beta_0^{(\alpha)} + \beta_1^{(\alpha)} x, \tag{2.7}$$

where $\beta_0^{(\alpha)}$ and $\beta_1^{(\alpha)}$ are the intercept and slope coefficients specific to the level α, and x is a given realization of the input variable X. The parameters are estimated by minimizing the Mean Pinball Loss (MPL) [9]:

$$\text{MPL} = \sum_{i=1}^{n} \rho_\alpha \left(y_i - \beta_0^{(\alpha)} - \beta_1^{(\alpha)} x_i \right), \tag{2.8}$$

where $\rho_\alpha(u)$ is the pinball loss function, also known as check function, defined as

$$\rho_\alpha(u) = \begin{cases} \alpha u & \text{if } u \geq 0; \\ (\alpha - 1)u & \text{if } u < 0. \end{cases} \tag{2.9}$$

At this point, we must highlight some key differences compared to traditional LR for mean prediction. Firstly, while LR focuses on estimating the mean of Y given X, LQR provides a more comprehensive view by estimating specific points of the conditional distribution of Y. This is particularly useful in cases like EPF, where the distribution of prices Y is asymmetric and the price time series exhibits heteroscedasticity, that is time-varying variance. A visual comparison of LR and LQR is provided in Figure 2.4. Secondly, minimizing the MSE is an analytical problem that can be solved in closed form, while minimizing the MPL, that is piecewise linear, requires the use of an algorithmic solver based on Linear Programming (LP). Finally, we need to reflect on the residuals, which in this case are referred to as quantile residuals at level α:

$$e_i^{(\alpha)} = y_i - \beta_0^{(\alpha)} - \beta_1^{(\alpha)} x_i, \quad i = 1, \ldots, n. \tag{2.10}$$

Here the $e_i^{(\alpha)}$'s are not Gaussian but instead follow a distribution known as the Asymmetric Laplace Distribution (ALD) [33]. The probability density function (pdf) of the ALD for a residual e at quantile level α is given by

$$f(e; \mu, \sigma, \alpha) = \frac{\alpha(1-\alpha)}{\sigma} \exp\left(-\frac{\rho_\alpha(e-\mu)}{\sigma}\right), \tag{2.11}$$

where

- μ is the location parameter;
- $\sigma > 0$ is the scale parameter;
- $\alpha \in (0,1)$ is the asymmetry parameter;
- $\rho_\alpha(u)$ is the check function as above.

Therefore, if one wishes to conduct a visual analysis of the residuals in an intuitive way, as is the case with mean regression, a Gaussian transformation of the residuals must be performed, called normalization, that we describe briefly below.

2.2.2 Quantile Residual Normalization

A straightforward derivation shows that the ALD Cumulative Distribution Function (CDF) is

$$F(e) = F(e;\mu,\sigma,\alpha) = \int_{-\infty}^{e} f(u;\mu,\sigma,\alpha)du = \begin{cases} \alpha \exp\left\{(\alpha-1)\dfrac{e-\mu}{\sigma}\right\}, & e < \mu, \\ 1-(1-\alpha)\exp\left\{-\alpha\dfrac{e-\mu}{\sigma}\right\}, & e \geq \mu. \end{cases} \tag{2.12}$$

Indeed, when $e = \mu$ we have

$$F(\mu) = 1-(1-\alpha)\exp(0) = \alpha, \tag{2.13}$$

so that the αth quantile is correctly given by the location parameter μ.
We now know from theory that the CDF has this form, but we do not know the value of the parameters μ and σ, and therefore they must be estimated. A direct estimation approach is, for the location μ, to take

$$\hat{\mu} = \texttt{quantile}_{\alpha}\left(\{e_i\}_{i=1}^{n}\right), \tag{2.14}$$

where $\texttt{quantile}_{\alpha}(\cdot)$ denotes the numerical estimation of the sample αth quantile of the residuals $\{e_i\}_{i=1}^{n}$, computed by sorting. Also, to directly estimate the scale parameter σ, one can tale the average pinball loss,

$$\hat{\sigma} = \frac{1}{n}\sum_{i=1}^{n} \rho_{\alpha}\left(e_i - \hat{\mu}\right). \tag{2.15}$$

These direct estimates are simple but not robust; they are typically used as starting values for a numerical algorithm that minimizes the Negative Log-Likelihood (NLL) function, following the well-established Maximum Likelihood Estimation (MLE) [32] statistical principles. In our case,

$$-\ell(\mu,\sigma) \propto n\log\sigma + \frac{1}{\sigma}\sum_{i=1}^{n}|e_i-\mu|. \tag{2.16}$$

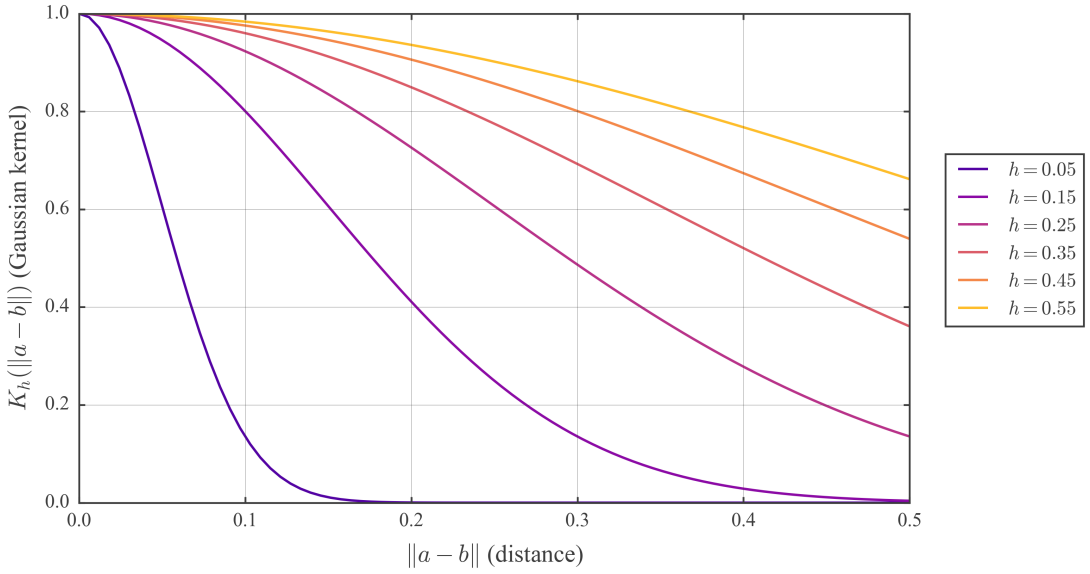

Fig. 2.5 **Visualization of the RBF Kernel with Varying Bandwidths.**
The figure illustrates the behavior of the Radial Basis Function (RBF) kernel, also known as Gaussian kernel, for different bandwidth (h) values, each represented by a different color. As the bandwidth decreases, the kernel fades more quickly.

Minimizing $-\ell(\mu,\sigma)$ with respect to μ and σ amounts to solving the MLEs problem. Once $\hat{\mu}$ and $\hat{\sigma}$ have been obtained, the centered residuals

$$\bar{e}_i = e_i - \hat{\mu}$$

are assumed to follow an ALD with parameters $(0,\hat{\sigma},\alpha)$. Their probability integral transform is given by

$$u_i = F\left(\bar{e};0,\hat{\sigma},\alpha\right).$$

Finally, applying the inverse standard normal CDF,

$$z_i = \Phi^{-1}(u_i),$$

yields transformed residuals z_i that are approximately standard normal, which is useful for model diagnostics. As we will see later, these allow the visual comparison of the residual distribution for different QR models in the way that is conventionally employed for mean regression models.

2.2.3 The Nadaraya-Watson Estimator

In this study, the selected approach for nonlinear point forecasting is the Nadaraya-Watson Estimator (NWE), also referred to as the KR estimator [11]. This is among the simplest models within the broader class of kernel methods (e.g., Support Vector Machines (SVMs), Gaussian Processes), and it was independently proposed by Nadaraya [34] and Watson [35] in 1964. The NWE is well suited to this study since it encapsulates the core concepts of nonlinear methods while remaining very simple. Nonlinear models such as SVMs rely on complex mathematical frameworks that often require significant parameter tuning, as well as a greter amount of available computational resources. As we will see in a moment, the NWE produces a flexible function approximation through A straightforward weighted averaging mechanism, which does not require high-dimensional optimization.

Like any other mean regression method, including LR, the NWE predicts the expected value of a dependent random variable Y by examining the relationship between the variable itself and an independent variable X. Also in this case, the objective is to estimate the conditional expectation of the response variable for a given input value x:

$$\hat{y}(x) = \mathbb{E}[Y|X = x]. \tag{2.17}$$

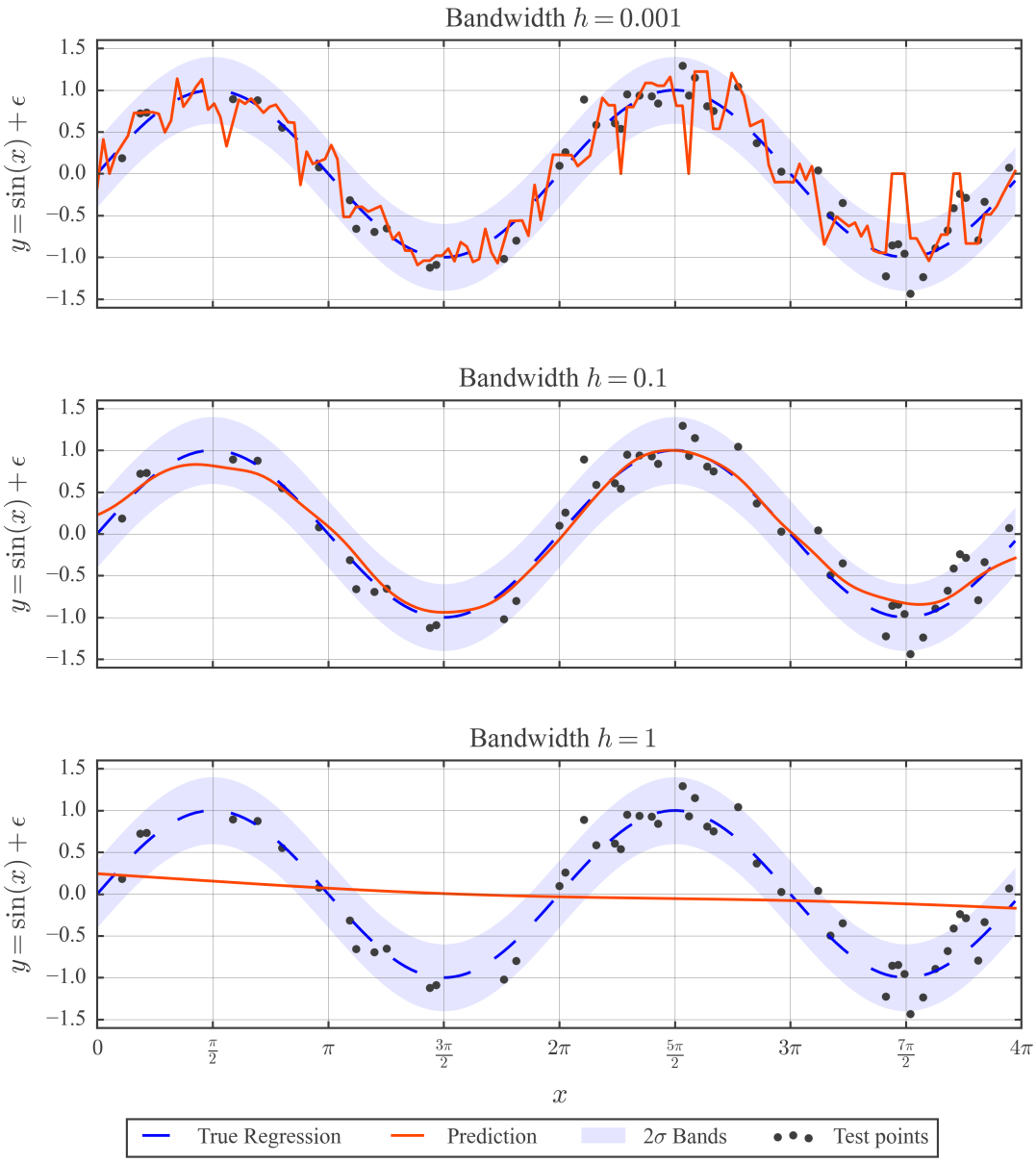

Fig. 2.6 **Behavior of The Nadaraya-Watson Estimator with Different Bandwidths**.
In the top panel, the Nadaraya-Watson Estimator (NWE) prediction (red) is shown for an excessively low h, leading to overfitting, where the estimator captures noise instead of the underlying function. In the central panel, the NWE effectively suppresses noise and reconstructs the true function, represented by the blue dashed line. Finally, in the lower panel, the bandwidth h is too high, causing the NWE to oversmooth and fail to capture the function's structure. In all three plots, the blue-shaded region represents the 2σ *bands, which, assuming Gaussian noise, contain approximately 95% of the data.*

Given training data $\{(x_i, y_i)\}_{i=1}^n$, the NWE provides a smooth estimate by averaging the observed values of y_i, weighted by kernel functions centered around each training point x_i:

$$\hat{y}(x) = \frac{\sum_{i=1}^n K_h(x, x_i) y_i}{\sum_{j=1}^n K_h(x, x_j)}, \tag{2.18}$$

where $K_h(x, x_i)$ is a kernel function with bandwidth parameter $h > 0$, which quantifies the similarity between x, the evaluation point, and x_i, the training data points. For this study, the Radial Basis Function (RBF), or Gaussian kernel, is adopted due to its widespread use in both engineering and statistics. For arbitrary inputs $a, b \in \mathbb{R}$, the RBF kernel is defined as:

$$K_h(a, b) = \exp\left[-\frac{(a-b)^2}{2h^2}\right]. \tag{2.19}$$

A visualization of the RBF kernel is available in Figure 2.5, where we show its behavior for different bandwidth values, each represented by a different color. For smaller bandwidths, the kernel tends to fade more quickly. To illustrate the rationale of the NWE, consider the system defined by the following relationship:

$$Y = \sin(X) + \varepsilon, \tag{2.20}$$

where $X \in [0, 2\pi]$ and $\varepsilon \sim N(0, \sigma^2)$ represents Gaussian noise with zero mean and fixed standard deviation σ. Figure 2.6 depicts an application of the NWE to $n = 100$ training samples generated from Equation (2.20), with $\sigma = 0.2$ and different bandwidth values. In the top panel, we see in red the NWE prediction for an excessively low h value: the estimator overfits the training data and captures noise instead of the underlying pattern. In the central panel, the NWE effectively suppresses noise, reconstructing the underlying function, as denoted by the blue dashed line. Finally, in the lower panel, the bandwidth h is too high, and the NWE fails to capture the structure of the function. Hence, the bandwidth parameter h plays a critical role in determining the model's performance and fit to the data. Bandwidth selection, extensively studied in the literature [36], remains a central challenge for the NWE.

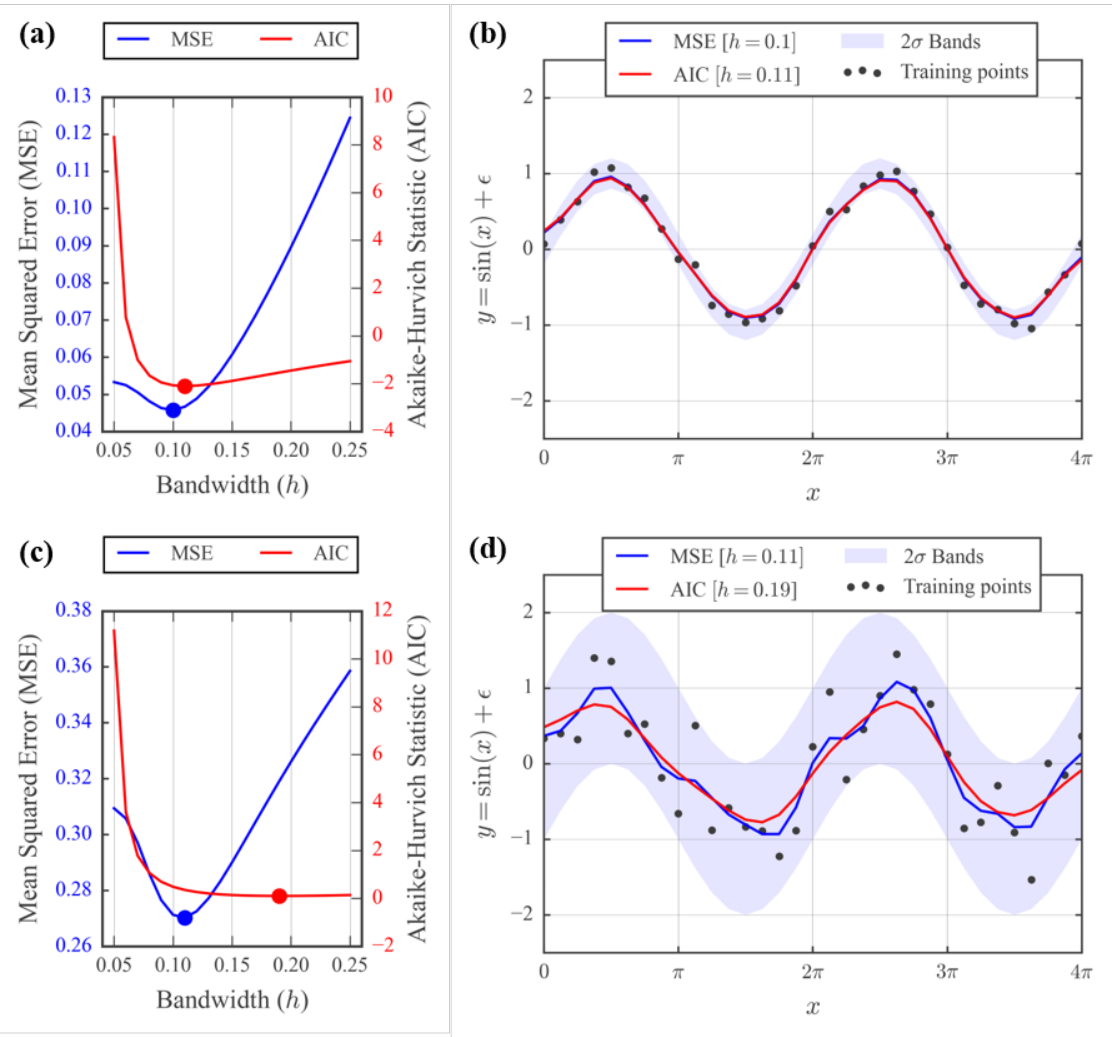

Fig. 2.7 **Comparison of Bandwidth Selection Using MSE and AIC Criteria.** .
Comparison of bandwidths selected using the Mean Squared Error (MSE, blue) and Akaike Information Criterion (AIC, red) criteria for the Nadaraya-Watson Estimator, as applied to the sinusoidal regression problem defined by Equation (2.20).
In the low-noise regime of subplots ***(a)*** *and* ***(b)****, with standard deviation* $\sigma = 0.1$*, both criteria yield similar bandwidths. However, in the high-noise regime (subplots* ***(c)*** *and* ***(d)****,* $\sigma = 0.5$ *), the MSE selects a smaller bandwidth, leading to overfitting, while the AIC chooses a larger bandwidth, effectively smoothing the noise and providing a better approximation of the true regression function.*

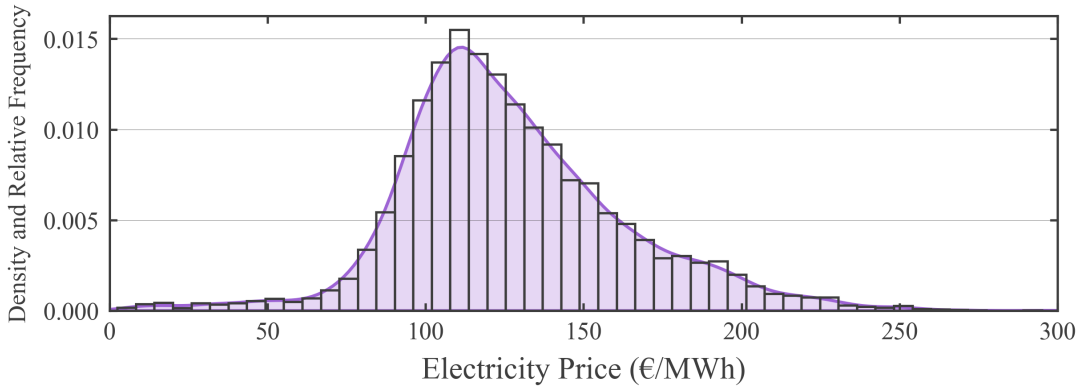

Fig. 2.8 **Kernel Density Estimation for Electricity Prices (2023).**
The figure presents an application of the Kernel Density Estimation (KDE) to Italian hourly electricity prices for year 2023. On the x-axis, prices are expressed in euro. The y-axis measures both the estimated probability density function $\hat{f}(x)$, and the relative frequency of the observations, shown by the histogram.

While minimizing the cross-validated MSE is a conventional method, it may be unreliable in scenarios with limited and noisy data. To address this, the Akaike-Hurvich Information Criterion (AIC) [37] is utilized, formulated as follows:

$$\begin{aligned} \min_h \quad & \mathrm{AIC}(h) = \log(\hat{S}^2) + \frac{1+\frac{T(H)}{n}}{1-\frac{T(H)+2}{n}}, \\ \text{where} \quad & \hat{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_h(x_i))^2 \quad \text{(training MSE)}, \\ & H = \big(K_h(x_i,x_j)\big)_{i,j=1}^{n} \quad (n \times n \text{ kernel matrix}), \end{aligned} \tag{2.21}$$

and $T(H)$ denotes the trace of matrix H. Figure 2.7 illustrates a comparison of bandwidths selected using the MSE (blue) and AIC (red) criteria for the regression problem defined by Equation (2.20). In low-noise regimes (e.g., $\sigma = 0.1$), the two criteria yield similar bandwidths. However, in high-noise regimes (e.g., $\sigma = 0.5$), the MSE selects a lower bandwidth, leading to overfitting, while the AIC prefers a larger bandwidth, effectively smoothing the noise and better approximating the true regression.

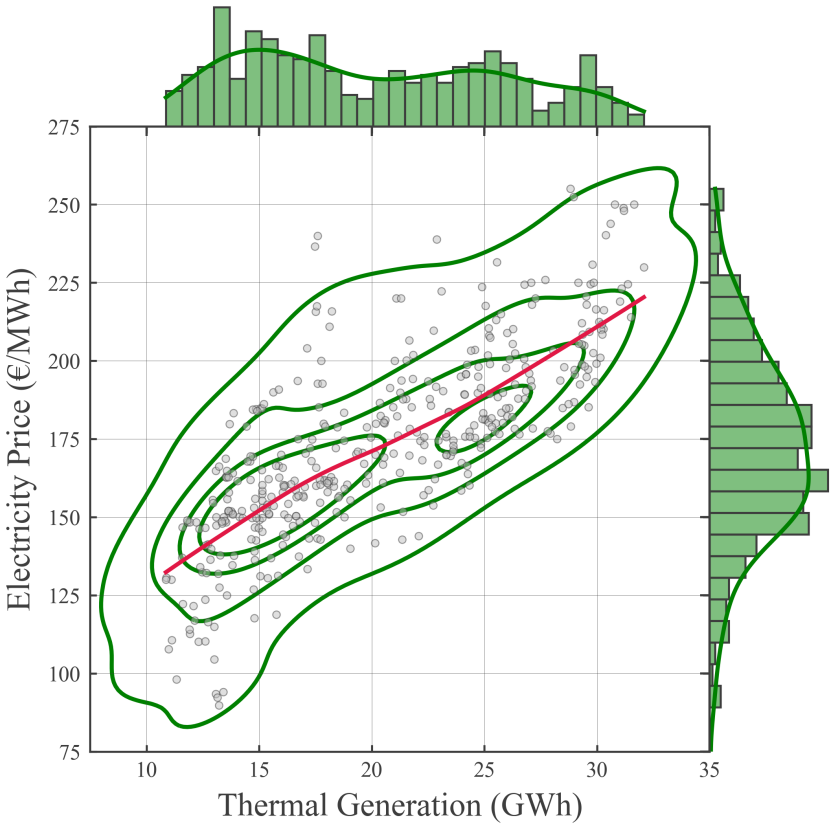

Fig. 2.9 **Bivariate Kernel Density Estimation of Electricity Prices and Thermal Generation**
The figure presents a bivariate Kernel Density Estimation (KDE) of Italian electricity prices (€/MWh) and national thermal generation (GWh), a strong predictor of price. The grey circles represent observed values for January 2023. The right and top panels show the corresponding univariate KDEs and histograms, as in Figure 2.8. In the central plot, the green contour lines depict the bivariate KDE, while the red curve represents the Nadaraya-Watson Estimator regression.

Kernel Density Estimation

In some of the introductory visual analyses (c.f.r. Figure 2.3) we employed Kernel Density Estimation (KDE). It is a non-parametric technique used to estimate the probability density function (pdf) of a continuous random variable, tightly related to the NWE as they both rely on the well-established Parzen window theory [38].
KDE can be viewed as an extension of histogram-based estimation which provides a smoother and continuous representation of the distribution. Given a real-valued continu-

ous random variable X with an unknown distribution, the objective is to estimate:

$$f(x) \geq 0 \quad \forall x \in \mathbb{R} \quad \text{s.t.} \quad \mathbb{P}(X \leq s) = \int_{-\infty}^{s} f(s)ds \quad \forall s \in \mathbb{R}. \tag{2.22}$$

Given independent and identically distributed samples $\{x_i\}_{i=1}^{n}$, KDE estimates $f(x)$ by weighting each observation using a kernel function $K_h(x, x_i)$, analogous to that used in the NWE:

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} K_h(x, x_i), \tag{2.23}$$

where as usual a bandwidth parameter $h > 0$ controls the smoothing scale and determines the influence of each point x_i on the final estimate.
In the case of KDE, the distributions are not conditioned on any predictor variable, making the estimation of the bandwidth a more straightforward task compared to KR with the NWE. Since the goal is to approximate the unconditional distribution of the data, standard heuristic rules can be applied to select an appropriate bandwidth with satisfactory results. In this study, we adopt Scott's Rule [39]:

$$\hat{h} = n^{-\frac{1}{p+4}}, \tag{2.24}$$

where n is the number of samples and p is the number of variables. Let us introduce two examples below .
Figure 2.8 illustrates the application of KDE to the Italian hourly electricity prices for the year 2023. The x-axis represents electricity prices in euros, while the y-axis shows both the estimated probability density function $\hat{f}(x)$ nd the relative frequency of observed prices, visualized through the histogram. Notably, the bandwidth h of the kernel function plays the same role as the number of bins in the histogram. Furthermore, Figure 2.9 presents a bivariate KDE of prices and national thermal generation (GWh), which is a strong predictor of price. The black dots correspond to observed values for January 2023. The right and top panels display the corresponding univariate KDEs and histograms, following the same approach as in the previous Figure 2.8. In the central plot, the blue contour lines represent the estimated bivariate density, while the orange curve shows the NWE regression.

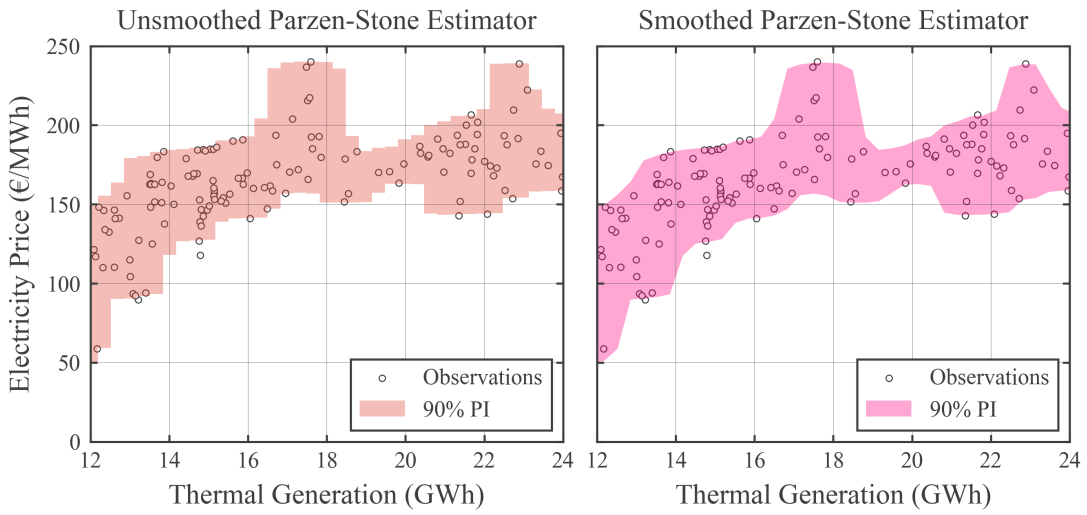

Fig. 2.10 **Comparison of Unsmoothed vs. Smoothed Parzen-Stone Estimator**.
The figure depicts 90% Prediction Intervals (PI) for electricity price (€/MWh) as a function of thermal generation (GWh). On the left, the PI (in orange) is obtained using the basic Parzen-Stone estimator, which does not smooth the response variable. On the right, the PI (in pink) represents the smoothed version, where smoothing is applied using a Gaussian cumulative distribution function with a variance of 1. The data cover the third week of January 2023.

2.2.4 The Parzen-Stone Estimator

While the NWE estimates the conditional expectation, and thus provides a point forecast, such forecasts are inherently limited in representing the uncertainty of continuous-valued variables. This limitation is particularly relevant in forecasting tasks such as EPF, where high volatility renders point estimates less informative. Estimating the full conditional probability distribution of the target variable addresses this limitation.
The conditional Cumulative Distribution Function (CDF) for a target variable Y given a realization x of X is defined as:

$$F(y|x) = \mathbb{P}(Y \leq y|X = x), \quad y \in \mathbb{R}. \tag{2.25}$$

Any conditional CDF can be expressed as a conditional expectation [40]:

$$F(y|x) = \mathbb{E}[I(Y \leq y)|X = x], \tag{2.26}$$

where $I(A)$ is the indicator function, which equals 1 if A is true, and 0 otherwise. Consequently, the NWE can estimate the CDF as:

$$\tilde{F}(y|x) = \frac{\sum_{i=1}^{n} K_h(x,x_i) I[y_i \leq y]}{\sum_{j=1}^{n} K_h(x,x_j)}. \tag{2.27}$$

As discussed in [36], this estimator, here referred to as the Parzen-Stone Estimator (PSE) does not smooth the dependent variable Y. To address this, the Smoothed PSE extends the above by incorporating a continuous cumulative distribution $G_{h_0}(y)$:

$$\hat{F}(y|x) = \frac{\sum_{i=1}^{n} K_h(x,x_i) G_{h_0}(y-y_i)}{\sum_{j=1}^{n} K_h(x,x_j)}, \tag{2.28}$$

where $h_0 > 0$ is the bandwidth for Y. In this work, G is taken as the standard Gaussian CDF:

$$G_{h_0}(y) = \frac{1}{\sqrt{2\pi}h_0} \int_{-\infty}^{y} \exp -\frac{u^2}{2h_0^2} du. \tag{2.29}$$

The PSE belongs to a class of nonparametric regression methods described in [41], based on Parzen window theory. A visual comparison between the unsmoothed and smoothed PSEs is provided Figure 2.10. The 90% PIs for electricity prices, in €/MWh, are estimated as a function of thermal generation, expressed in GWh, based on data from January 2023. The left panel presents the PI (shown in orange) derived using the standard PSE, which does not incorporate smoothing for the response variable. In contrast, the right panel displays the smoothed PI (in pink), where the smoothing is Gaussian with a bandwidth of value $h_0 = 1$.

2.3 Univariate Analyses

In this section, we describe the variables considered for this study, outlining their sources and justifying their inclusion in the final regression models. To anticipate, the final predictor set is comprised of

- the previous day's hourly electricity prices, in €/MWh;
- the previous day's average NG prices (Gas), in €/MWh;
- the forecasts for solar generation (Solar), in GWh;
- the forecasts for thermal generation (Thermal), in GWh.

We also conduct a univariate regression analysis for both included and excluded variables, providing initial insights into their potential as regressors. Here, price is the endogenous variable, while all other variables that may be included as regressors are treated as exogenous.
As was discussed in the previous Section, selecting appropriate predictor variables is an open problem, with no universal solutions that are independent of the specific market and time frame. Our variable selection process began with a set of theoretically grounded choices, which were refined through empirical data analysis. Since market prices are determined by the interaction of demand and supply, our initial focus was on variables that could serve as proxies for these factors. Demand was represented by national electrical load, a natural choice. On the supply side, we considered thermal and photovoltaic generation, given their relevance to the current Italian energy landscape. Additionally, the price of NG was included, as it remains the dominant fossil fuel for electricity generation in Italy. Conversely, national electrical load was excluded from the multivariate regression model, as it offered no performance improvements in either point or probabilistic forecasting.
Given the other goal of analysis—to evaluate the performance of nonlinear models in EPF—we include a comparison of results achievable with linear methods. Interestingly, the final results show no significant differences in performance between linear and nonlinear models, although the latter offer qualitatively better fits. The analyses presented here cover data from January to August 2023. The data sources are reported below, separately for each variable.

2.3.1 Day-Ahead Electricity Price

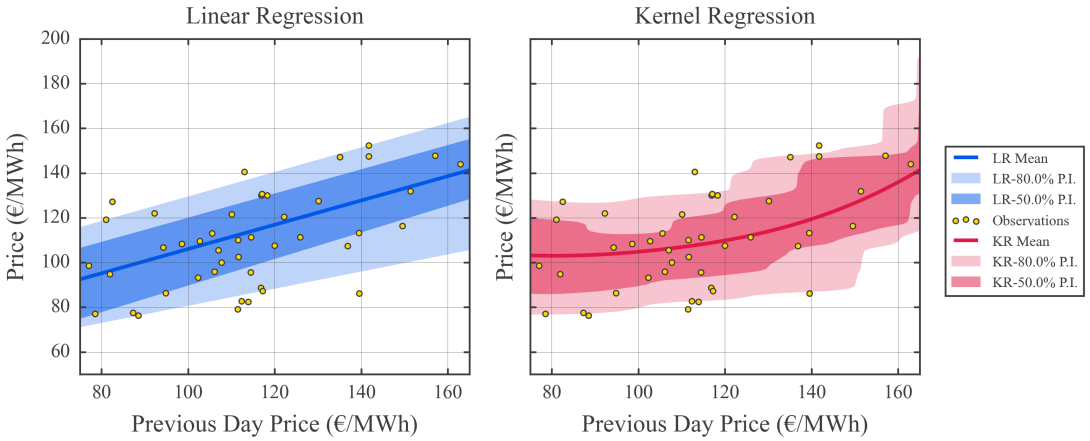

Fig. 2.11 **Univariate Analysis: Forecasting Wednesdays at 12:00 PM Using Previous Day Prices**. *The chart compares two regression models—Linear Regression (LR) on the left, in blue and Kernel Regression (KR) on the right, in red—for predicting Wednesday prices at 12:00 PM, covering data from 2023. Both methods present mean regression results and Prediction Intervals (PIs) at 50% (spanning the interquartile range) and 80% (extending from the 10th to the 90th percentiles). The similarity between the outputs of these two models stems from the clear dependency between consecutive daily prices. In most forecasting applications, LR stands out as a recommended baseline model due to its simplicity and effectiveness.*

The target variable in this study is the equilibrium price in the Mercato del Giorno Prima (MPG), the Italian day-ahead market, with its past values included as regressors. In Italy, this price is known as the Prezzo Unico Nazionale (PUN), the unique national price[2]. Let d represent the day of the week, numbered from 0 (Monday) to 6 (Sunday), with $(d-1)$ denoting the previous day. The goal is to predict the price at hour h on day d using the price at the same hour on the previous day, $(d-1)$. For illustrative purposes, we focus on predicting the price for Wednesdays ($d=3$) at noon ($h=12$).

As shown in Fig. 2.11, the results of LR are depicted in blue on the left, while KR results are shown in red on the right. For both models, the mean regression and central PIs at 50% and 80% are reported. The 50% PI has bounds at the 25th and 75th percentiles, while the 80% PI spans the 10th and 90th percentiles. Despite the differing methodologies of KR and LR, their univariate regression results are highly similar, reflecting a clear

[2]Starting from January 1, 2025, the PUN as a market price has been abolished in favor of zonal prices divided into seven geographical macro-areas. However, it still exists as an index and is used to parameterize supply contracts; it is calculated as the weighted average of the zonal prices.

correlation between the prices on days d and $(d-1)$. In the linear case, the model serves as an effective baseline for evaluating more complex models [28].
Although [17] emphasized the potential benefits of including prices from two days $(d-2)$ and one week $(d-7)$ prior to the prediction day, we found no empirical evidence to support their inclusion. These lags were expected to better capture daily and weekly seasonality, but their addition did not improve performance in our models.
Price data were sourced from the website of Gestore dei Mercati Energetici (GME), a state-owned corporation managing Italy's commodity exchanges, including the MGP and day-ahead NG markets, known as MGP-GAS [42]. Data are freely available and regularly updated.

2.3.2 Natural Gas Day-Ahead Price

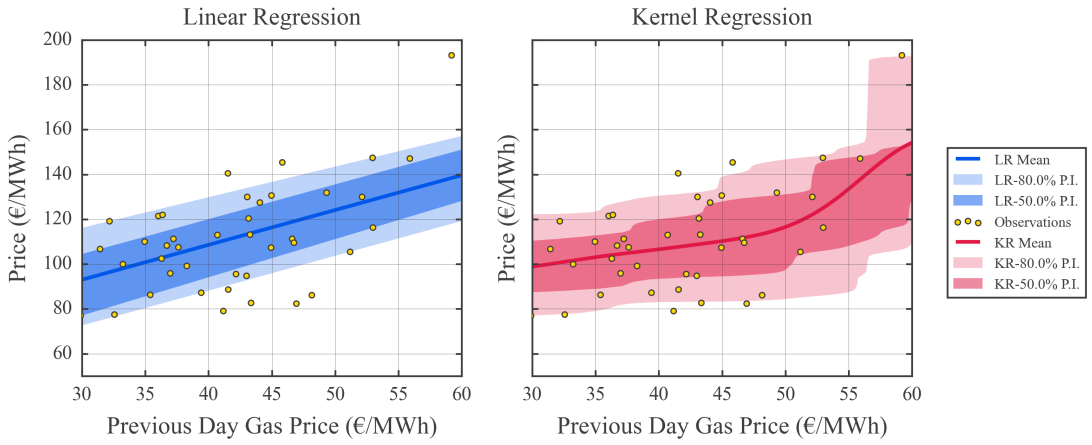

Fig. 2.12 **Univariate Analysis: Forecasting Wednesdays at 12:00 PM Using Gas Prices**.
The chart compares Kernel Regression (KR, in red) and Linear Regression (LR, in blue) models for forecasting electricity prices using the average natural gas price from the previous day. The analysis spans 2023 (Source: MGP-GAS). Both models present mean predictions and Prediction Intervals (PIs) at 50% (interquartile range) and 80% (extending from the 10th to the 90th percentiles). While LR assumes a simple linear relationship, KR captures a slight non-linear trend, especially for natural gas prices exceeding €50, where electricity price variations accelerate. Gas is a significant predictor, considering that Italy's electricity generation remains heavily reliant on thermoelectric plants powered by natural gas, despite the growth in photovoltaic and wind energy systems.

Thermoelectric generation using NG remains the backbone of Italy's electricity production, accounting for roughly 40% of total generation. Consequently, electricity prices

are strongly correlated with NG prices, as illustrated in Fig. 2.12. KR (red, right panel) captures a nonlinear trend, showing an accelerated increase in electricity prices when gas prices exceed €50. KR also predicts wider PIs compared to LR.
Here, we use the spot price of NG from the day-ahead MGP-GAS market on the day prior to prediction. Unlike electricity prices, gas prices are reported as daily averages, reflecting the structure of the MGP-GAS market, since it does not feature an hourly resolution. The growing liquidity of this market ensures that MGP-GAS prices partially reflect costs associated with gas-based electricity generation. These data were also obtained from the GME website [42]. It is worth noting that Italy's reliance on natural gas is heightened by the absence of active nuclear power plants; in countries with nuclear energy, a weaker correlation would be observed.

2.3.3 Thermal Generation Forecast

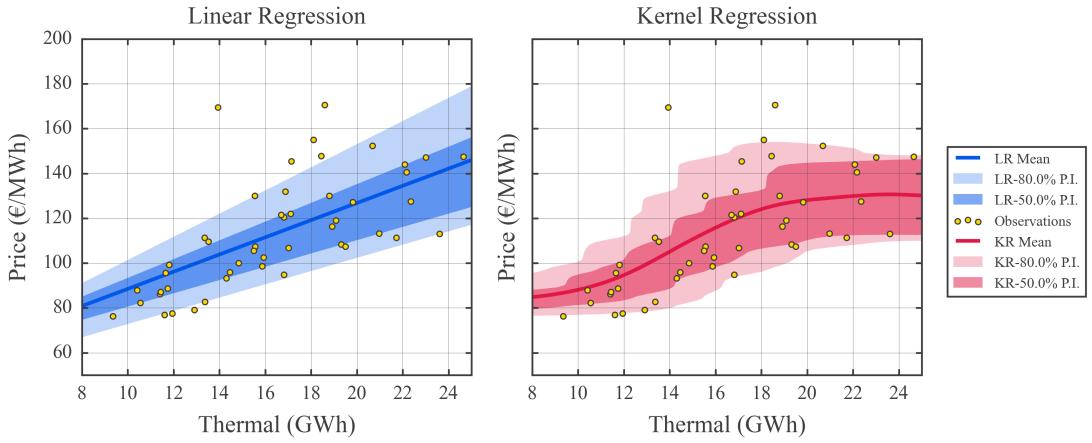

Fig. 2.13 **Univariate Analysis: Forecasting Wednesdays at 12:00 PM Using Thermal Generation**. *This figure highlights the significant impact of thermoelectric generation, which accounts for 40% of Italy's electricity production, on electricity price fluctuations. Notably, price volatility intensifies during periods of high thermal generation, reflecting reduced contributions from alternative sources and amplifying the sensitivity of prices to even minor changes in thermal supply. Both Kernel Regression (KR, shown in red) and Linear Regression (LR, shown in blue) capture this trend, with relatively similar results. As with previous analyses, the data span 2023.*

Thermal generation data were included for similar reasons as NG prices. However, actual generation levels may vary due to factors such as plant-specific operating limits, maintenance schedules, and unexpected outages. As shown in Fig. 2.13, higher levels

of thermal generation are associated with increased electricity price volatility, likely reflecting reduced availability from alternative sources. KR (red) and LR (blue) produce similar results for this variable.
Thermal generation data, available on an hourly basis, were sourced from Terna, the state-owned operator of Italy's transmission grid [43]. In this study, we assume perfect forecasts for thermal generation, as short-term predictions for this variable are typically accurate.

2.3.4 Photovoltaic Generation Forecast

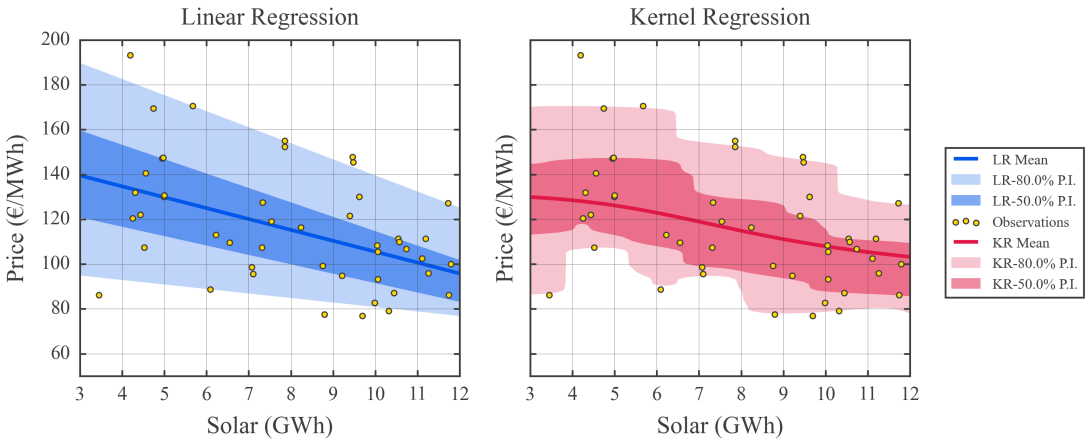

Fig. 2.14 **Univariate Analysis: Forecasting Wednesdays at 12:00 PM Using Photovoltaic Generation**. *With photovoltaic systems accounting for 12% of Italy's total installed capacity [5], the trend is the reverse of that observed with thermal generation. Higher photovoltaic supply correlates with lower prices and reduced volatility. This pattern is reflected consistently in both Kernel Regression (KR, shown in red) and Linear Regression (LR, shown in blue), with minimal differences between the two models.*

Photovoltaic generation reflects electricity produced by solar installations, a key renewable resource in the Mediterranean region. Investments in photovoltaic capacity have grown significantly, driven by the region's favorable conditions for solar energy.
As shown in Fig. 2.14, high levels of photovoltaic generation are associated with lower electricity prices and reduced volatility. Both KR (red) and LR (blue) capture this relationship effectively, with minimal differences between the two models.
Photovoltaic generation forecasts, assumed to be perfect as with thermal generation, were obtained from Terna [43]. These data are also updated daily and have hourly resolution.

2.3.5 Excluded Variables

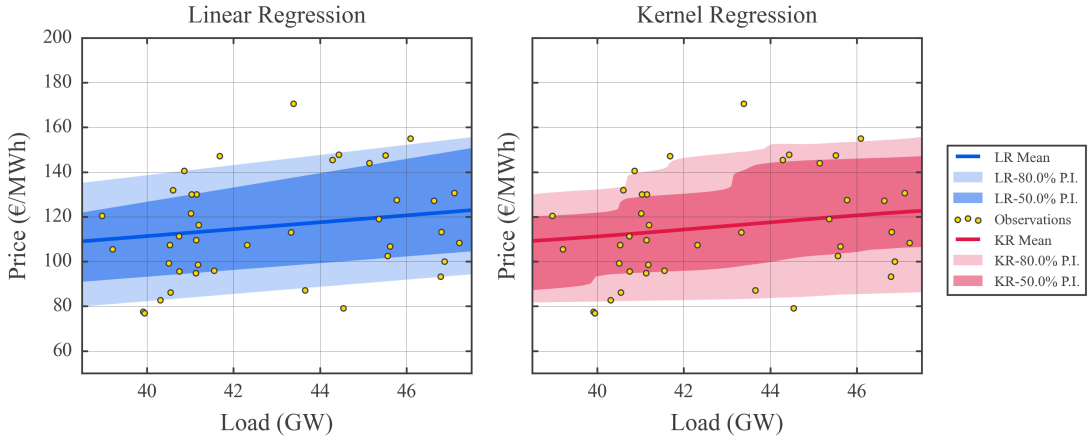

Fig. 2.15 **Univariate Analysis: Forecasting Wednesdays at 12:00 PM Using National Electric Load**. *This analysis examines the relationship between electricity prices and national electric load, a commonly used predictor alongside past price values. Surprisingly, the correlation observed between the load and electricity prices is relatively weak. This is evident in the results of both Linear Regression (LR, shown in blue on the left) and Kernel Regression (KR, shown in red on the right).*

Certain variables were excluded from the final regression models despite they could be relevant in principle. National electrical load, for example, is a widely used predictor in EPF [16]. However, as shown in Fig. 2.15, load exhibited no strong correlation with electricity prices in our dataset. It is worth noting that while the load has historically been a significant predictor, this relationship appears weaker in the current Italian market. Over the past five years, electricity prices have undergone drastic changes, whereas the national electric load has remained relatively stable, except during exceptional periods such as the COVID-19 lockdowns [5].

Similarly, net foreign exchange data, which capture cross-border energy imports and exports, were excluded due to a lack of significant correlation with electricity prices in our dataset (Fig. 2.17). The diminished impact of this variable may reflect broader disruptions in European energy markets caused by the Russo-Ukrainian conflict.

Lastly, hydroelectric generation (Hydro) showed limited influence on price dynamics, as shown in Fig. 2.16. Both hydroelectric and cross-border exchange data were sourced from Terna [43].

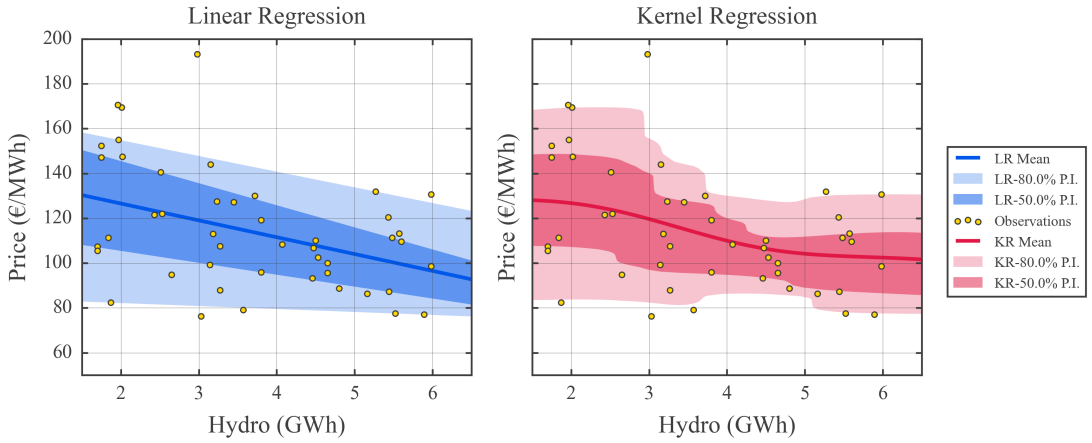

Fig. 2.16 **Univariate Analysis: Forecasting Wednesdays at 12:00 PM Using Hydro Generation**. *Hydroelectric generation displays trends similar to solar generation, as higher output leads to reduced prices and lower volatility. This is illustrated in both Linear Regression (LR, shown in blue) and Kernel Regression (KR, shown in red), with KR offering a more nuanced representation. Despite this, hydro generation was excluded from the final model due to its limited predictive contribution. It is worth noting that drought conditions in recent years, particularly in the Mediterranean, have diminished its relevance, unlike in Nordic electricity markets, where Hydro remains a key factor.*

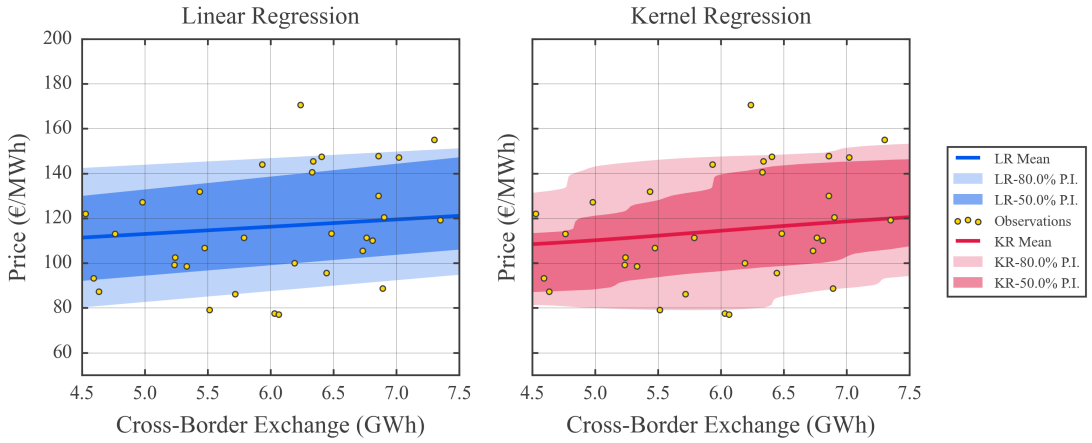

Fig. 2.17 **Univariate Analysis: Forecasting Wednesdays at 12:00 PM Using Cross-Border Exchange**. *The variable represents cross-border electricity flows, which connect Italy to countries such as France, Switzerland, Austria, Slovenia, Montenegro, and Greece via 25 interconnection lines [43]. Italy often relies on imports, particularly from countries like France and Switzerland, where nuclear energy is cheap and abundant, especially when electricity availability from renewables is low. Conversely, during sunny days, Italy often exports cheap photovoltaic electricity. However, in this dataset, exchanges exhibit no clear correlation with electricity prices, limiting its utility for predictive models*

.

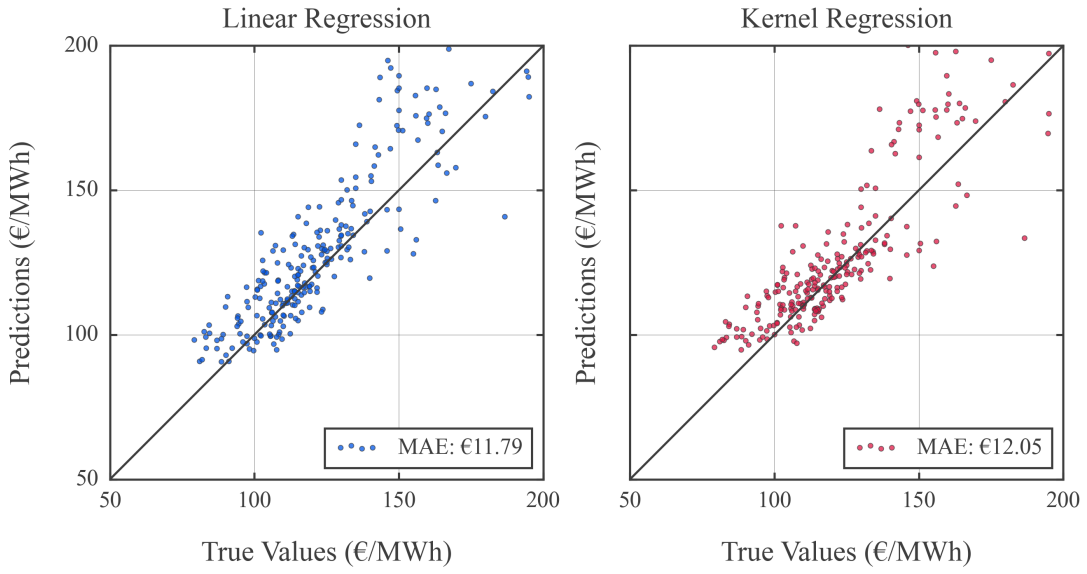

Fig. 2.18 **Multivariate Regression Results for September 2023. Test Period: Peakload**.
The performance during peakload hours is quite similar for both models. However, their accuracy decreases at higher price levels, which is expected given the increased volatility.

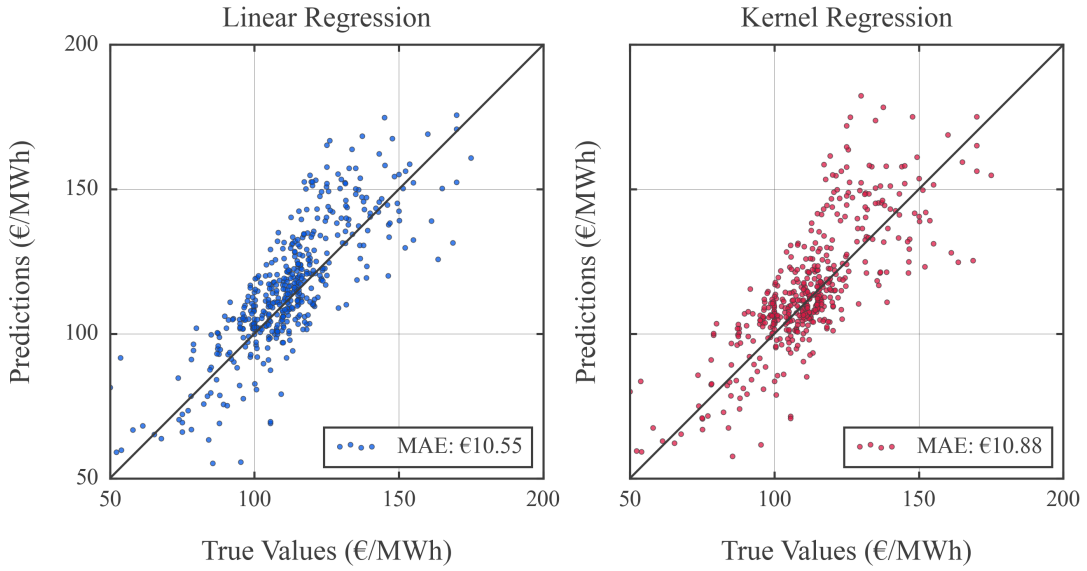

Fig. 2.19 **Multivariate Regression Results for September 2023. Test Period: Baseload.**
For baseload, the situation is very similar to peakload, with slightly better average performance.

.

2.4 Multivariate Analysis

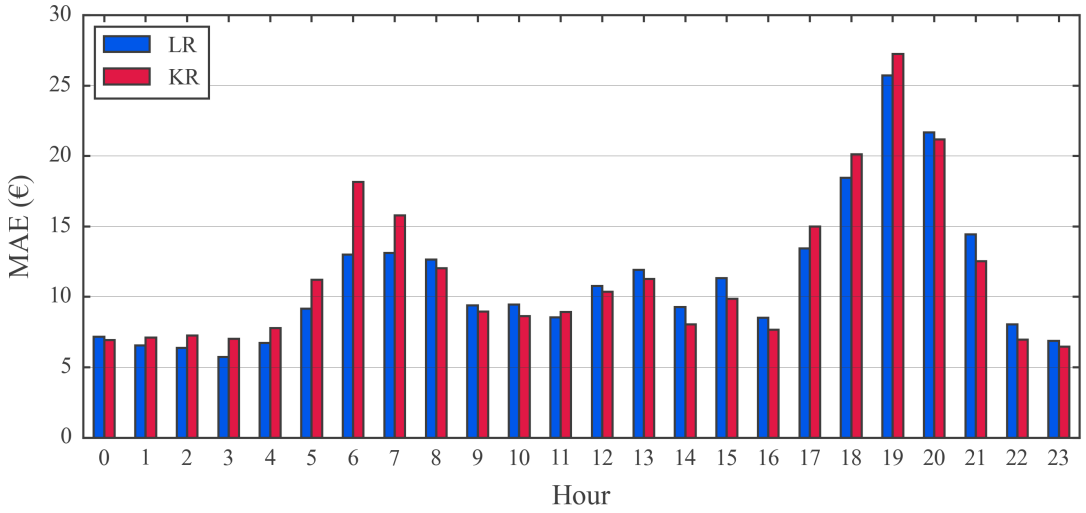

Fig. 2.20 **Point Forecasting Results. Error by Hour of Day.**
The figure illustrates the predictive performance for mean regression, subdivided by hours of the day. The predictive accuracy of both models degrades in late evening, when demand oscillations are more frequent. Despite Linear Regression (LR, in blue) and Kernel Regression (KR, in Red) are different in nature, the achieve very similar results.

In electricity markets, it is common to divide the hourly periods of the week into two groups: peak and off-peak, based on an average load that is generally higher or lower. In Italy, the peak period includes all hours from 8:00 AM to 8:00 PM, Monday to Friday. The remaining hours and the entire weekend constitute the off-peak period, also known as the baseload period. Since many forward contracts are based on this division, and the day-ahead market serves to make load adjustments compared to what was obtained in advance through forwards, it is interesting to follow this division to assess the performance of models. Particularly in industrial EPF, the predictive performance on the peak is more interesting because higher prices generally form due to higher demand.

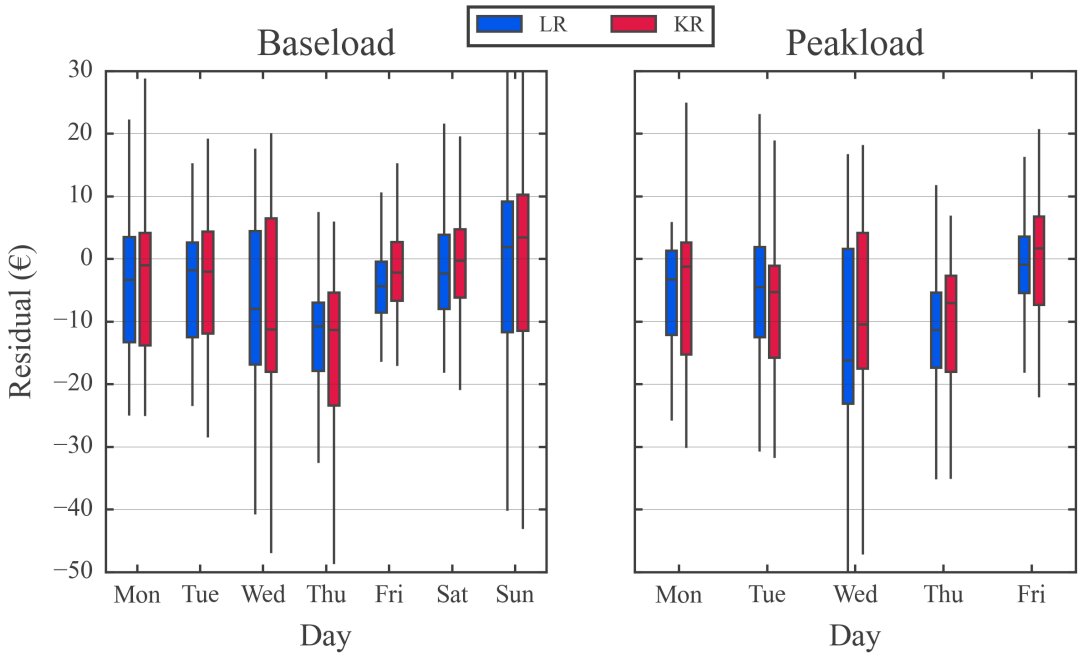

Fig. 2.21 **Point Forecasting Results. Residual Analysis.**
The figure presents a residual analysis for the test period considered in the multivariate model analysis. The plot focuses on point forecasting. The left subplot corresponds to baseload, while the right subplot represents peakload. In each subplot, the y-axis shows the residuals, or forecasting errors, expressed in euros, while the x-axis represents data grouping by day of the week. Linear Regression (LR) is shown in blue, and Kernel Regression (KR) in red. In this case as well, the results are very similar.

Following this distinction, we begin by presenting the results for point forecasting. As before, values related to KR in the figures will be depicted in red, while those for LR models will be in blue. The predictive accuracy of both linear and non-linear models is presented for peakload in Fig. 2.18 and for baseload in Fig. 2.19, covering all days from Monday to Sunday ($d = 0,\ldots,6$) and all hours of the day ($h = 0,\ldots,23$). As an initial general observation, the models demonstrate very similar accuracy in both cases, around €11 for baseload and approximately €12 during peakload periods. This result can be considered satisfactory, as it aligns with the accuracy of forecasting models reported in literature and tested on Italian market data from previous periods (pre-2020). For greater detail, Figure 2.20, shows the predictive performance subdivided by hours of the day. Both models show a decrease in predictive accuracy during the late evening hours, when demand fluctuations are more frequent. Despite the inherent differences between LR (in blue) and KR (in red), both models achieve very similar results. Furthermore, Figure 2.21

presents a residual analysis. The left panel corresponds to baseload, while the right panel represents peakload. In both cases, the y-axis displays residuals (forecasting errors) in euros, while the x-axis groups data by day of the week. Overall, the two models exhibit very similar performance and no evident pattern in residuals is present.

Moving on to the more recent topic of probabilistic forecasts, we can start by looking at the MPL for each quantile level considered: $\alpha = 0.05, 0.25, 0.75, 0.95$. The results are shown in Fig. 2.22, segmented for baseload and peakload periods. Let us begin by noting that, in both cases, the quantiles that are more challenging to predict are those corresponding to the central levels (25th and 75th percentiles), while the highest quantile (95th percentile) is generally more difficult to predict than the lowest quantile (5th percentile). This latter observation was expected, as the distribution of electricity prices is asymmetric, with higher price levels exhibiting greater volatility. In this case as well, the performance of KR and LR is nearly equivalent. This consideration is further supported by the supplementary analyses presented in Figures **??**.

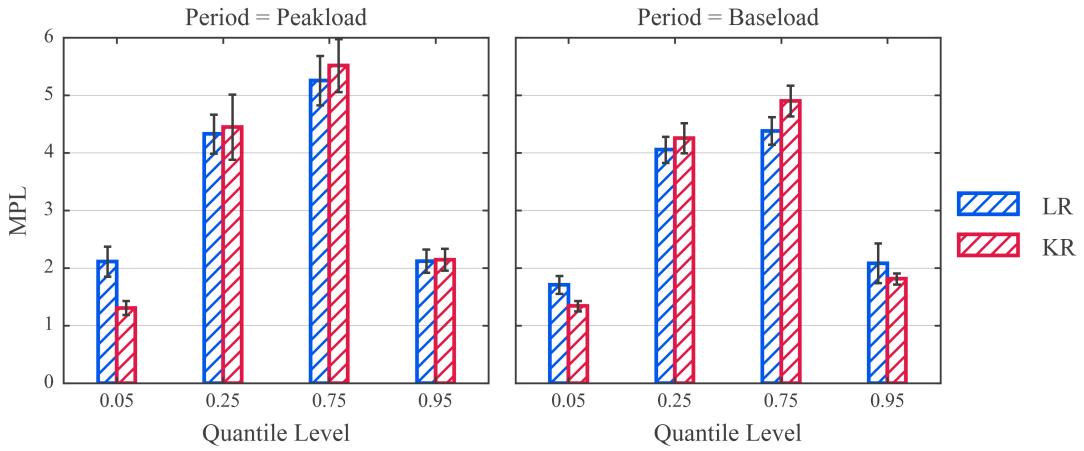

Fig. 2.22 **Probabilistic Forecasting. MPL across Quantiles for Baseload and Peakload Test Periods.** *The barplots illustrate the Mean Pinball Loss (MPL), with Linear Regression (LR) results shown in red and Kinear Regression (KR) in blue. Black lines above the bars represent estimation errors. The most challenging quantiles to predict are consistently found at the central levels (25th and 75th percentiles), while the 95th percentile tends to be more difficult to estimate than the 5th percentile. This outcome is expected due to the asymmetric distribution of electricity prices, where higher values exhibit greater volatility. In this case as well, linear and nonlinear models demonstrate very similar performance.*

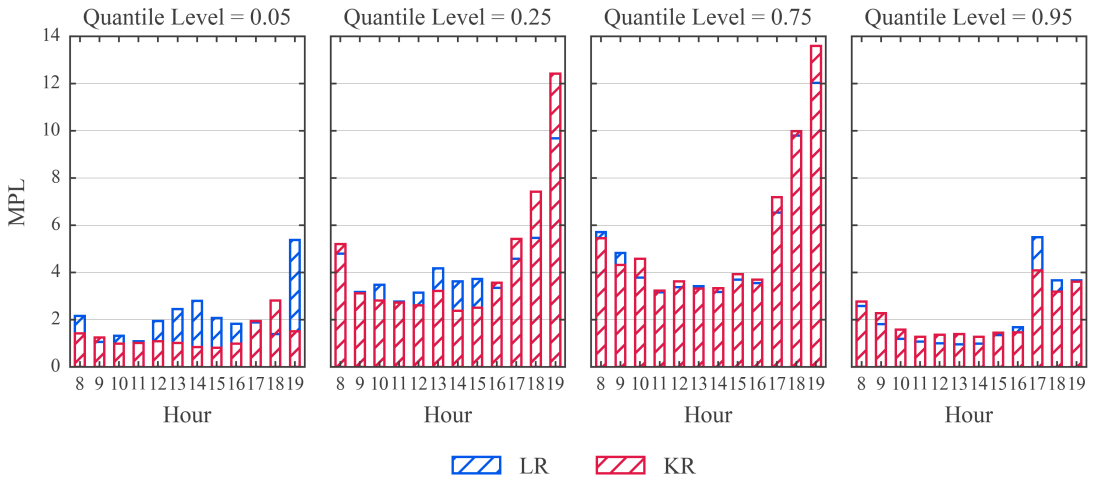

Fig. 2.23 **Probabilistic Forecasting Results. MPL across Quantiles and Peakload Hours.**
The figure illustrates the predictive performance during peakload hours in the test period of September 2023. In this case as well, both models exhibit reduced predictive accuracy during late evening hours across all quantile levels.

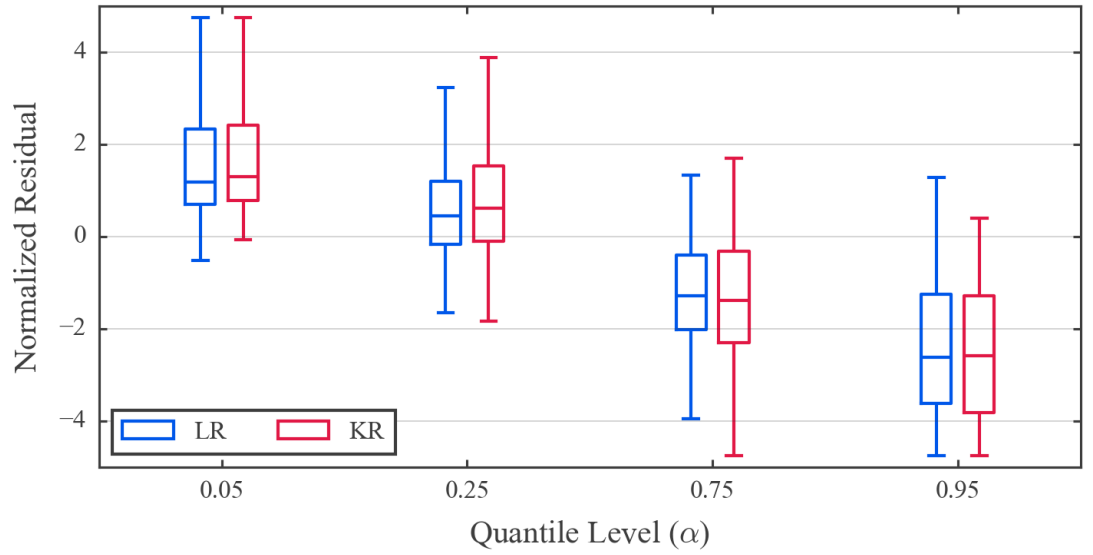

Fig. 2.24 **Probabilistic Forecasting Results. Residual Analysis.**
The figure presents a quantile residuals analysis for the multivariate models. In this case, it is important to note that the residuals are normalized and therefore follow an approximately Gaussian distribution with a mean of 0 and a variance of 1. The data are grouped by quantile level. The differences between Linear Regression (LR, in blue) and Kernel Regression (KR, in red) are even smaller if compared to the case of mean regression. Notice the overestimating tendency present for both LR and KR at the upper quantile.

Chapter 3

Model Averaging in Point and Probabilistic Electricity Price Forecasting

While the main variables driving electricity price formation are well known, effectively combining their predictive capabilities into a single forecasting model presents significant challenges. The predictive performance of factors such as electrical load or hourly generation can vary widely depending on market conditions or other contextual influences that are difficult to predict or quantify in advance. This variability means that the effectiveness of a given predictor often depends on situational factors that may not be evident beforehand.

These issues are further amplified in the context of probabilistic forecasting. For instance, when forecasting the quantiles of future price distributions, different predictors tend to exhibit varying levels of accuracy across different quantile levels. To address these complexities, this Chapter explores the potential of Model Averaging (MA) as an innovative solution. Instead of relying solely on the best-performing individual model, MA involves considering a diverse set of models and aggregating their forecasts using various techniques.

We conduct an in-depth analysis of the Electricity Price Forecasting (EPF) problem, starting with mean predictions and progressing to quantile predictions. The analysis begins with simple, individual forecasting models, encompassing both linear and non-linear

approaches. We evaluate the predictive strength of individual variables and combinations of variables. Building on this, we then explore a variety of model aggregation strategies, from straightforward approaches to cutting-edge methods.
The results of our study are based on recent data from the Italian day-ahead electricity market and include a comparison with widely used ensemble forecasting techniques. In both point forecasting and probabilistic forecasting, MA guarantees substantial improvements.

3.1 Combining Electricity Price Forecasts

When one faces a complex decision about the future, it is essential to her to ensure that this choice is as rational and well-informed as possible. For instance, consider the scenario of a student nearing the end of high school, faced with the decision of which university to attend in the coming years. The information that the individual already possesses plays a critical role, as personal inclinations are often the most decisive factors. However, it can be equally valuable to seek input from friends and acquaintances. In fact, gathering additional perspectives can help guide a more informed and thoughtful choice. Hence, the final decision will be a synthesis of one's own reflections and the insights gained from others. Here we will discuss an example of how this reasoning, derived from everyday life, can be applied to the field of energy forecasting.
The idea that a forecast can be made more robust through the synthesis of a pool of different predictions began to gain attention in the academic community as early as 1906 [44], when Victorian scientist Francis Galton visited a livestock fair in Plymouth, UK. During the event, he participated in a competition to estimate the weight of a ox—a common activity in rural settings where each participant submits their best guess of the animal's weight. At the end of the contest, the ox is weighed, and the person whose estimate is closest to the actual weight wins. On that occasion, Galton observed that the median of the roughly 800 estimates made had an error of less than 1% compared to the actual weight of the animal. From these preliminary observations, over more than a century of research, the rigorous formalization of a statistical method known as MA emerged. As the name suggests, this approach involves combining the predictions of multiple models using techniques that can vary in nature: from simple arithmetic averages to regression-based approaches that employ highly complicated algorithms [44, 45].

Specifically, MA has evolved to a powerful tool in modern forecasting applications, which consist of making predictions based on historical data sequences. As a remark, these techniques are referred to in the literature not only as MA — a term rooted in statistics and Machine Learning (ML) — but also as consensus forecasting, forecast combinations, and ensemble methods. One area where this methodology has been reported to be especially valuable is EPF [46, 17, 20, 12, 47], which consists in the estimation of future wholesale electricity prices [16]. This practice gained significance with the global shift toward energy market privatization in the 1990s. Across the West, competitive markets have largely replaced state monopolies, meaning operators now contend with prices determined not by regulatory authorities but by the dynamics of supply and demand [17]. Being able to predict these outcomes with an acceptable degree of error provides a competitive advantage, as it allows market operators to make more informed decisions on trading strategies, optimize production schedules, and mitigate financial risks.

Due to the highly volatile nature of electricity markets—driven by a range of factors like demand fluctuations, weather conditions, and the growing integration of renewable energy sources—it is often challenging to identify a single forecasting model (i.e., a specific combination of input variables and a statistical algorithm) that performs well across all scenarios. This variability means that different models may excel under specific conditions but fail under others. In the first place, performance depends on the predictors used: incorporating solar generation makes sense for daytime price forecasting, yet it would be meaningless for nighttime price predictions. Furthermore, variability also arises from the methods employed. A linear model, for example, might be sufficient to capture the correlation between gas prices and electricity prices, while solar generation impacts electricity prices in a non-linear way, requiring a different approach.

The first systematic study on this topic is [46], where the authors test various MA techniques against a single best model selected basing upon the performance achieved on past data. The authors examine different aggregation schemes and, by conducting tests on price datasets from four countries, report that using the arithmetic mean turns out to be the winning approach, despite its simplicity. However, the study was written at a time when probabilistic forecasts were not yet widely popular in EPF research, and therefore it is limited to point forecasts. While point forecasts provide a single estimated value (often, the expectation, or mean, of the target variable), they could fall short in volatile contexts where precise predictions are difficult to achieve [20]. Probabilistic forecasting

becomes critical, for instance by allowing us to estimate Prediction Intervals (PIs) within which future prices are likely to fluctuate, with given probability. PIs could be employed in scenario planning by utility companies, or in other risk mitigation practices. As a result, the need for both point and probabilistic approaches in energy forecasting has been increasingly acknowledged in the contemporary literature; see [20] and [12] for extensive reviews on this topic. In the European context, which we will focus on here, the uncertainty in electricity markets has surged due to two factors. On the one hand, the rapid and growing integration of many non-programmable renewable energy sources, which create frequent imbalances between demand and supply. On the other hand, and more recently, the energy crises following the Russia-Ukraine conflict.

Various studies have recently explored methods to enhance probabilistic EPF, often drawing on techniques previously popularized in the broader field of ML. Notable among these are efforts involving Neural Networks (NNs) [48, 47, 49, 23, 50] and Gradient Boosting [47], which have shown promise for improving forecast accuracy and reliability. However, the exploration of MA within probabilistic EPF remains limited [44, 51]. We believe it is worthwhile to deepen our understanding of probabilistic EPF, as predictor selection becomes even more complex in probabilistic forecasting than in point forecasting; the relevance of each predictor may shift depending on the quantile being forecasted. For instance, demand-related variables might be more effective for lower quantiles, while supply-side factors could improve predictions for upper quantiles. This study aims to shed light on this area by examining the effectiveness of different model averaging techniques. Specifically, we focus on the Italian day-ahead market, which is often used as a reference by international researchers due to its transparency and accessibility [13].

The present Chapter is structured as follows. In the next Section 3.2, we define the point and probabilistic forecasting problems, and present the methods used, including algorithmic details and the techniques used for aggregating forecasts within the MA framework. Additionally, we introduce the error and performance metrics employed in the study. In Section 3.3, we conduct an analysis of standalone models—i.e., without averaging—to identify the best predictors and predictor combinations for our problem, based on the different models used. Section 3.4 is the core of the study, as it examines the performance of ensemble models aggregated using methods ranging from simple to state-of-the-art techniques, achieving an 11% improvement in point forecasting error and a 9% improvement in probabilistic forecasting.

3.2 Methodology

In this section, we provide a brief overview of the methods and algorithms used in this Chapter. While the descriptions are self-contained, we have included several references for Readers interested in exploring the topics in more depth. Below, all point forecasting methods will focus on mean prediction (i.e. mean regression), which we will introduce first. Similarly, the probabilistic forecasting methods, introduced afterward, will all be based on Quantile Regression (QR). We chose this approach because quantile prediction is one of the most general forms of probabilistic forecasting, as it allows for modeling the entire distribution of the variable of interest [9]. Moreover, it directly enables the generation of arbitrary prediction PIs, which are the most common form of probabilistic forecasts found in the EPF literature [20]. One objective of this study is to compare the performance of structurally similar models for both mean and quantile forecasting. As we will see shortly, this approach is applied to both standalone models and ensemble models.

3.2.1 Mean Forecasting

We focus on the day-ahead market: assuming it is day d, the objective is to predict the expected value of electricity price $Y^{d+1,h}$ for day $(d+1)$ at scheduling period h. Numbering the days of the week from 0 to 6 (Monday to Sunday), and since the Italian market is discretized into hourly load periods, $h = 0,\ldots,23$, we obtain a total of $7 \times 24 = 168$ separate regression models for each method. For all d and h, we have a training set $\{(\mathbf{x}_i^{d,h}, y^{d+1,h})\}_{i=1}^{n}$ of $n = n(d,h)$ samples, defined as follows.

- The input $\mathbf{x}_i^{d,h}$ is a real-valued vector of predictors with size p, available at day d:

$$\mathbf{x}_i^{d,h} = \left[x_{i,(1)}^{d,h}, \ldots, x_{i,(p)}^{d,h}\right] \in \mathbb{R}^p, \quad \forall i = 1,\ldots,n.$$

  The size is constant for a given model, though it may vary when considering different models. For example, univariate models can be considered where the dimension is one ($p = 1$), or bivariate models where the dimension is two ($p = 2$), and so on. The predictors can include past values of electricity price or exogenous variables (see the next Section for detail).

- The output $y_i^{d+1,h}$ is a real scalar representing the electricity price for the day $(d+1)$ during the hourly period h. In other words, $y_i^{d+1,h}$ is a realization of the random variable of electricity prices, $Y^{d+1,h}$.

All the point forecasting models considered here provide estimates of the conditional expectation of the next day's price $Y^{d+1,h}$, given some realization $\mathbf{x}^{d,h}$ of the predictors $\mathbf{X}^{d,h}$ from the previous day d:

$$\hat{y}(\mathbf{x}^{d,h}) \simeq \mathbb{E}[Y^{d+1,h}|\mathbf{X}^{d,h} = \mathbf{x}^{d,h}]. \tag{3.1}$$

It is worth noting that this approach is the one that has almost always been followed in the point forecasting literature, with few exceptions, like for instance those based on Least Absolute Deviation (LAD) [52], which estimate the (conditional) median of the response variable. To simplify the notation, we will no longer explicitly reference d and h, unless the context requires it.

Linear Regression

The most classic and well-established model in any regression analysis, whether inferential or predictive, is Linear Regression (LR). In this model, the response variable is estimated as a linear combination of the input variables:

$$\hat{y}_{\mathrm{LR}}(\mathbf{x}) = \sum_{k=1}^{p} \theta_k x_{(k)} = \boldsymbol{\theta}^\top \mathbf{x}, \quad \theta_j \in \mathbb{R} \quad \forall k = 1, \ldots, p. \tag{3.2}$$

The LR is the simplest method belonging to the class of parametric models: its functional form is pre-specified, as in equation above (3.2), and the parameter vector $\boldsymbol{\theta}$ is determined from the training data $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$. Specifically, the optimal values are determined according to the least squares criterion, namely the minimization of the predictions' Mean Squared Error (MSE) over the training data:

$$\min_{\theta} \frac{1}{n} \sum_{i=1}^{n} \left(\hat{y}_{\mathrm{LR}}(\mathbf{x}_i) - y_i\right)^2. \tag{3.3}$$

The functional form of the estimator is linear, and the function to minimize, known as the loss function, is quadratic: the optimal parameters can be calculated analytically. LR

in this form is extensively covered in any graduate-level statistics textbook, such as [32]. For readers interested in exploring its applications in the field of ML, we recommend the manuals [11] and [10]. For an introductory exposition in the context of forecasting, see [53].

Kernel Regression

In this study, the nonlinear baseline for point forecasting is the Kernel Regression (KR) estimator, also known as the Nadaraya-Watson Estimator [11]. Given training data $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$, smooth estimates are obtained by taking the weighted average of the observed response variable y_i, with weights determined by kernel functions centered at each data point $\mathbf{x}_i$:

$$\hat{y}_{\mathrm{KR}}(\mathbf{x}) = \frac{\sum_{i=1}^n K_h(\mathbf{x}, \mathbf{x}_i) y_i}{\sum_{j=1}^n K_h(\mathbf{x}, \mathbf{x}_j)}, \tag{3.4}$$

where $K_h(\mathbf{x}, \mathbf{x}_i)$ is a kernel function with bandwidth vector $h = \left[h_{(1)}, \ldots, h_{(p)}\right]$, with positive entries, $h_{(j)} > 0$, $\forall j = 1, \ldots, p$. The kernel is used to measure the similarity between $\mathbf{x}$, the point where the estimation is made, and $\mathbf{x}_i$, the points or observations from the training data. Specifically, we use the Radial Basis Function (RBF), commonly known as the Gaussian kernel, which is a popular choice in literature. For arbitrary inputs $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^p$, the RBF reads

$$K_h(\mathbf{x}_1, \mathbf{x}_2) = \exp\left[-\frac{1}{2}\sum_{k=1}^{p}\left(\frac{x_{1,(k)} - x_{2,(k)}}{h_{(k)}}\right)^2\right]. \tag{3.5}$$

The bandwidth vector $\mathbf{h} = \left[h_{(1)}, \ldots, h_{(p)}\right]$ controls the model's behavior, directly affecting the goodness of fit with respect to the data. In fact, although KR is a non-parametric model [11], meaning it does not assume a predefined functional relationship between the input and output as in LR, it still relies on certain parameters to govern its behavior. In the case of kernel-based methods, the bandwidth parameters σ are crucial. These parameters must be optimized through various techniques, which can range widely in nature—from analytical rules that assume data normality and asymptotic limits [36] to information criteria [37] and Cross-Validation (CV) of the chosen error metric [11]. Here, we employ CV as our optimization approach. Given the time series nature of the forecasting task, we specifically apply time series CV [54], which uses an incremental

training dataset structure: each successive validation step trains on a slightly larger portion of the historical data, reflecting real-world data accumulation over time.
The error metric minimized throughout the optimization process is the Mean Absolute Error (MAE), which will also be the primary metric for the final model evaluation. Given observed values y_j and corresponding predictions $\hat{y}_j$, for $j = 1, \ldots, n$, it reads

$$\text{MAE} = \frac{1}{n} \sum_{j=1}^{n} |y_j - \hat{y}_m|. \tag{3.6}$$

The chosen optimization algorithm is the established Nelder-Mead simplex method [55]. It is worth noting that KR methods are relatively more popular in the statistical and econometric literature than in engineering fields. For a solid introduction to KR, see [36]; for a perspective from the field of ML, refer to [11] and [10].

3.2.2 Quantile Forecasting

In addition to estimating the expected value of electricity price $Y^{d+1,h}$ for day $(d+1)$ at scheduling period h, we can also predict specific quantiles of the price distribution. For quantile forecasting, the objective is to estimate the conditional quantiles of $Y^{d+1,h}$. This can provide broader perspective on the uncertainty associated with the predictions. For instance, it allows the construction of PIs with arbitrary associated probability.
Quantiles of a real-valued random variable Y are defined as follows. For a given quantile level $\alpha \in (0,1)$, the α-quantile of Y, denoted as $Q_Y(\alpha)$, is the value $y^{(\alpha)}$ such that:

$$\mathbb{P}(Y \leq y^{(\alpha)}) = \alpha. \tag{3.7}$$

This quantile represents a point in the distribution of Y below which a proportion α of the observations are expected to fall. Extending this concept to conditional quantiles, for any realization x of the predictor variable X, the conditional quantile of Y given $X = x$ is defined as:

$$Q_{Y|X}(\alpha|x) = \inf\{y \in \mathbb{R} : \mathbb{P}(Y \leq y|X = x) \geq \alpha\}. \tag{3.8}$$

In our setting, where we forecast electricity prices for each day $d+1$ and period h, the objective of quantile regression is to estimate the conditional quantile of $Y^{d+1,h}$ at a

specific level α, given observed values of the predictors from the previous day d and denoted as $\mathbf{x}^{d,h}$. Therefore, we aim to obtain estimates

$$\hat{y}^{(\alpha)}(\mathbf{x}^{d,h}) \simeq Q_{Y^{d+1,h}|\mathbf{X}^{d,h}}(\alpha|\mathbf{x}^{d,h}). \tag{3.9}$$

Here, $\hat{y}^{(\alpha)}(\mathbf{x}^{d,h})$ represents the estimate of the α-th conditional quantile of the electricity price for day $(d+1)$ and hourly period h based on the observed values of the predictors $\mathbf{x}^{d,h}$. As with mean forecasting, this analysis will be carried out across $7 \times 24 = 168$ separate quantile regression models, per each level α.
For example, in order to construct a central 90% PI we select two quantile levels, namely $\alpha = 0.05$ for the lower bound and $\alpha = 0.95$ for the upper bound. These quantile levels correspond to the 5th and 95th percentiles of the forecasted distribution, respectively. Once we estimate the corresponding quantiles $y^{(0.05)}$ and $y^{(0.95)}$ for the electricity price $Y^{d+1,h}$, the central 90% PI is given by the interval:

$$[y^{(0.05)}, y^{(0.95)}] \tag{3.10}$$

This interval indicates that there is a 90% probability, sometimes termed confidence level [56], that the actual electricity price $Y^{d+1,h}$ for day $(d+1)$ at scheduling period h will fall within this range. To simplify the notations, we will omit explicit references to d and h in what follows.

Linear Quantile Regression

The Linear Quantile Regression (LQR) model extends the concept of LR by estimating not the mean but specific quantiles of the response variable distribution. LQR is the first method developed to rigorously make such estimates, and was formalized by Koenker [9]. In LQR, the conditional α-quantile, for $\alpha \in (0,1)$, of the response variable Y is modeled as a linear combination of input variables, similar to the form used in LR (3.2):

$$\hat{y}^{(\alpha)}_{\text{LQR}}(\mathbf{x}) = \sum_{k=1}^{p} \theta_k x_{(k)} = \boldsymbol{\theta}^\top \mathbf{x}, \quad \theta_k \in \mathbb{R} \quad \forall k = 1, \ldots, p. \tag{3.11}$$

Instead of minimizing the MSE, as in LR, the LQR approach minimizes the Mean Pinball Loss (MPL), which is specifically suited to quantile estimation; see [57] and [9] for an

in-depth discussion. For any quantile level α, the Pinball Loss (PL) is defined as follows:

$$\rho_\alpha\left(\hat{y}^{(\alpha)}, y\right) = \begin{cases} \alpha \cdot |y - \hat{y}^{(\alpha)}|, & \text{if } y \geq \hat{y}^{(\alpha)}, \\ (1-\alpha) \cdot |y - \hat{y}^{(\alpha)}|, & \text{if } y < \hat{y}^{(\alpha)}, \end{cases} \tag{3.12}$$

where y is the observed valued and $\hat{y}^{(\alpha)}$ is the corresponding prediction. Here, the difference

$$y - \hat{y}^{(\alpha)} \tag{3.13}$$

is termed quantile residual at level α.
The LQR optimization criterion, therefore, seeks to minimize the MPL across all training observations $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$:

$$\min_\theta \frac{1}{n} \sum_{i=1}^{n} \rho_\alpha\left(\hat{y}_{\text{LQR}}^{(\alpha)}(\mathbf{x}_i), y_i\right). \tag{3.14}$$

Unlike standard LR, the LQR minimization problem does not have an analytical solution due to the piecewise nature of the MPL function. However, the MPL function remains convex in θ, allowing for efficient optimization through numerical methods such as those commonly employed in linear programming [58].

Kernel Quantile Regression

For any real-valued target variable Y and any specific observation $\mathbf{x}$ of the input variable $\mathbf{X}$, the conditional Cumulative Distribution Function (CDF) is defined as

$$F(y|\mathbf{x}) = \mathbb{P}(Y \leq y | \mathbf{X} = \mathbf{x}), \quad y \in \mathbb{R}. \tag{3.15}$$

Notably, the CDF of Y given $\mathbf{X} = \mathbf{x}$ can also be expressed as the conditional expectation of an indicator function, specifically

$$F(y|\mathbf{x}) = \mathbb{E}[I(Y \leq y) | \mathbf{X} = \mathbf{x}], \tag{3.16}$$

where $I(A)$ takes a value of 1 if A is true, and 0 otherwise. This relationship provides a basis for estimating $F(y|\mathbf{x})$ using the KR estimator [41], as introduced in eq. (3.4):

$$\hat{F}(y|\mathbf{x}) = \frac{\sum_{i=1}^{n} K_h(\mathbf{x}, \mathbf{x}_i) I(y_i \leq y)}{\sum_{j=1}^{n} K_h(\mathbf{x}, \mathbf{x}_j)}, \tag{3.17}$$

where the numerator accounts for the similarity-weighted cumulative probability that $Y \leq y$ for each observed y_i, while the denominator normalizes these weights a cross all n data points. Also in this case, K_h, is a kernel function of bandwidth h, set to the RBF in our study. This family of estimators was extensively discussed in [41], including nearest-neighbors methods and their convergence properties. A good introduction is provided in [36].
Similarly to the case of KR for mean regression, the bandwidth parameters h govern the behavior of the model, thus requiring proper tuning. Given a specific quantile level $\alpha \in (0,1)$, the bandwidth values can be optimized through CV, where the objective function to be minimized is the MPL. Again, the employed optimization method is Nelder-Mead.

3.2.3 Model Averaging for Mean Regression

Let us now introduce the methods that form the primary focus of this study: those based on MA, starting as usual from the mean regression task. Each of these methods is founded on the principle of starting with a set of individual forecasting models, known collectively as an ensemble.
The ensembles can be heterogeneous, consisting of models that may differ in nature (e.g. LR, KR), or homogeneous, representing instances of the same statistical model trained with varying sets of input variables [44]. Another possibility, not considered in this study, concerns ensembles of several instances of the same model trained on different sets. More generally, the ensemble size, denoted by M, can vary depending on the forecasting scenario and computational resources.
There are countless possibilities for aggregating forecasts. The work [44] provides an extensive review. Broadly, we can categorize these approaches into two main types: voting and stacking. Voting methods aggregate predictions through a simple statistic, such as the (possibly weighted) mean, or the median. They are termed "voting" because,

in the context of ensemble classifiers, the final prediction represents the majority vote among the models in the set. In contrast, stacking methods are more advanced: they employ an additional statistical model that takes the outputs of the ensemble models as inputs and generates the final forecast. Despite this approach enables a dynamic weighting of predictions, potentially leading to more accurate results, its superiority is not guaranteed in practice.

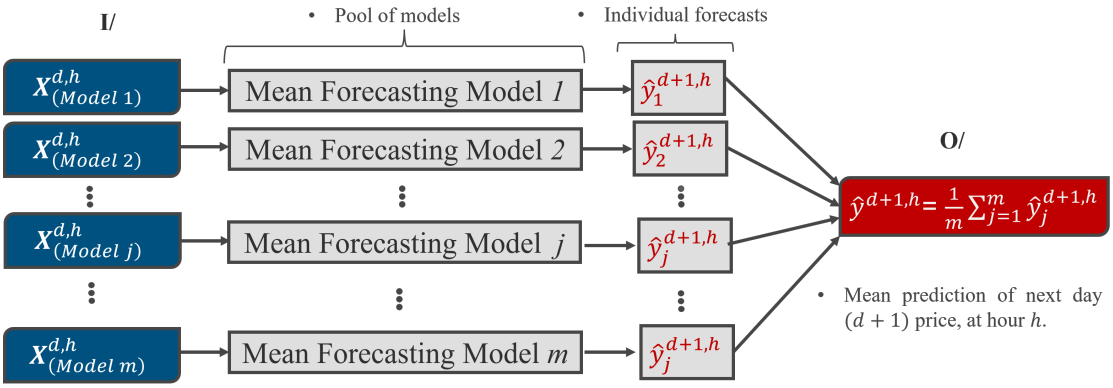

Fig. 3.1 **Schematic Diagram for Voting Mean Regression.**
The final estimate in this simple averaging scheme is calculated as the arithmetic mean of the individual predictions from each model in the pool. This approach, known as Voting Regression, has demonstrated strong empirical performance in ensemble forecasting, consistently outperforming many more advanced solutions that involve weighted averaging or regression-based techniques [44].

Voting Mean Regression

For Voting Mean Regression (VMR), the final forecast $\hat{y}^{d+1,h}$ is obtained by calculating the arithmetic mean of the individual predictions $\hat{y}_j^{d+1,h}$ for each model j in the ensemble:

$$\hat{y}^{d+1,h} = \frac{1}{M}\sum_{j=1}^{M} \hat{y}_j^{d+1,h}, \tag{3.18}$$

where M represents the total number of forecasts. This simple averaging approach has shown robust performance across a range of forecasting tasks [44], often outperforming more complex ensemble methods involving weighted averages or additional regression layers (i.e. those in the stacking family). A visual representation of VMR is presented in Figure 3.1.

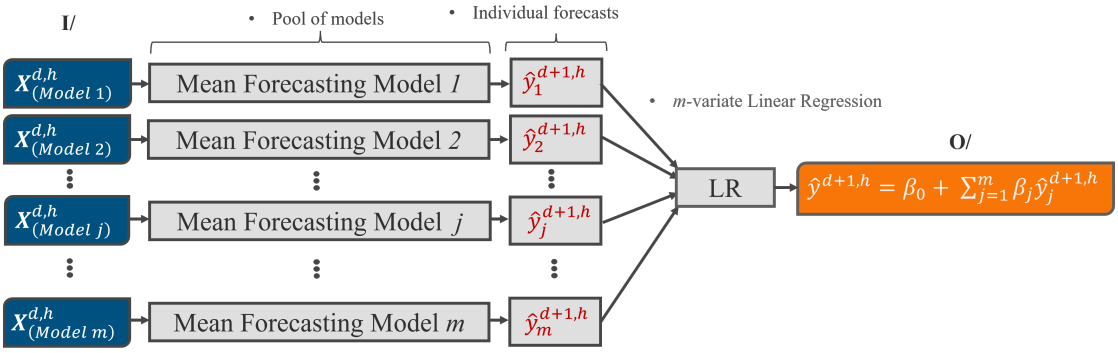

Fig. 3.2 **Schematic Diagram for Stacking Mean Regression with LR.**
The figure presents an illustration for the working principle of the Stacking Mean Regression. A set of diverse forecasts is aggregated by means of a regression method: Linear Regression (LR) in this case.

Stacking Mean Regression

For Stacking Mean Regression (SMR), in this study, the final forecast $\hat{y}^{d+1,h}$ is produced by training a simple LR that uses the individual predictions $\hat{y}_j^{d+1,h}$ from each model j in the ensemble as input features. This approach is represented by the following equation:

$$\hat{y}^{d+1,h} = \beta_0 + \sum_{j=1}^{M} \beta_j \hat{y}_j^{d+1,h}, \tag{3.19}$$

where β_0 is the intercept, and β_j are the weights assigned to each model's forecast, estimated by fitting the LR on historical prediction data. Although the stacking approach can, in principle, incorporate more advanced regression models than LR, this falls beyond the scope of this investigation, where our primary interest lies in understanding the conceptual and empirical aspects of the various aggregation methods. Notably, the term "stacking" originated within the field of NNs[59]. A visual schematization is presented in Figure 3.2.

Random Forest and Extra-Trees Regression

In this study, Random Forest (RF) serves as one of the ensemble baselines for point forecasting. We now briefly recall its functioning, referring the reader to [10] for an in-depth discussion. Given training data $\{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$, the RF estimator combines multiple decision trees, each constructed on different bootstrapped subsets of the training data and

with random subsets of predictors at each node. For an arbitrary input $\mathbf{x}$, the prediction from an RF model, $\hat{y}_{\mathrm{RF}}(\mathbf{x})$, is obtained by averaging the predictions from all M trees in the ensemble:

$$\hat{y}_{\mathrm{RF}}(\mathbf{x}) = \frac{1}{M}\sum_{m=1}^{M} \hat{y}_{\mathrm{T}^{(m)}}(\mathbf{x}), \tag{3.20}$$

where $\hat{y}_{\mathrm{T}^{(m)}}(\mathbf{x})$ represents the prediction made by the m-th decision tree; $m = 1,\ldots,M$. Each decision tree $\mathrm{T}^{(m)}$ is trained on a bootstrapped sample of the data $\{(\mathbf{x}_{i_m}, y_{i_m})\}_{i_m=1}^{n_m}$, where n_m is the number of points in the m-th bootstrapped sample. We recall that a bootstrap sample is a randomly selected subset of the data created by sampling with replacement from the original dataset, allowing observations to be repeated.
In the training phase, RF applies additional randomness by selecting a subset of predictors at each split, thus decorrelating the individual trees and enhancing the diversity of the ensemble. The algorithm typically uses the MSE, or the MAE (as in our case), to determine the optimal splits within each tree.
Each individual tree $\mathrm{T}^{(m)}$ in the ensemble can be seen as a partition of the input space $\mathbb{R}^p$, where each partition or leaf node contains the mean response value of the observations that fall within it. Thus, the prediction of a single tree for a given input $\mathbf{x}$ is:

$$\hat{y}_{\mathrm{T}^{(m)}}(\mathbf{x}) = \frac{1}{|\mathscr{L}^{(m)}(\mathbf{x})|}\sum_{i\in\mathscr{L}^{(m)}(\mathbf{x})} y_i, \tag{3.21}$$

where $\mathscr{L}^{(m)}(\mathbf{x})$ is the set of indices of the training points falling into the same leaf as $\mathbf{x}$ in the m-th tree, and $|\mathscr{L}^{(m)}(\mathbf{x})|$ denotes its cardinality.
The RF model's accuracy and robustness come from averaging over multiple, diverse trees, which reduces overfitting and variance compared to individual decision trees. Furthermore, RF naturally provides feature importance measures by examining the reduction in impurity (MSE, MAE, as in our case, or other error metrics) at each split across all trees in the ensemble. Here, we highlight that in the field of ML, RF emerged in the 1980s as one of the earliest methods to reveal to the scientific community the potential of ensemble approaches, thus becoming a standard among these types of methods.
This study also examines the Extremely Randomized Trees Regression algorithm (ETR, also known as the "Extra-Trees" algorithm), which differs from RF in the approach to constructing individual trees. Actually, it was first proposed in [60] as a variation to

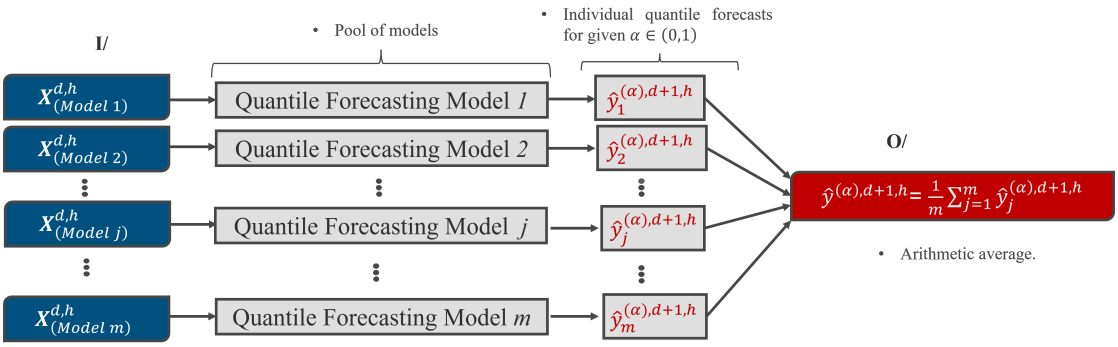

Fig. 3.3 **Schematic Diagram for Voting Quantile Regression.**
Similarly to Voting Mean Regression, final estimate in this scheme is calculated as the arithmetic mean of the individual predictions from each model in the pool.

RF. While both methods create an ensemble of decision trees, the ETR has two specific characteristics:

1. it uses the entire dataset rather than bootstrap samples;
2. it selects split points uniformly at random within the range of each feature, instead of choosing the most optimal split based on an impurity measure.

These adjustments generally lead to increased diversity among individual trees, potentially enhancing prediction performance by further decorrelating the trees in the ensemble. Thus, ETR may achieve lower variance for the overall prediction, although it tends to be more sensitive to noisy features and outliers than RF.

3.2.4 Model Averaging for Quantile Regression

We now introduce the MA methods for QR, which, as was previously discussed, is the chosen approach in this study for generating probabilistic forecasts. Although several studies focus on averaging techniques for PIs or CDFs (see [44] for a review), these are beyond the scope of our investigation.

Voting Quantile Regression

In Voting Quantile Regression (VQR), the approach is similar to VMR, but instead of averaging mean predictions, we take the arithmetic mean of the estimated quantiles $\hat{y}_j^{(\alpha)}$ at the desired quantile level α. The ensemble quantile prediction is given by:

$$\hat{y}^{(\alpha)}(\mathbf{x}) = \frac{1}{M} \sum_{j=1}^{M} \hat{y}_j^{(\alpha)}(\mathbf{x}). \tag{3.22}$$

A visual representation is provided in Fig. 3.3.

Quantile Regression Forest and Extremely Randomized Trees Quantile Regression

Introduced in 2006, Quantile Regression Forest (QRF) extends the traditional RF algorithm to estimate conditional quantiles of the response variable [61]. Given training data $\{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$, QRF uses a set of regression trees $\{T^{(m)}\}_{m=1}^{M}$ constructed similarly to RF, but with a focus on estimating quantiles instead of means. Although QRF is a considerably more recent approach than its mean regression counterpart, it has recently gained traction in probabilistic forecasting applications, as seen in studies like [56], where the authors use it as a benchmark model to compare against the method they propose. Similarly, Extra-Trees Quantile Regression (ETQR) modifies the Extra-Trees (Extremely Randomized Trees) algorithm to provide quantile estimates. However, ETQR differs in two key ways: (1) it does not use bootstrap samples, instead training each tree on the entire dataset, and (2) it selects split points within each feature at random, without attempting to optimize the impurity criterion.

Quantile Regression Averaging and Factor Quantile Regression Averaging

The last model, or rather family of models, that we introduce here is particularly important, as it represents the state-of-the-art in performance for MA applied to probabilistic forecasts [20, 18]. This method gained significant recognition during a major scientific forecasting competition, the Global Energy Forecasting Competition (GEFCOM), organized by the IEEE Power and Energy Society. The 2014 edition of GEFCOM was especially notable as it marked the first time that the competition focused exclusively on

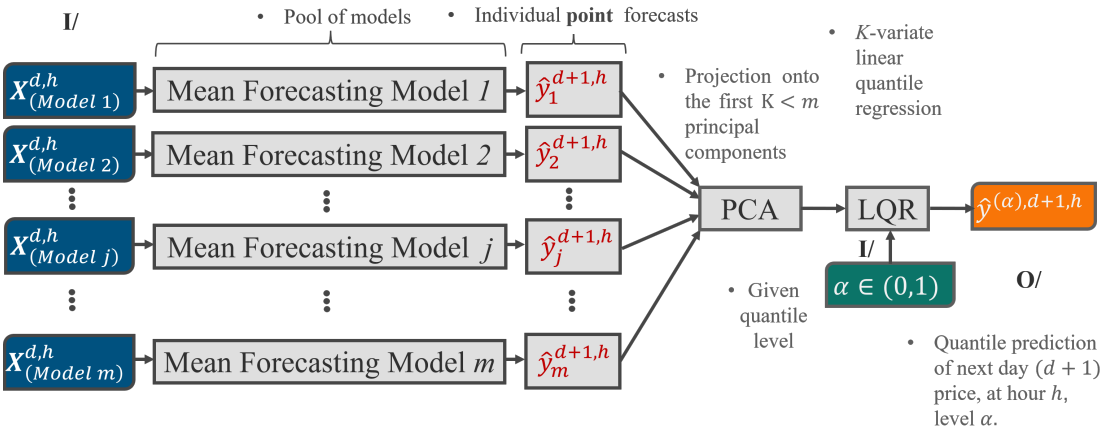

Fig. 3.4 **Schematic Diagram for Stacking Quantile Regression with FQRA.**
The figure presents an illustration for the working principle of the Factor Quantile Regression Averaging (FQRA) method [18]. A diverse set of point forecasts is aggregated into a probabilistic forecast by means of two steps. Firstly, the predictions are fed into a dimensionality reduction method, namely Principal Component Analysis (PCA). Secondly, a Linear Quantile Regression (LQR) is performed to yield quantile predictions at a given level.

probabilistic forecasting [62]. Among the top four solutions, two relied on the so-called Quantile Regression Averaging (QRA) method [20], bringing it to popularity in the research field. QRA-based methods combine outputs from multiple models using LQR to generate the final quantile forecast, directly fitting the target quantile level to a linear combination of point predictions from individual models. Therefore, it can be thought of as a stacking ensemble method. Since QRA is not a new concept but one whose application has only recently been explored, there is not yet a standard approach. In some studies (e.g. [56, 51]), ensembles of quantile forecasting models are used, whereas in others [20, 18] — including our study — the QR is applied solely for the aggregation phase. Indeed, according to influential authors like Nowotarski [20], QRA is primarily intended to leverage the expertise that the scientific community has developed in generating point forecasts, while it is known that there is comparatively less knowledge in the newer field of probabilistic forecasting. Further, it is important to remember that usually mean regression methods involve less computational burden than their quantile versions. Given a set of M individual models, each producing predictions $\hat{y}_j(\mathbf{x})$ for a given input $\mathbf{x}$, the QRA ensemble forecast for quantile level α is expressed as:

$$\hat{y}^{(\alpha)}_{\text{QRA}}(\mathbf{x}) = \beta_0 + \sum_{j=1}^{M} \beta_j \hat{y}_j(\mathbf{x}), \tag{3.23}$$

where the coefficients β_j are estimated via quantile regression; namely by minimizing the MPL for level α. This approach allows QRA to incorporate diverse models or input configurations while directly targeting the quantile level of interest.
However, QRA has a drawback. As was explained above, the calculation of the optimal LQR coefficients relies on convex optimization algorithms. Since the ensemble of models used in applications like EPF often presents a high degree of similarity (see the next section for a detailed numerical comparison), the forecasts generated by these models tend to exhibit high correlation [45]. This can challenge the optimization algorithm, especially when there are many degrees of freedom, that is, when the ensemble includes a large number M of individual models. Such a redundancy can potentially lead to the calculation of suboptimal solutions, as a result of algorithmic failures.
Factor Quantile Regression Averaging (FQRA) [18] is an extension to QRA designed to address redundancy and instability in the optimization process. By applying Principal Component Analysis (PCA) to the predictions of individual models, FQRA reduces dimensionality before performing QR on the transformed predictor space. This preprocessing step stabilizes the linear programming optimization, mitigating issues related to collinearity among the ensemble's model outputs.
Given M individual models $\hat{y}_j$, FQRA follows these steps:

- Construct the $n \times M$ matrix of predictions,

$$\hat{\mathbf{Y}} = \left(\hat{y}_j(\mathbf{x}_i)\right)_{ij}; \quad i = 1,\ldots, \quad n, j = 1,\ldots,M. \tag{3.24}$$

- Apply PCA to $\hat{\mathbf{Y}}$ to extract the leading K principal components; usually we limit to the first one ($K = 1$).

- Perform LQR on the reduced component space, estimating the quantile at a given level $\alpha \in (0,1)$ as

$$\hat{y}^{(\alpha)}_{\text{FQRA}} = \gamma_0 + \sum_{k=1}^{K} \gamma_k \text{PC}_k, \tag{3.25}$$

  where PC_k are the selected principal components, and γ_k are coefficients estimated to optimize the quantile loss for the target level α.

This approach is intended to preserve essential predictive information while reducing the impact of noise. A visual representation is presented in Figure 3.4.

3.3 Standalone Analysis

Before delving into the analysis of ensemble methods, it is essential to assess the performance of individual, or standalone, models. This analysis provides insights into the predictors available and helps identify the most effective ones, whether used alone or in combination. Initially, we focus on estimating the expected value, or mean, followed by quantiles.
Recall that our objective is predicting the electricity price for the day $d+1$ at the hour h. To start, we list the available predictors, namely a set of variables supposed to be related to our target:

- (past) electricity price, using the previous day's value for the same hour h of day d, expressed in €/MWh (**Price**);
- previous day's natural gas price in €/MWh, averaged daily, as Italy's day-ahead gas market does not operate on an hourly basis (**Gas**);
- load forecast for day $d+1$ at hour h, expressed in GW (**Load**);
- solar generation forecast for day $d+1$ at hour h, in MWh (**Solar**);
- thermal generation forecast for day $d+1$ at hour h, in MWh (**Thermal**);
- hydroelectric generation forecast for day $d+1$ at hour h, in MWh (**Hydro**);
- scheduled cross-border exchanges for day $d+1$ at hour h, in MWh (**Forex**).

All the data are freely downloadable from the sources [43] and [42]. It should be noted that, in this study, the forecasts for load and generation actually correspond to the true observed values, thus amounting to perfect forecasts. This assumption is justified by the high short-term predictability of these variables, which can be estimated with sufficient accuracy. Numerical weather predictions have not been included, as this information is indirectly embedded within the renewable energy generation variables, which capture the impact of weather on non-dispatchable sources [63]. For a more extensive discussion on the role and relevance of these variables within the Italian electricity market, in both point and probabilistic forecasting, refer to [64].

3.3.1 Mean Forecasting Results (Standalone Models)

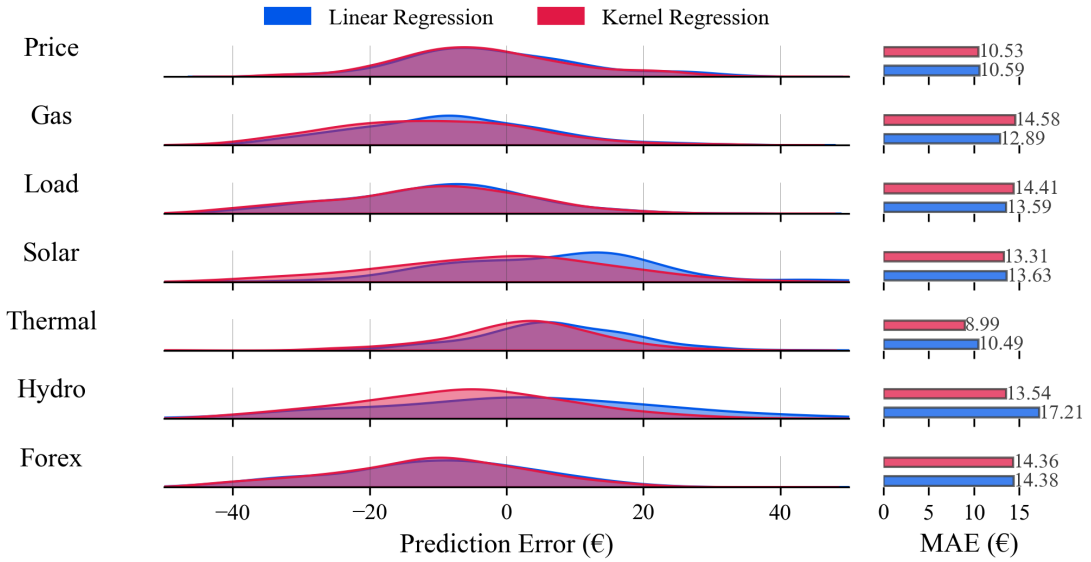

Fig. 3.5 **Error Distribution Analysis for Univariate Point Forecasting Models.**
The figure provides an analysis of the error distribution for univariate mean forecasting models (each using a single predictor). Each row corresponds to a different predictor, with the predictor's name displayed on the left. In the center of each row, Kernel Density Estimation plots show the error distributions: the blue curve represents the errors for the Linear Regressions, and the red curve represents the errors for the Kernel Regressions. On the right, the Mean Absolute Error (MAE) for each model, reported in euro.

The following results focus on a test period limited to peakload hours, specifically from 8 AM to 8 PM, on weekdays (Monday to Friday) during May and June 2024. May 1st was omitted as it is a public holiday. The training data covers two distinct periods: January to August 2021 and January 2023 to April 2024. These periods were chosen to exclude intervals with irregular price patterns, as discussed in [64].
We begin by conducting an initial evaluation of the potential input variables. Figure 3.5 displays the error distributions for univariate mean forecasting models, each utilizing a single predictor. Each row represents a different predictor, labeled on the left. The central portion of each row features Kernel Density Estimation (KDE) plots that depict the error distributions: the blue curve corresponds to errors from the univariate LR model, while the red curve represents errors from the univariate KR model. On the right, bar plots show the MAE for each model, measured in euros. The goal is to achieve an error distribution that is symmetric around zero, tightly clustered, and characterized by a low average error. We refer the Reader back to Chapter 2 for an introduction

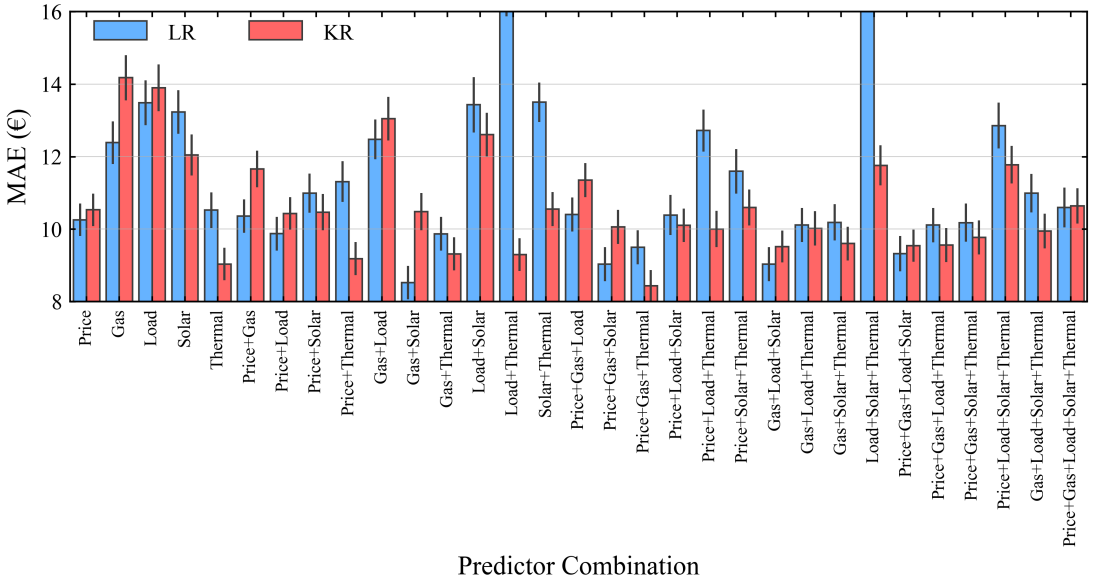

Fig. 3.6 **Comparison of Mean Absolute Error for Different Predictor Combinations.**
This summary chart presents the Mean Absolute Error (MAE), expressed in euro, for various predictor combinations. The blue bars represent the results from Linear Regression (LR), while the red bars show the results from Kernel Regression (KR). The black lines above each bar indicate the standard error, which is the standard deviation divided by the square root of the number of observations.

to KDE. Among the predictors, Thermal stands out as the most effective, delivering the lowest average errors and the most concentrated error distributions for both linear and nonlinear models. Unsurprisingly, the Price variable follows closely. Gas, another strong predictor, exhibits a natural correlation with thermal generation, as the majority of Italy's thermoelectric plants rely on natural gas. In this case, LR outperforms KR. Load demonstrates a less competitive performance: its predictive relevance has diminished over time due to increasing price volatility, while Load maintains stable and seasonal patterns. Solar generation, while associated with a low average error, shows a broader error distribution, reflecting its inherent variability. On the other hand, Forex and Hydro perform poorly and are consequently excluded from further consideration.
Moving forward, our goal is to explore the interactions among the remaining predictors by analyzing the behavior of multivariate models. To this end, we compare LR with KR across various predictor combinations. We first discuss the summary of average errors before delving into the detailed error distributions. Figure 3.6 uses a barplot to display the MAE for models built using different input configurations. The x-axis lists the predictor combinations, such as "Load" for a univariate model using only electrical load, and

"Load+Thermal" for a bivariate model incorporating both load and thermal generation. The y-axis represents the MAE in euros. Blue bars indicate results for LR models, while red bars represent KR models. Black lines above each bar illustrate the Standard Error (SE). The best-performing predictor combination is "Price+Gas+Thermal," achieving an average MAE of €8.84 across LR and KR. Specifically, the linear model yields an MAE of € 9.22, while the nonlinear model achieves a notably lower € 8.46, which also represents the minimum MAE for KR. Interestingly, the top-performing linear model is the "Gas+Solar" combination, a relationship not evident in the univariate analysis. However, this model exhibits high variance, despite its low mean error.
Generally, the characteristics of individual predictors persist when combined in multivariate models. For instance, "Solar" tends to introduce outliers in the error distribution due to its high variability, particularly when paired with "Load." Combinations such as "Load+Solar" and "Load+Solar+Thermal" perform the worst, demonstrating that including more predictors does not always improve model performance. For example, the five-variable KR model, which includes all predictors, underperforms by 20% compared to the best KR model. Table 3.1 provides a comprehensive summary of average errors, including the MAPE and the MSE as additional metrics.

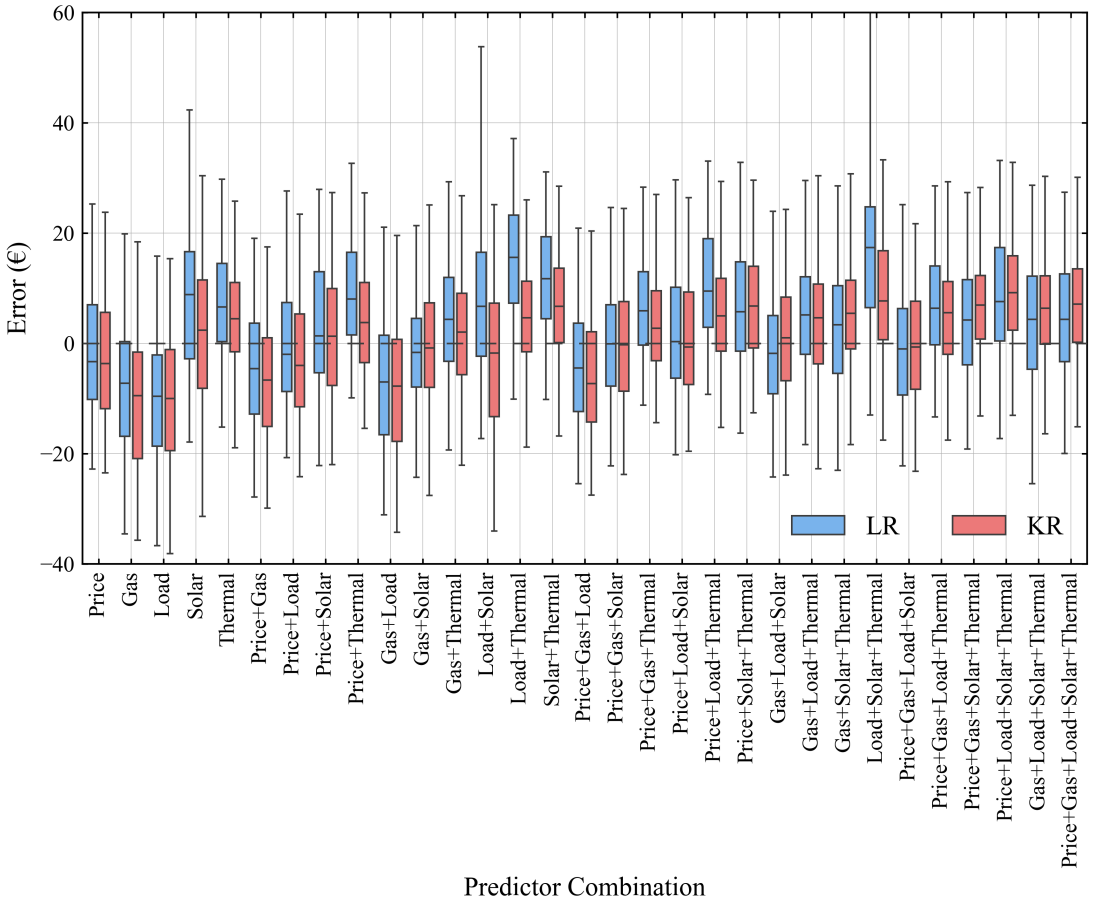

Fig. 3.7 **Comparison of Prediction Errors for Different Predictor Combinations.**
The figure shows the prediction errors for models with different combinations of predictors. The x-axis represents the predictor combinations, where, for example, "Load" refers to a univariate model using only electrical load as input, and "Load+Thermal" refers to a bivariate model using both load and thermal generation as inputs. The y-axis shows the model errors. For each combination of predictors, two boxplots are presented: the left (blue) boxplot represents errors from Linear Regression (LR) models, and the right (red) boxplot represents errors from Kernel Regression (KR) models. The boxplots display the intertquartile range, with whiskers extending to cover the central 95% of the data distribution.

To complement, Figure 3.7 provides a detailed view of the error distributions for each predictor combination. Along the x-axis, the predictor combinations are labeled, while the y-axis measures model errors. Each combination is represented by two boxplots: the left (blue) boxplot corresponds to LR models, and the right (red) boxplot corresponds to KR models. The boxplots display the InterQuartile Range (IQR), with whiskers extending from the 2.5th percentile to the 97.5th percentile, covering the central 95% of

the data distribution. This visualization underscores the variability observed in the error distributions, particularly for combinations involving "Solar" and "Load."
From a broader perspective, the performance differences among the models are relatively modest. Some models exhibit a tendency to overestimate (indicated by a negative median error), while others tend to underestimate (positive median error). As we will explore in subsequent sections, aggregating predictions using a MA approach can help mitigate these biases by balancing over- and under-estimations.

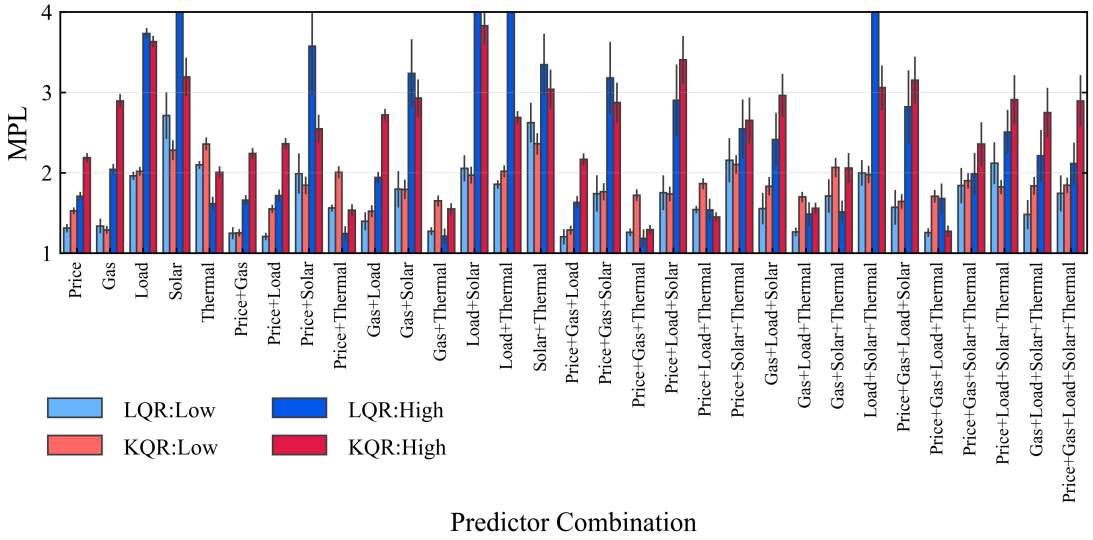

Fig. 3.9 **Comparison of Mean Pinball Loss for Different Predictor Combinations.**
This chart illustrates the Mean Pinball Loss (MPL) for various predictor combinations, with a layout similar to the previous MAE comparison. For each predictor combination, four bars are presented: lighter shades represent Low quantiles (0.05), while darker shades correspond to High quantiles (0.95). Blue shades indicate Linear Regression (LR), and red shades represent Kernel Regression (KR). Black lines represent standard errors.

3.3.2 Quantile Forecasting Results (Standalone Models)

We now shift our focus to individual models for the two quantile regression tasks: predicting the 0.05 quantile (referred to as "Low") and the 0.95 quantile (referred to as "High"). For each task, separate regressions are constructed for both LQR and KQR, following the procedures outlined in Section 3.2.
We start with the average performance of each model, in terms of PL. Figure 3.9 presents a

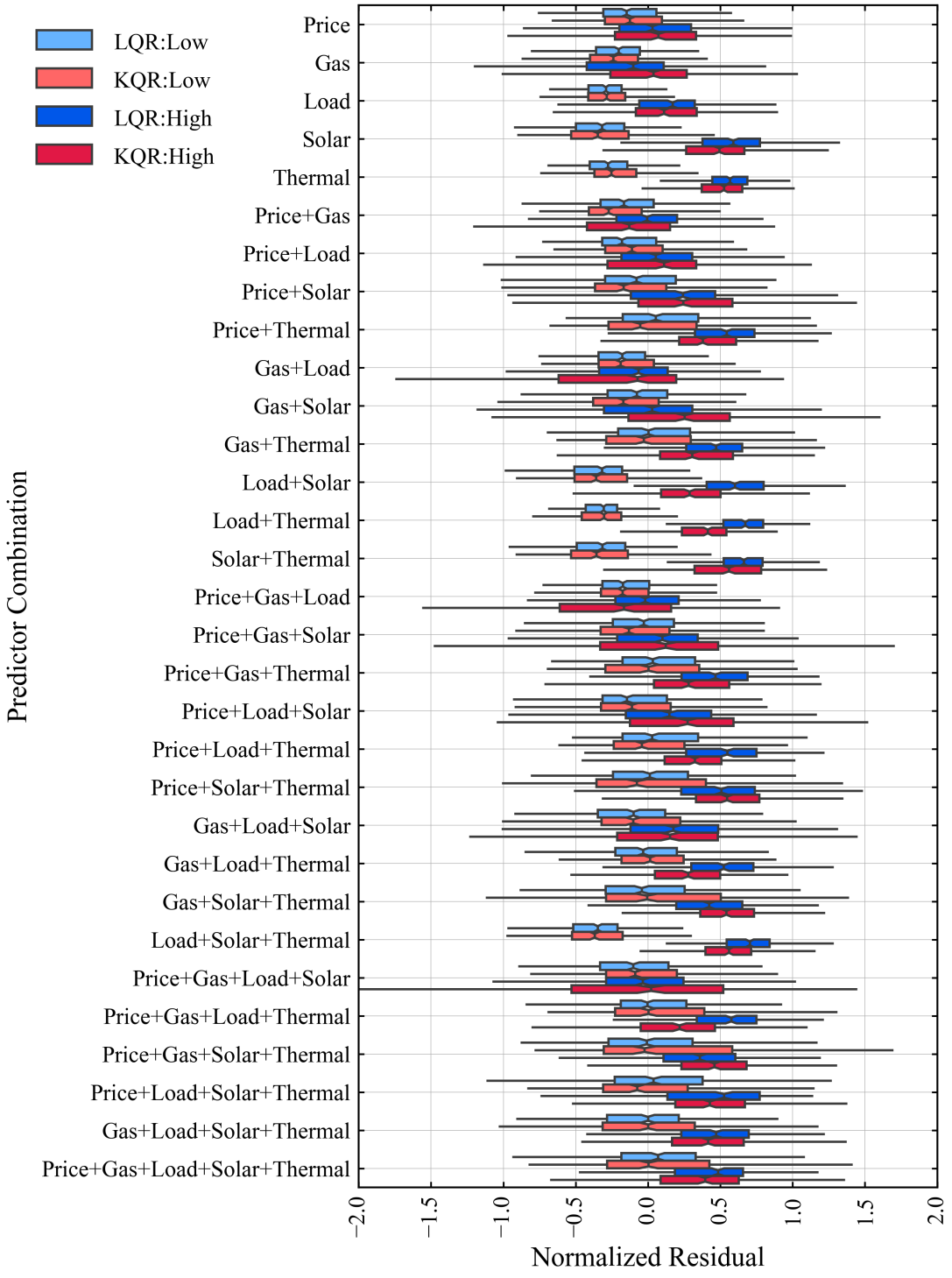

Fig. 3.8 **Comparison of Normalized Quantile Residuals for Different Predictor Combinations.** *This chart examines the distributions of the normalized quantile residuals at levels 0.05 (Low) and 0.95 (High). Blue represents Linear Quantile Regression (LQR), while red represents Kernel Quantile Regression (KQR). Light shades indicate the residuals for the 0.05 quantile, while darker shades indicate those for the 0.95 quantile.*

comparative summary of the MPL for various predictor combinations. Each combination is depicted with four bars: lighter shades represent the Low quantiles (0.05), and darker shades represent the High quantiles (0.95). Blue tones indicate LQR results, while red tones indicate KQR results, maintaining the same shading scheme as before. The layout is similar to the MAE comparison in Figure 3.6. Additionally, Table 3.2 provides detailed performance metrics for each model. For Low quantiles, several predictor combinations achieve near-optimal performance. The KQR model performs best with the "Price + Gas" combination (MPL = 1.25), while the LQR model excels with "Price + Gas + Load" (MPL = 1.21). However, Solar emerges as the poorest predictor for Low quantiles, exhibiting wide error distributions whenever included. For High quantiles, the results show increased variance and a greater divergence between LQR and KQR models for univariate predictors compared to Low. The optimal results for High are achieved with the "Price + Gas + Thermal" combination, which also performed best for mean prediction. This combination yields MPL values of 1.29 for KQR and 1.19 for LQR. The quadrivariate combination "Price+Gas+Load+Thermal" slightly reduces the MPL for KQR to 1.27.

Next, we examine the detailed error distributions for each predictor combination and regression type (LQR or KQR). Figure 3.8 presents the distributions of quantile residuals at levels 0.05 and 0.95. for each individual model. Blue represents LQR, while red represents KQR, with lighter shades indicating residuals for Low and darker shades for High. Outliers are excluded from this plot to improve readability.The results for Low quantiles consistently exhibit lower variance than those for High quantiles across nearly all predictor combinations. This suggests that the lower quantile is generally easier to forecast than the upper quantile, likely due to the greater volatility of electricity prices at higher levels. This observation is further supported by Figure 3.9, where MPL values for Low quantiles are generally lower and lack extreme peaks, while High quantiles show a higher risk of substantial forecasting errors if suboptimal predictor combinations are chosen. In summary, the "Price + Gas + Thermal" combination stands out as the most robust for both Low and High quantiles, while Solar consistently underperforms, particularly for Low quantiles. The increased variance and divergence in results for High quantiles expresses the greater difficulty in forecasting high price levels.

3.3.3 Overall Considerations

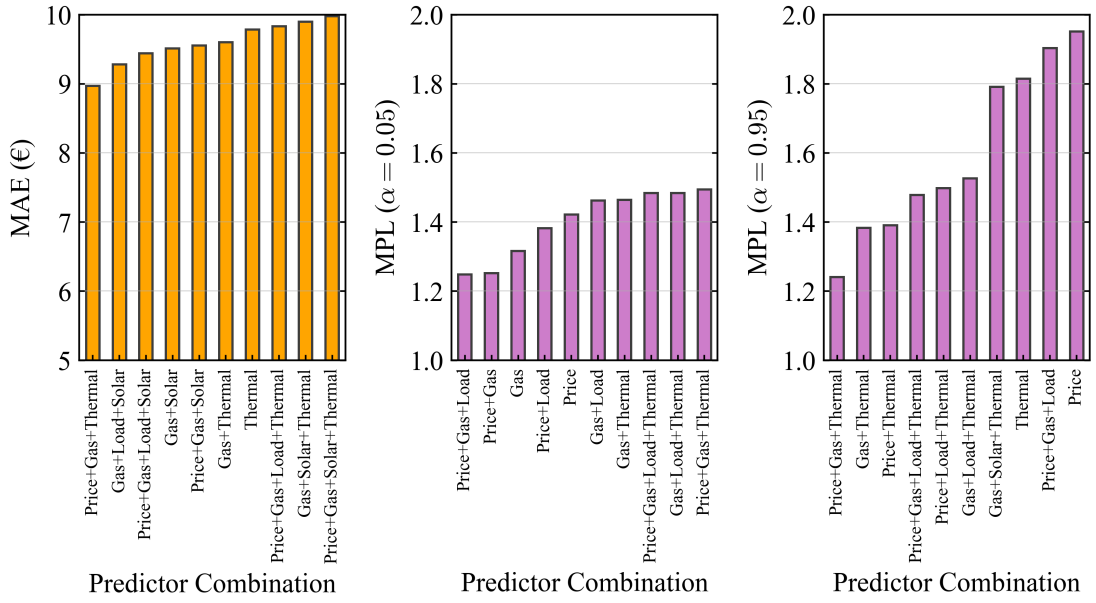

Fig. 3.10 **Best Ten Predictor Combinations sorted by increasing Error Metrics.**
This figure presents three barplots, one for each forecasting task: mean regression, level 0.05 quantile regression, and level 0.95 quantile regression, from left to right. On the x-axis of each barplot are the ten predictor combinations with the best average performance (of linear and kernel regression), arranged in decreasing order of performance. The first plot (mean regression) shows the Mean Absolute Error (MAE) on the y-axis, expressed in euro, while the last two plots (quantile regressions) show the Mean Pinball Loss (MPL). Notably, for both the mean and 0.05 quantile regressions, the top 10 predictor combinations are relatively similar, while for the 0.95 quantile regression, only the first six combinations perform well, with Thermal emerging as the best single predictor in this case.

Having individually analyzed the three forecasting tasks—mean, Low (0.05 quantile), and High (0.95 quantile)—we now consolidate our findings by referencing Figure 3.10. This figure displays three barplots: mean regression, 0.05 quantile regression, and 0.95 quantile regression, arranged from left to right. The x-axis of each barplot lists the ten predictor combinations with the best average performance, ranked in descending order of effectiveness. The mean regression plot (first, in orange) uses the y-axis to represent MAE in euros, while the quantile regression plots (in purple) use the y-axis to display MPL. For mean and Low regressions, the performance of the top 10 predictor combinations is relatively consistent, with variables such as Load and Price frequently appearing in these combinations. In contrast, for High regression, only the top 5 or 6 combinations show comparable performance, with Thermal emerging as the strongest

individual predictor in this context. Notably, Load is a recurring top predictor for the Low (0.05) task, likely due to its stability and periodic nature, which helps in capturing the lower bounds of price behavior under typical demand conditions. On the other hand, Solar is primarily effective for mean forecasting, as its inherent variability tends to average out over time, making it less reliable for predicting extreme quantile movements. Variables like Price, Gas, and Thermal consistently deliver strong results across all tasks.

3.4 Ensemble Analysis

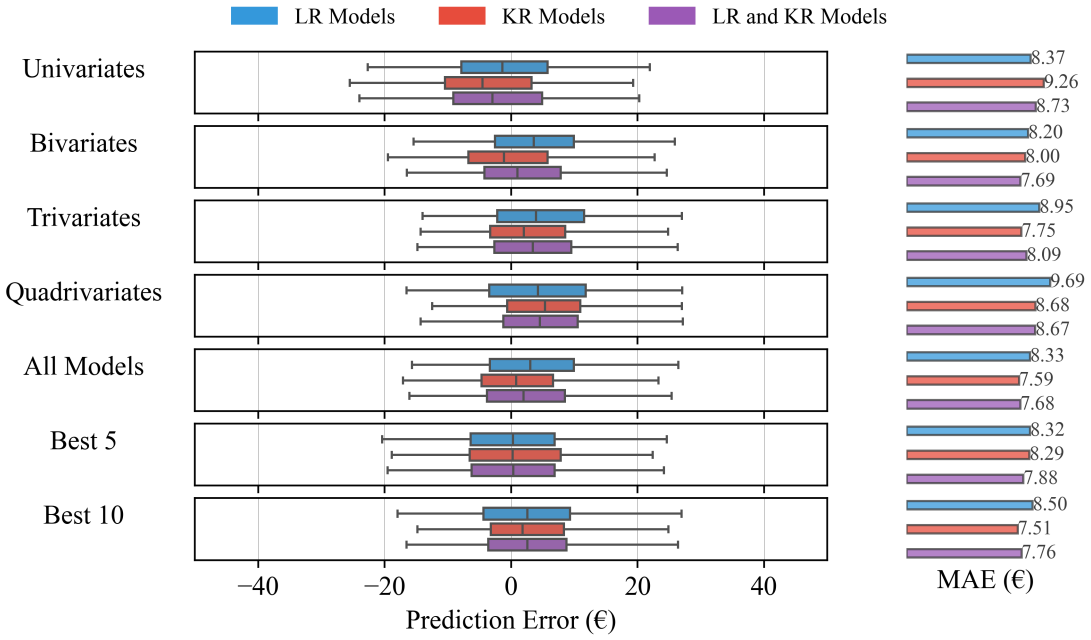

Fig. 3.11 **Error Distribution Comparison for Voting Mean Regression Ensembles.**
This figure compares the error distribution for the Voting Mean Regression (VMR) method. Each row corresponds to a different ensemble configuration, based on the predictors used: the first row represents univariate ensembles, the second bivariate, followed by trivariate, quadrivariate, and "All Models," which includes all possible predictor combinations. The "Best 5" and "Best 10" ensembles include the models formed from the top 5 and 10 predictor configuration, respectively. For each configuration, three boxplots illustrate the test error distributions: blue for the averaging of Linear Regression (LR) models, red for the averaging of Kernel Regression (KR) models, and violet for the averaging of both LR and KR models together. On the right side of each row, three bars represent the Mean Absolute Error (MAE) for each ensemble, colored in line with the respective boxplots, expressed in euro.

After analyzing standalone models and identifying the best performers, along with their strengths and weaknesses, we can now move to the main focus of this work: ensembles. We will follow the same structure as in the previous exposition, first discussing results for mean forecasting, followed by those for quantile forecasting. The division of datasets into training and test periods is identical to that in the prior section.

We will also follow a structured evaluation framework for the ensemble methods, based on fixed model ensembles. This begins with univariate ensembles, followed by bivariate, trivariate, and quadrivariate ensembles, concluding with the "All Models" ensemble that includes all possible predictor combinations, per each model. Additionally, the "Best 5" and "Best 10" ensembles contain only the top 5 and top 10 predictor combinations, respectively, for each forecasting task (Mean, Low, and High).

3.4.1 Mean Forecasting Results (Ensemble Models)

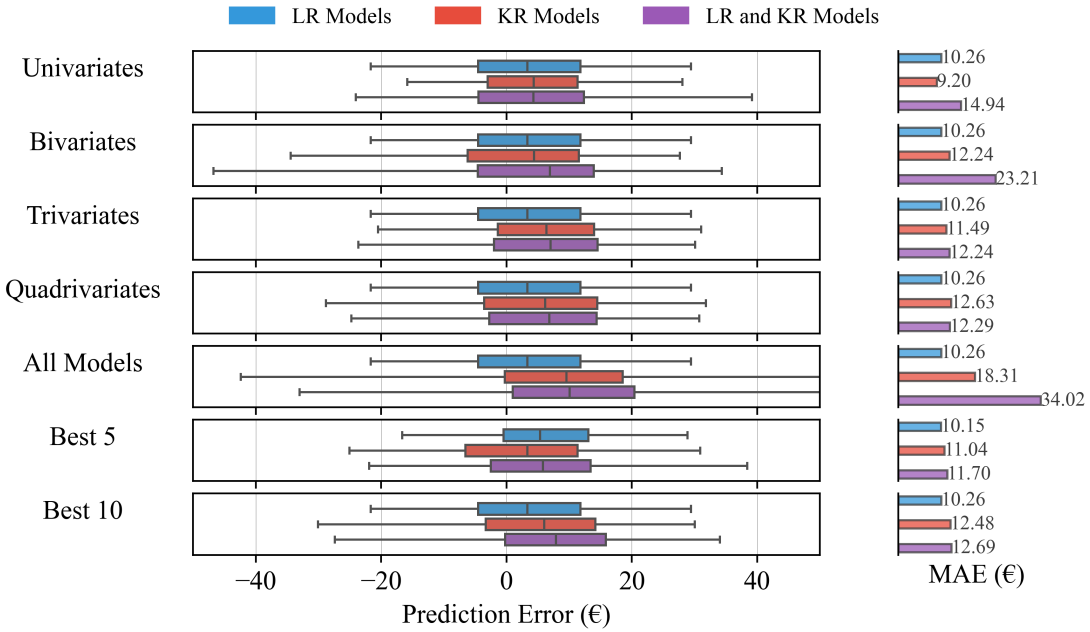

Fig. 3.12 **Error Distribution Comparison for Stacking Mean Regression Ensembles.**
This figure compares the error distribution for the Stacking Mean Regression (SMR) method, which averages the predictions from an ensemble of models using the least squares criterion. The layout is the same as in the previous figure.

Starting with the simplest MA approach, VMR, we recall that this method outputs estimates that are simply the arithmetic mean of the individual forecasts. Results are displayed in Figure 3.11. Each ensemble's performance is shown using three boxplots, which illustrate the error distribution for different regression types: blue for the averaging of LR, red for KR, and violet for the ensemble that combines both LR and KR. Similarly to the previous figures, each box represents the IQR of the corresponding data, while whiskers cover the central 95% of the distribution. To the right of each row, three bars indicate the MAE for each ensemble (color-coded to match the respective boxplots), in euros. Notably, this approach achieves reduced average errors in multiple instances compared to the performance of the best individual models. In nearly all cases, the ensemble forecasts are either better than, or at least as accurate as, the best individual forecasts, with the only exception being the ensemble of quadrivariate LRs (MAE = € 9.69). The strength of this method for mean forecasting lies not only in its ability to reduce the MAE but also in the symmetry it induces in the error distribution, effectively eliminating prediction outliers. This outcome practically realizes the intuitive principle that averaging multiple forecasts should lead to the cancelation of over- and under-estimation errors, resulting in a more stable and balanced prediction. The best overall result is achieved by the "Best 10" KR models, yielding an MAE of €7.51, which corresponds to an 11% reduction in MAE (i.e., a decrease of €0.93) compared to the best-performing individual model, which, as noted, is the KR model with predictors Price, Gas, and Thermal. Even more noteworthy is that similarly strong results are obtained by averaging all KR models without selection, yielding an MAE of €7.59. Another interesting observation is that including linear models in the VMR ensembles consistently worsens forecasts, whether they are used alone or in combination with nonlinear KR models. From the perspective of error distribution symmetry, the "Best 5" ensemble achieves the most balanced results, showing a median error near zero, regardless of whether LR, KR, or LR+KR models are used. Nevertheless, the average error appears generally higher. This result is consistent with the findings of [45], stating that simple averaging schemes (mean, median, etc.) tend to perform well in the presence of many individual models.

Moving on to the stacking approach, indicated here as SMR, the results are notably less favorable, see Figure 3.12. The layout, colors, and metrics are consistent with the previous figure. First and foremost, these outcomes serve as a reminder that a more advanced method—in this case, calculating averaging weights via LR—is not always

more efficient in practice than a simpler approach, such as the arithmetic mean in VMR. Compared to the previous results, the situation is reversed: in all cases, the distribution of errors shows a mean and variance higher than those of the best individual models, with particularly unfavorable results for the "All Models" ensemble. Additionally, even considering only the top-performing models (i.e., the "Best 5" and "Best 10" ensembles) fails to yield improvements. This is likely due to a limitation of LR: its low suitability for handling problems with many highly correlated variables, as is the case with the outputs of the various models in this context. These results align with findings from previous studies [45] and [46], which suggest that, for point EPF applications, regression-based MA approaches are generally suboptimal.

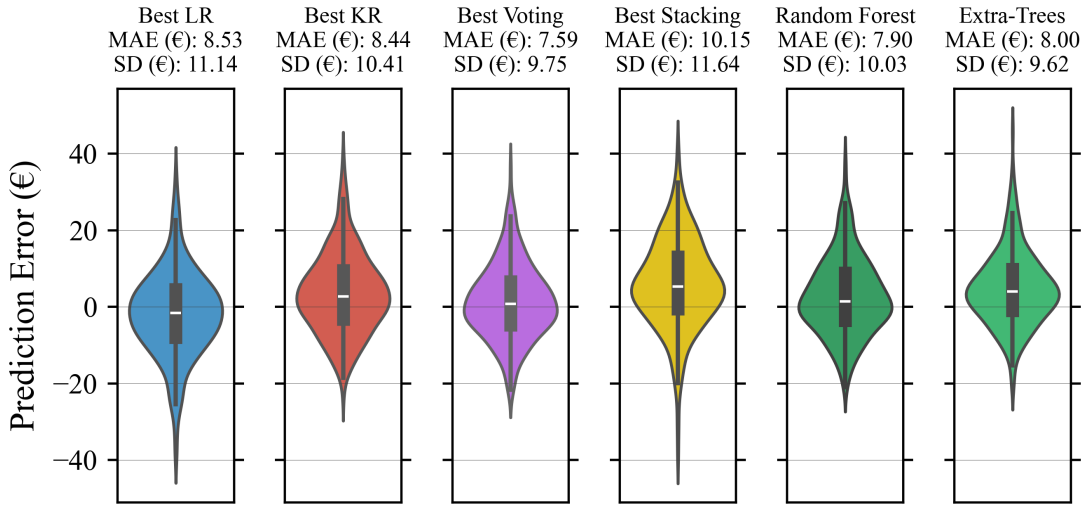

Fig. 3.13 **Overall Comparison of the Best Mean Forecasting Models**.
This figure presents a comparison of the top forecasting models' performance. The results show the overall error distribution across all the peakload hours of the test period, spanning May and June 2024, Monday to Friday. Each column corresponds to a different model, with the Mean Absolute Error (MAE) and the errors' Standard Deviation (SD), in euro, displayed on top, under the model's name. The color scheme is as follows: blue for the best Linear Regression (LR) model, red for the best Kernel Regression (KR) model, violet for the Voting regression (here, the average of all KR models), yellow for (linear) Stacking, dark green for Random Forest, and light green for Extra-Trees. Within each violin plot, a dark boxplot shows the interquartile range, with a white notch marking the median.

To summarize, we can first refer to Figure 3.13, which provides a comparison across the top-performing mean forecasting models. The figure displays the error distribution for each model, specifically across all peakload hours during the test period, covering May and June 2024, from Monday to Friday. Each column corresponds to a specific

model, with MAE and Standard Deviation (SD) of errors (in euros) indicated on top beside the model name. The color scheme will also be used in the following figures, where blue represents the best LR model, red represents the best KR model (see the previous Section 3.3), violet represents the VMR (average across all KR models), yellow denotes SMR, dark green represents RF, and light green represents ETR. Each violin plot illustrates the distribution's density, with a dark boxplot inside to mark the interquartile range and a white line highlighting the median. As noted previously, the best-performing solution overall is the MA approach via VMR, excelling both in minimizing mean error and in maintaining a balanced error distribution. Following this method are tree-based approaches, which perform well considering their relatively low requirements in terms of tuning. Both RF and ETR exhibit similar performance, with RF showing a slightly lower mean error (MAE of €7.90 and SD of 10.03), while ETR features a smaller variance (MAE of €8.00, SD of €9.62). All these methods, except for the SMR, achieve a 5-10% error reduction over the best standalone models.

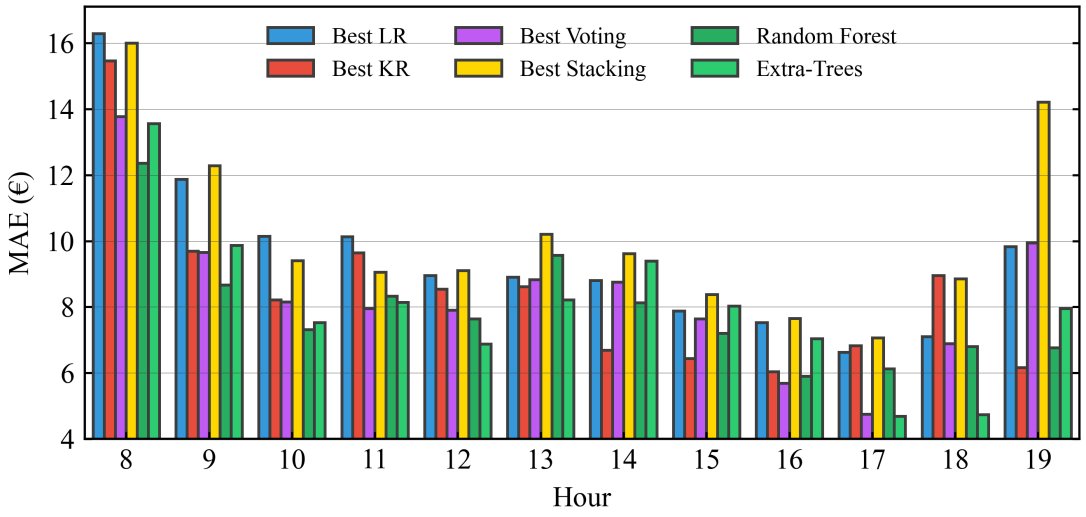

Fig. 3.14 **Analysis of Top Forecasting Models During Peakload Hours, Mean Regression**.
This barplot illustrates the hourly performance of the best forecasting models during peakload hours of the test period. The x-axis represents the hours of the day, while the y-axis shows the Mean Absolute Error (MAE), in euro. The comparison includes six models: the best-performing individual Linear Regression (LR), using "Gas+Solar" as inputs (blue), the top-performing individual Kernel Regression (KR), using "Price+Gas+Thermal" (red), the best Voting regression model (the average of all KR models, violet), Stacking (yellow), Random Forest (dark green), and Extra-Trees (light green). Both Random Forest and Extra-Trees achieve acceptable performance without requiring extensive model selection efforts.

Finally, a more detailed comparison of the hourly MAE for peakload prices is shown in Figure 3.14. Starting with individual models, LR performs best in the morning hours, but its accuracy decreases in the evening — a predictable outcome since it includes Solar as one of the predictors. KR shows the opposite trend, with higher accuracy in the central hours and towards the evening. Once again, VMR stands out as the top-performing model, with the lowest variability in MAE across different hours. This result underscores the competitiveness of simple arithmetic averaging in forecasting, as seen in various studies and competitions, including the M4 competition [65]. RF and ETR also show limited hourly variability, though with slightly higher MAEs (it is worth recalling that both these tree-based methods utilize all available predictors as inputs, achieving strong performance without requiring extensive model selection efforts). SMR performs particularly poorly at 8 AM and 7 PM (19, in the figure), corresponding to the first and last scheduling periods, respectively.

3.4.2 Quantile Forecasting Results (Ensemble Models)

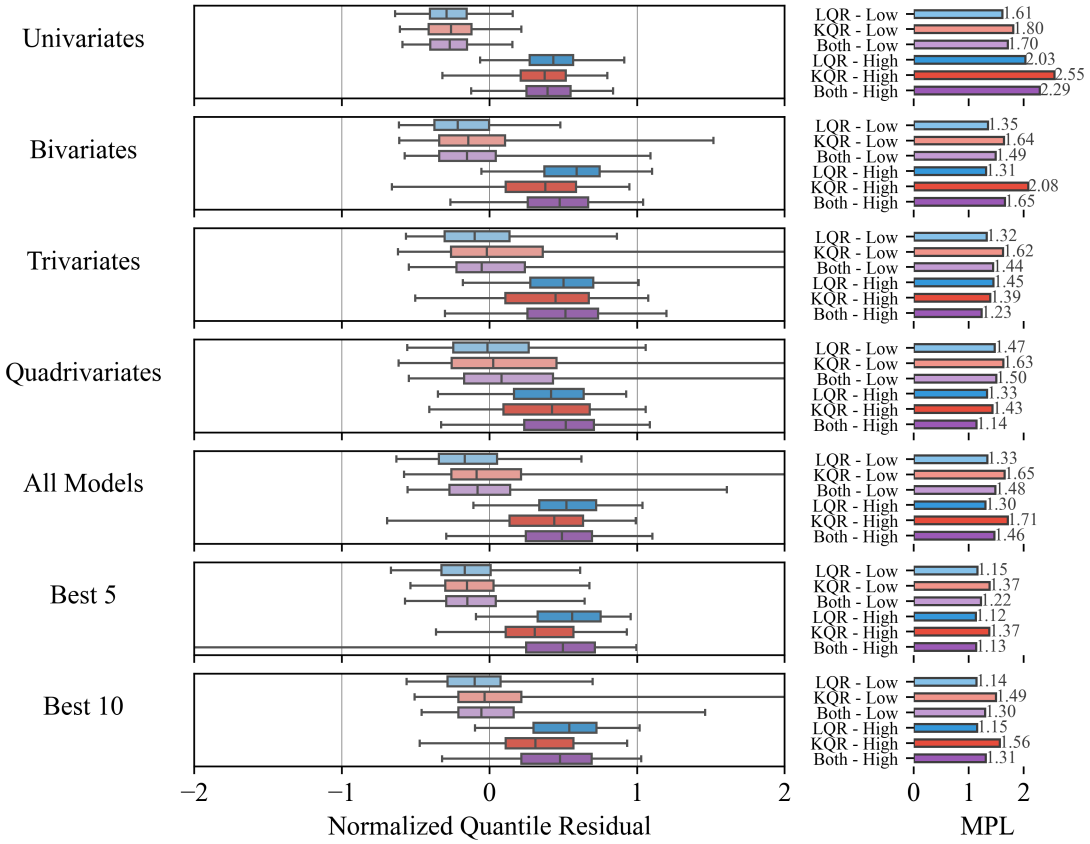

Fig. 3.15 **Error Distribution Comparison for Voting Quantile Regression.**
This figure compares the error distributions for the Voting Quantile Regression (VQR) approach. Each row represents different ensembles based on predictor combinations. In each, six boxplots illustrate the distribution of quantile residuals for the 0.05 (Low) and 0.95 (High) quantiles. Blue plots represent the averaging of Linear Quantile Regression (LQR) models, red represents ensembles of Kernel Quantile Regressions (KQR), and violet represents the combined LQR and KQR models. Lighter shades correspond to the 0.05 quantile, while darker shades represent the 0.95 quantile. To the right of each row, bars indicate the Mean Pinball Loss (MPL) for each ensemble, with colors matching the boxplots.

Let us now examine the results for the more current research challenge of MA methods for quantile prediction. As in previous sections, we focus on two distinct tasks: forecasting a lower quantile, termed "Low," which corresponds to the $\alpha = 0.05$ quantile level, and an upper quantile, termed "High," corresponding to the $\alpha = 0.95$ level. The MA method using simple arithmetic mean, referred to here as VQR, appears less robust for quantile prediction than for mean forecasting, although excellent results are achievable with

careful model selection within the ensemble. In contrast, stacking-based methods, here called SQR, demonstrate exceptional robustness across both tasks, achieving the lowest error levels in certain cases.

Starting with the VQR approach, we analyze the results shown in Figure 3.15. This figure presents a comparison of the error distribution for VQR-based ensembles. Each row corresponds to a different ensemble type, progressing from univariate to bivariate, trivariate, and quadrivariate ensembles, culminating in the "All Models" ensemble, which includes all available predictor combinations. Additionally, two targeted ensembles, "Best 5" and "Best 10," feature only the models formed from the top five and top ten predictor combinations, respectively. This layout is similar to those used for mean regression results, for clarity, but has one substantial difference in that it presents two series of residual distributions: one per each predicted quantile. Each row in the figure contains six boxplots displaying the distribution of quantile residuals for the 0.05 (Low) and 0.95 (High) quantiles. The color scheme mirrors previous figures: blue represents the averaging of LR models, red represents KR models, and violet denotes the ensemble combining both LR and KR models. Lighter shades denote the 0.05 quantile, while darker shades correspond to the 0.95 quantile. To the right of each row, a set of bars indicates the MPL for each ensemble, color-coded to match the corresponding boxplots. In contrast to mean forecasting, the VQR approach does not deliver a pronounced advantage in terms of accuracy, unless individual predictors are carefully selected. This indicates that predictor selection is essential to unlocking the potential of VQR, as was suggested in both [44] and [20]. Starting with Low (in lighter shades), it is essential to note that the top individual models, both LQR and KQR, already achieved strong performance, with MPLs of 1.21 and 1.25, respectively. An improvement can be achieved through averaging the linear model ensembles, particularly "Best 5" and "Best 10", which yield very similar MPLs of 1.15 and 1.14, respectively. If we consider High, the gains are similar. Compared to the best-performing individual model with the "Price + Gas + Thermal" predictor combination, which had an MPL of 1.19 for the LR model, the "Best 5 LR" ensemble achieves an MPL of 1.12, while "Best 10" reaches 1.15. The "Best 5 LR" result, in particular, reflects a 6% improvement. Notably, unlike mean forecasting — where KR-based averaging generally excelled — it is the LR-based ensembles that perform best for quantile forecasting.

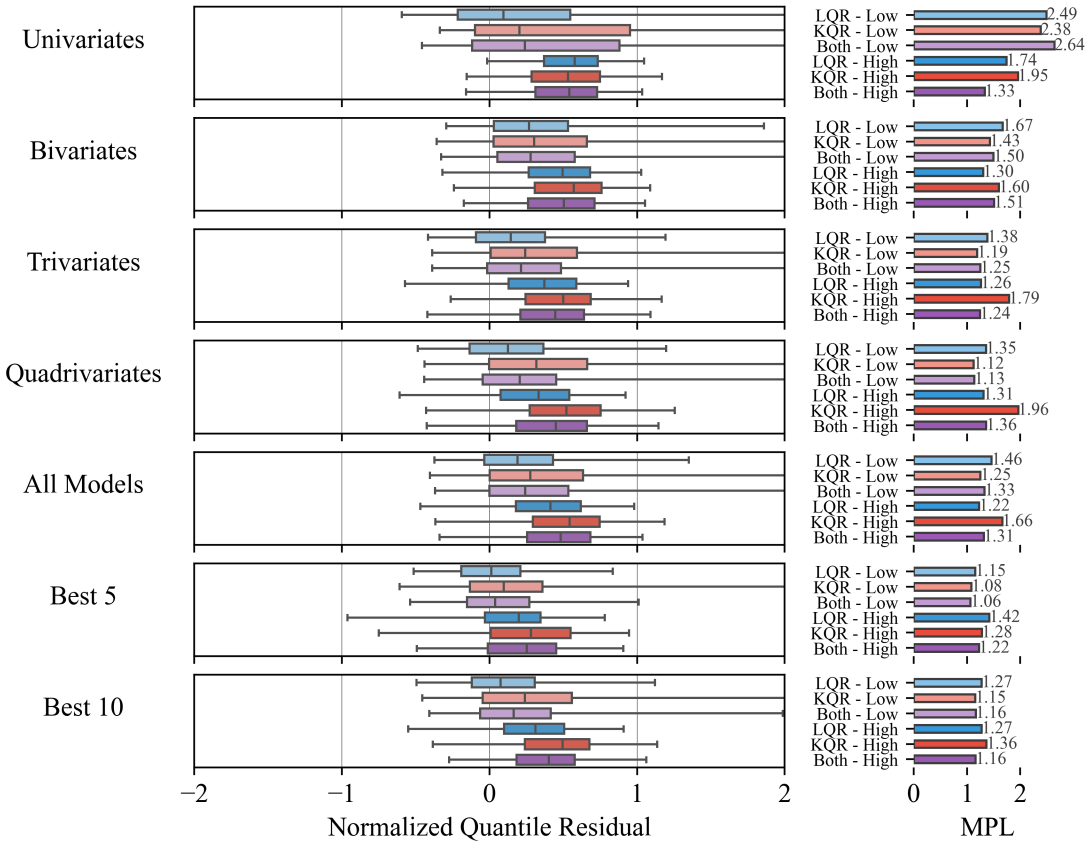

Fig. 3.16 **Error Distribution Comparison for Stacking Quantile Regression.**
This figure illustrates the error distribution for Stacking Quantile Regression (SQR) using Factor Quantile Regression Averaging (FQRA) [18]. In this method, the individual model outputs are first used as inputs to a Principal Component Analysis (PCA). The PCA reduces the input dimensionality by projecting onto a single principal component. As in the previous figure, each row represents an ensemble based on different predictor combinations. For each ensemble, six boxplots display the error distribution for the 0.05 (Low) and 0.95 (High) quantiles. Blue corresponds to ensembles of Linear Quantile Regressions (LQR), red to Kernel Quantile Regressions (KQR), and violet to the combination of both LQR and KQR models, with lighter shades for the 0.05 quantile and darker shades for the 0.95 quantile. On the right side of each row, colored bars indicate the Mean Pinball Loss (MPL), and correspond to the colors of the boxplots.

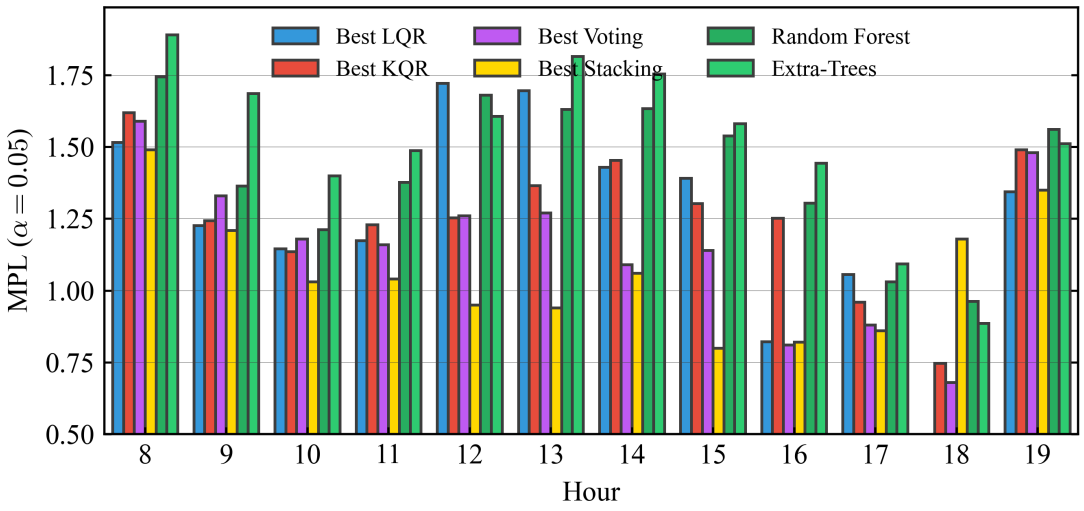

Fig. 3.17 **Analysis of Top Forecasting Models During Peakload Hours,** $\alpha = 0.05$.
This barplot presents the hourly performance of the best forecasting models during peakload hours (8 AM to 8 PM) from Tuesday to Thursday in May and June 2024, focused on the 0.05 quantile regression task. The x-axis represents the hours of the day, while the y-axis shows the Mean Pinball Loss (MPL). The comparison includes the top individual Linear Quantile Regression (LQR) model using "Price+Gas+Load" as inputs (light blue), the best Kernel Quantile Regression (KQR) model using "Price+Gas" (light red), the best Voting regression model (the average of all LQRs, violet), Random Forest (dark green), Extra-Trees (light green), and the Stacking model (yellow), using the quadrivariates KR ensemble.

We can now move to SQR, where, as a remark, we use FQRA as the aggregation approach. In FQRA, outputs from individual models first undergo PCA, which reduces dimensionality by projecting predictions onto a single principal component PC. This PC then serves as the single predictor in a univariate LQR model. All the detail is given in the last part of Section 3.2. Results are visible in Figure 3.16. We employ the same layout as in the VQR case: for each ensemble configuration, six boxplots show error distributions for the 0.05 and 0.95 quantiles, with the same coloring scheme as above. To the right, colored bars denote the MPL. For the Low quantile, the lowest loss is achieved by stacking the forecasts of the "Best 10 LR+KR" ensemble, with an MPL of 1.06, representing a 12% reduction compared to the best-performing individual model. It is important to stress that, for this task, the SQR approach reached the overall lowest loss, emerging as the best method. Conversely, for the High quantile, the VQR method performs slightly better than the SQR. Here, the lowest MPL for stacking is achieved by the "Best 5LR+KR" ensemble, with a value of 1.16 — representing a slight 4% reduction compared to the top individual model. Interestingly, the best results for SQR consistently

arise when combining both linear and nonlinear models, whereas for VQR, linear models tended to outperform others in quantile predictions. This underscores SQR's strength in leveraging diverse model types within an ensemble for improved quantile accuracy. However, even with this approach, selecting the right models for the ensemble is crucial. Notably, the "All Models" ensembles perform at the same level or worse than the best individual model.

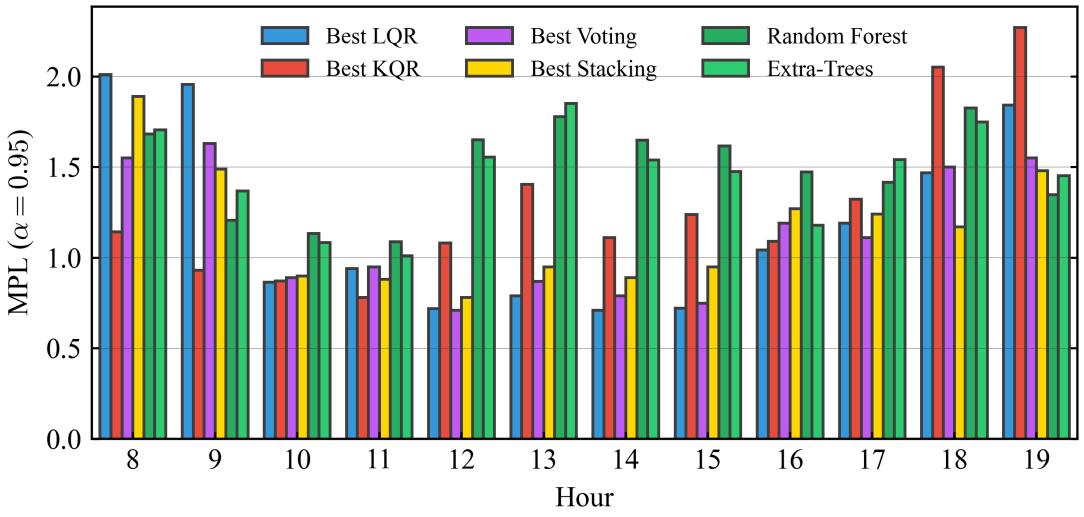

Fig. 3.18 **Analysis of Top Forecasting Models During Peakload Hours,** $\alpha = 0.95$.
This barplot presents the hourly performance of the best forecasting models during peakload hours (8 AM to 8 PM) from Tuesday to Thursday in May and June 2024, focused on the 0.95 quantile regression task. The layout is the same as in the previous figure.

For a more granular perspective, we can examine the hourly MPL details in Figure 3.17 and Figure 3.18, which break down the MPL values for each peakload hour, as done previously for the mean forecasts, specifically for the Low and High quantiles, respectively. Broadly speaking, both VQR (Voting) and SQR (Stacking) perform similarly overall, with both approaches successfully mitigating certain spikes in loss observed with individual LR or KR models. For the Low quantile, VQR tends to perform better in the morning but underperforms in the afternoon, with the opposite pattern seen in SQR. However, this distinction does not apply to the High quantile, where performance differences between VQR and SQR are less pronounced. Notably, tree-based models perform poorly in both quantile forecasting tasks, despite being highly effective for mean forecasting.

Predictor Combination	Model	MAE (€)	MAPE (%)	RMSE (€)
Price	LR	10.59	11.0	13.29
	KR	10.7	11.2	13.35
Gas	LR	12.89	13.5	15.83
	KR	14.76	15.6	17.79
Load	LR	13.59	14.5	17.17
	KR	14.35	15.5	18.05
Solar	LR	13.63	13.2	17.0
	KR	12.55	12.9	15.85
Thermal	LR	10.49	9.9	13.14
	KR	9.43	9.1	12.34
Price+Gas	LR	10.71	11.2	12.95
	KR	11.98	12.7	14.36
Price+Load	LR	10.24	10.5	13.34
	KR	10.64	11.1	13.35
Price+Solar	LR	11.05	10.8	14.63
	KR	10.41	10.6	13.31
Price+Thermal	LR	11.5	11.2	15.05
	KR	9.48	9.5	12.58
Gas+Load	LR	12.74	13.2	15.58
	KR	13.73	14.2	16.75
Gas+Solar	LR	9.13	9.2	11.86
	KR	10.87	11.1	13.79
Gas+Thermal	LR	9.44	9.0	12.28
	KR	9.16	8.9	11.89
Load+Solar	LR	13.47	12.6	18.61
	KR	12.92	13.4	16.45
Load+Thermal	LR	16.43	15.7	19.31
	KR	9.75	9.5	12.81
Solar+Thermal	LR	13.33	12.6	15.97
	KR	10.94	10.7	13.62
Price+Gas+Load	LR	10.67	11.1	13.15
	KR	11.83	12.4	14.23
Price+Gas+Solar	LR	9.38	9.3	12.18
	KR	10.17	10.5	12.71
Price+Gas+Thermal	LR	9.22	8.8	12.2
	KR	**8.46**	**8.2**	**11.2**
Price+Load+Solar	LR	10.69	10.4	14.45
	KR	10.11	10.3	12.79
Price+Load+Thermal	LR	12.76	12.4	16.2
	KR	10.6	10.5	13.98
Price+Solar+Thermal	LR	12.05	11.6	16.1
	KR	10.42	10.1	13.55
Gas+Load+Solar	LR	9.41	9.4	12.3
	KR	9.81	9.9	12.4
Gas+Load+Thermal	LR	9.77	9.4	12.63
	KR	10.32	10.2	13.08
Gas+Solar+Thermal	LR	9.97	9.4	13.1
	KR	9.78	9.6	12.41
Load+Solar+Thermal	LR	19.07	17.9	24.78
	KR	12.3	12.1	15.63
Price+Gas+Load+Solar	LR	9.49	9.4	12.65
	KR	9.92	10.0	12.56
Price+Gas+Load+Thermal	LR	9.79	9.4	12.81
	KR	10.18	10.0	13.26
Price+Gas+Solar+Thermal	LR	10.07	9.5	13.63
	KR	9.84	9.4	12.73
Price+Load+Solar+Thermal	LR	13.11	12.6	17.1
	KR	11.86	11.8	14.88
Gas+Load+Solar+Thermal	LR	10.75	10.2	14.18
	KR	10.42	10.2	13.23
Price+Gas+Load+Solar+Thermal	LR	10.45	9.9	14.2
	KR	11.12	10.8	14.7

Table 3.1 **Performance Summary for Linear and Kernel Mean Regression Models.**
The table reports Mean Absolute Error (MAE), expressed in euros, Mean Absolute Percentage Error (MAPE), and Root Mean Squared Error (RMSE), in euros, for each LR and KR model considered in the analysis, with best values in bold.

Predictor Combination	Model	MPL Low (0.05)	MPL High (0.95)
Price	KQR	1.53	2.19
	LQR	1.31	1.71
Gas	KQR	1.29	2.9
	LQR	1.34	2.04
Load	KQR	2.02	3.63
	LQR	1.96	3.73
Solar	KQR	2.28	3.2
	LQR	2.71	5.7
Thermal	KQR	2.36	2.01
	LQR	2.1	1.62
Price+Gas	KQR	1.25	2.24
	LQR	1.25	1.66
Price+Load	KQR	1.55	2.36
	LQR	1.21	1.72
Price+Solar	KQR	1.85	2.55
	LQR	1.99	3.58
Price+Thermal	KQR	2.01	1.54
	LQR	1.56	1.25
Gas+Load	KQR	1.52	2.72
	LQR	1.4	1.94
Gas+Solar	KQR	1.8	2.93
	LQR	1.8	3.24
Gas+Thermal	KQR	1.65	1.55
	LQR	1.28	1.21
Load+Solar	KQR	1.97	3.83
	LQR	2.06	4.67
Load+Thermal	KQR	2.02	2.69
	LQR	1.86	4.77
Solar+Thermal	KQR	2.36	3.04
	LQR	2.63	3.35
Price+Gas+Load	KQR	1.29	2.17
	LQR	1.21	1.64
Price+Gas+Solar	KQR	1.77	2.87
	LQR	1.74	3.18
Price+Gas+Thermal	KQR	1.72	1.29
	LQR	1.26	1.19
Price+Load+Solar	KQR	1.74	3.41
	LQR	1.76	2.9
Price+Load+Thermal	KQR	1.87	1.45
	LQR	1.55	1.54
Price+Solar+Thermal	KQR	2.11	2.65
	LQR	2.16	2.55
Gas+Load+Solar	KQR	1.84	2.96
	LQR	1.56	2.42
Gas+Load+Thermal	KQR	1.7	1.56
	LQR	1.27	1.49
Gas+Solar+Thermal	KQR	2.07	2.06
	LQR	1.71	1.52
Load+Solar+Thermal	KQR	1.98	3.06
	LQR	2.0	9.57
Price+Gas+Load+Solar	KQR	1.65	3.15
	LQR	1.57	2.82
Price+Gas+Load+Thermal	KQR	1.71	1.27
	LQR	1.26	1.68
Price+Gas+Solar+Thermal	KQR	1.9	2.36
	LQR	1.84	1.99
Price+Load+Solar+Thermal	KQR	1.83	2.91
	LQR	2.12	2.51
Gas+Load+Solar+Thermal	KQR	1.84	2.75
	LQR	1.48	2.21
Price+Gas+Load+Solar+Thermal	KQR	1.85	2.9
	LQR	1.75	2.12

Table 3.2 **Performance Summary for Linear and Kernel Quantile Regression Individual Models.** *The table summarizes the Mean Pinball Loss (MPL) values for Kernel Quantile Regression (KQR) and Linear Quantile Regression (LQR) models across different predictor combinations.*

Chapter 4

Conclusion

The problem of accurately predicting electricity prices is widely acknowledged as a complex issue. Over the past three decades, research spanning electrical engineering and statistics has identified the key factors contributing to this complexity. In our small way, this Thesis has attempted to retrace some of these challenges and present possible solutions inspired by the most recent research developments. We do not claim to have produced a comprehensive treatment on Electricity Price Forecasting (EPF)— a definitive work will likely never exist, given the incredibly dynamic nature of this vast field.
In fact, it has grown increasingly intricate over the years, as the fundamental and, consequently, statistical characteristics of electricity prices continue to evolve. In earlier periods, when power generation relied almost entirely on thermal sources, volatility—defined as the variance over short time intervals—was significantly lower. Although even then there was a degree of weather dependency—for example, prices were (and still are) typically higher on very cold winter days due to increased heating demand—it was nothing compared to what has been observed over the past fifteen years, largely due to the high penetration of Non-Programmable Renewable Energy Sources (NPRES). Thus, while in the pre-renewable era weather conditions influenced electricity prices primarily through their impact on demand, they now affect both demand and supply—an effect that has become increasingly pronounced due to the exponential growth of NPRES across Europe. Taking photovoltaic energy as a particularly illustrative example, it is well known that in regions where it accounts for a significant share of installed capacity, it often generates more electricity than the actual demand during the central hours of

sunny days. This surplus leads to a depression in electricity prices—sometimes with unpredictable outcomes. Even less predictable is the generation capacity of onshore wind power, which is subject to greater variability. This leads to sudden fluctuations in electricity supply, destabilizing market prices. To all this, we must add—last but not least—the impact of climate change, which is making accurate weather forecasting increasingly difficult: this problem alone could easily be the subject of a dedicated PhD dissertation.

Before the advent of renewable energy, almost all of the research in EPF focused on point forecasting—that is, predicting a single, most likely value for future electricity prices (e.g. "Tomorrow's price will be 100 €/MWh"). However, starting roughly a decade ago, the scientific community working in this field began to recognize that, rather than competing to marginally improve the accuracy of point forecasts—which would likely remain unsatisfactory due to the market's inherent unpredictability—it was more effective to shift towards probabilistic forecasting. This approach involves estimating a full distribution of possible future prices, thereby capturing the uncertainty and variability of the market in a more informative and realistic way (e.g. "Tomorrow's price will fluctuate between 90 €/MWh and 110 €/MWh with 95% probability"). We have attempted to illustrate this point in Chapter 2, where we emphasize what are arguably the most fundamental aspects of EPF: data analysis and data understanding. For this reason, we chose to limit ourselves to presenting results obtained using basic linear and nonlinear regression techniques, applied to both point and probabilistic forecasting. We discussed the quantitative variables available to a forecaster aiming to predict electricity prices in the Italian market—how they correlate with price movements, why certain variables matter, and which ones can be discarded. The extensive visual analyses and results show that, once a thorough preliminary data analysis is carried out, the choice of statistical algorithm tends to impact the forecasts only marginally. After all, the vast literature on neural networks and Machine Learning (ML) algorithms in EPF—when authored by credible researchers—often highlights their only modest superiority over simple linear regression models.

However, the complexities of EPF today go far beyond the challenges posed by NPRES and the shift toward probabilistic approaches. What led us to explore this topic in greater depth were recent events and contingencies—most notably the surge in inflation and the Russia-Ukraine war, which have severely destabilized European electricity markets. As a result, EPF became so complex that it proved nearly unmanageable for certain Italian

energy companies, some of which faced significant financial losses and, in extreme cases, came close to bankruptcy. It is worth recalling that similar dynamics have already occurred in deregulated energy markets; the first of these—the Californian market—was also the first to collapse disastrously. There would be much to say about the (de)merits of privatization and deregulation, but that discussion would likely require a third thesis, in addition to this one and the aforementioned work on weather forecasting.

Returning to the current state of the market, it is clear that we are facing increasingly irregular and volatile price series shaped by broad structural and geopolitical transformations. In Chapter 3, we explored the potential of Model Averaging (MA) as a flexible and data-driven solution to improving forecasting accuracy. Consistent with prior literature, we found that simple aggregation methods, such as the arithmetic mean of forecasts from diverse models, significantly reduce prediction errors. Interestingly, these simple approaches often outperformed more sophisticated ensemble techniques, such as regression-based stacking. In the probabilistic domain, however, the task proved more challenging. While basic averaging still offered benefits, more advanced methods were required in order to achieve optimal performance. Notably, tree-based ML methods (such as Random Forest), which performed well in point forecasting, failed in the probabilistic setting. This highlights the limitations of off-the-shelf ML applications and suggests avoiding a "feed input and check results" approach.

We chose to focus only on those aspects we considered most relevant for understanding the contemporary EPF problem, always aiming to produce a text that remains accessible even to non-experts. To that end, we supplemented our discussion of methods and results with visual analyses wherever possible. This approach has value not only from a didactic perspective but also from a professional one: when developing business-oriented EPF solutions, model interpretability is a crucial requirement. Still, also for this same reason, we deliberately avoided the use of more advanced ML techniques.

The question of real-world applicability naturally points toward future developments that we could not explore in detail here. Among the most promising directions is the so-called financial evaluation of forecasts. This concept refers to assessing the quality of forecasting models not only through statistical metrics such as the Mean Absolute Error (MAE), but through their actual financial impact on trading strategies. A model that performs slightly worse statistically may still yield better economic outcomes—such as higher profit margins or reduced risk exposure—when applied in a market context. To perform this kind of evaluation, it is common to develop energy trading simulators,

where decision-making modules are fed with different forecasting methods, and the final outcomes are then assessed in terms of financial returns. As for alternative ways of evaluating predictions—specifically probabilistic forecasts—there exists an entire class of metrics beyond the commonly used pinball loss. Among these, interval-based evaluation metrics are particularly promising and may be explored in future work. Of special interest are those designed to assess the robustness of forecasts, i.e., their ability to maintain reliability under changing conditions or across different market regimes.
Another industrially relevant direction would be the development of "daily retraining" counterparts for the implemented models. This refers to the practice of updating model parameters on a daily basis as new data becomes available, rather than relying on a static model trained once and used indefinitely. Similarly, a topic that remains unexplored in this work—but certainly worthy of further investigation—is the family of transformations applied to both input and output variables. Despite their potential to improve model performance, stability, or interpretability, there is still no consensus in the current literature on which transformations are most appropriate in the context of EPF. Logarithmic scaling, normalization, and differencing are just a few examples of techniques whose effects can vary significantly depending on the specific forecasting task and market conditions.
In conclusion, we can state that EPF stands as a representative case of those application domains where computational intelligence can be effectively employed. The degrees of freedom inherent to the problem are virtually infinite, which precludes the possibility of constructing a fully exhaustive treatment—something that academic research often aims for but cannot always achieve.

References

[1] Thomas Piketty and Arthur Goldhammer. *Capital in the Twenty-First Century*. Harvard University Press, 2014.

[2] Italian National Institute of Statistics (Italian: Istituto nazionale di statistica; Istat). Indici dei Prezzi al Consumo. Aspetti Generali e Metodologia di Rilevazione. Edizione 2022. https://www.istat.it/produzione-editoriale/il-mercato-del-lavoro-2020-una-lettura-integrata/, 2022.

[3] Mario Draghi. The future of European competitiveness: Report by Mario Draghi. https://commission.europa.eu/topics/strengthening-european-competitiveness/eu-competitiveness-looking-ahead_en, 2024.

[4] European Commission. European Social Fund Plus. https://european-social-fund-plus.ec.europa.eu/en/react-eu.

[5] Energy & Strategy (Politecnico di Milano). Electricity Market Report 2023. https://www.energystrategy.it/es-download/, 2023. Accessed: 2025-04-25.

[6]

[7] International Energy Agency. World energy outlook 2020, 2020. Licence: CC BY 4.0.

[8] Paul Joskow. California's Electricity Crisis. *Oxford Review of Economic Policy*, 17(3):365–388, 2001.

[9] Roger Koenker. *Quantile Regression*. Econometric Society Monographs. Cambridge University Press, 2005.

[10] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., New York, NY, USA, 2001.

[11] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, 1 edition, 2007.

[12] Tao Hong, Pierre Pinson, Yi Wang, Rafał Weron, Dazhi Yang, and Hamidreza Zareipour. Energy Forecasting: A Review and Outlook. *IEEE Open Access Journal of Power and Energy*, 7:376–388, 2020.

[13] Silvia Golia, Luigi Grossi, and Matteo Pelagatti. Machine Learning Models and Intra-Daily Market Information for the Prediction of Italian Electricity Prices. *Forecasting*, 5(1):81–101, 2023.

[14] Piano Nazionale Integrato per l'Energia e il Clima (the Italian national energy and climate plan. https://commission.europa.eu/energy-climate-change-environment/implementation-eu-countries/energy-and-climate-governance-and-reporting/national-energy-and-climate-plans_en, 2023. Accessed: 2025-04-25.

[15] Energy & Strategy (Politecnico di Milano). Renewable Energy Report 2024. https://www.energystrategy.it/es-download/, 2024. Accessed: 2025-04-25.

[16] Rafal Weron. *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*. Number hsbook0601 in HSC Books. Hugo Steinhaus Center, Wroclaw University of Technology, 2006.

[17] Rafal Weron. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, 30(4):1030–1081, 2014.

[18] Bartosz Uniejewski and Rafał Weron. Regularized quantile regression averaging for probabilistic electricity price forecasting. *Energy Economics*, 95:105121, 2021.

[19] Ties Van Der Heijden, Jesus Lago, Peter Palensky, and Edo Abraham. Electricity Price Forecasting in European Day Ahead Markets: A Greedy Consideration of Market Integration. *IEEE Access*, 9:119954–119966, 2021.

[20] Jakub Nowotarski and Rafał Weron. Recent advances in electricity price forecasting: A review of probabilistic forecasting. *Renewable and Sustainable Energy Reviews*, 81:1548–1568, 2018.

[21] Xin Lu, Jing Qiu, Gang Lei, and Jianguo Zhu. An Interval Prediction Method for Day-Ahead Electricity Price in Wholesale Market Considering Weather Factors. *IEEE Transactions on Power Systems*, pages 1–11, 2023.

[22] Chenxu Zhang, Yong Fu, and Lin Gong. Short-Term Electricity Price Forecast Using Frequency Analysis and Price Spikes Oversampling. *IEEE Transactions on Power Systems*, 38(5):4739–4751, 2023.

[23] Guangchun Ruan, Jianxiao Wang, Haiwang Zhong, Qing Xia, and Chongqing Kang. Improving Sample Efficiency of Deep Learning Models in Electricity Market. *IEEE Transactions on Power Systems*, 38(5):4761–4773, 2023.

[24] Mahmood Hosseini Imani, Ettore Bompard, Pietro Colella, and Tao Huang. Forecasting Electricity Price in Different Time Horizons: An Application to the Italian Electricity Market. *IEEE Transactions on Industry Applications*, 57(6):5726–5736, 2021.

[25] Peru Muniain and Florian Ziel. Probabilistic forecasting in day-ahead electricity markets: Simulating peak and off-peak prices. *International Journal of Forecasting*, 36(4):1193–1210, 2020.

[26] Aitor Ciarreta, Blanca Martinez, and Shahriyar Nasirov. Forecasting electricity prices using bid data. *International Journal of Forecasting*, 39(3):1253–1271, 2023.

[27] Angelica Gianfreda, Francesco Ravazzolo, and Luca Rossini. Comparing the forecasting performances of linear models for electricity prices with high RES penetration. *International Journal of Forecasting*, 36(3):974–986, 2020.

[28] Katarzyna Maciejowska and Rafal Weron. Short- and mid-term forecasting of baseload electricity prices in the UK: The impact of intra-day price relationships and market fundamentals. HSC Research Reports HSC/15/04, Hugo Steinhaus Center, Wroclaw University of Technology, 1 2015.

[29] Konstantin Dragomiretskiy and Dominique Zosso. Variational mode decomposition. *IEEE Transactions on Signal Processing*, 62(3):531–544, 2014.

[30] James W. Taylor. Evaluating quantile-bounded and expectile-bounded interval forecasts. *International Journal of Forecasting*, 37(2):800–811, 2021.

[31] Pavlos S. Georgilakis. Market Clearing Price Forecasting in Deregulated Electricity Markets Using Adaptively Trained Neural Networks. In Grigoris Antoniou, George Potamias, Costas Spyropoulos, and Dimitris Plexousakis, editors, *Advances in Artificial Intelligence*, pages 56–66, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

[32] George Casella and Roger L Berger. *Statistical inference*, volume 2. Duxbury Pacific Grove, CA, 2002.

[33] Keming Yu and Jin Zhang and. A three-parameter asymmetric laplace distribution and its extension. *Communications in Statistics - Theory and Methods*, 34(9-10):1867–1879, 2005.

[34] E. A. Nadaraya. On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142, 1964.

[35] G. S. Watson. Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics; Series A*, 1964.

[36] Qi Li and Jeffrey Scott Racine. *Nonparametric Econometrics: Theory and Practice*, volume 1 of *Economics Books*. Princeton University Press, 2006.

[37] Clifford M. Hurvich, Jeffrey S. Simonoff, and Chih-Ling Tsai. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(2):271–293, 1998.

[38] Emanuel Parzen. On Estimation of a Probability Density Function and Mode. *The Annals of Mathematical Statistics*, 33(3):1065 – 1076, 1962.

[39] D.W. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization*. A Wiley-interscience publication. Wiley, 1992.

[40] E. Çınlar. *Probability and Stochastics*. Graduate Texts in Mathematics. Springer New York, 2011.

[41] Charles J. Stone. Consistent Nonparametric Regression. *The Annals of Statistics*, 5(4):595 – 620, 1977.

[42] Gestore dei Mercati Energetici. https://www.mercatoelettrico.org. Accessed: 2025-04-25.

[43] Terna Download Center. https://www.terna.it/en/electric-system/transparency-report/download-center. Accessed: 2025-04-25.

[44] Xiaoqian Wang, Rob J. Hyndman, Feng Li, and Yanfei Kang. Forecast combinations: An over 50-year review. *International Journal of Forecasting*, 39(4):1518–1547, 2023.

[45] Mary E. Thomson, Andrew C. Pollock, Dilek Önkal, and M. Sinan Gönül. Combining forecasts: Performance and coherence. *International Journal of Forecasting*, 35(2):474–484, 2019.

[46] Jakub Nowotarski, Eran Raviv, Stefan Trück, and Rafał Weron. An empirical comparison of alternative schemes for combining electricity spot price forecasts. *Energy Economics*, 46:395–412, 2014.

[47] Nadeela Bibi, Ismail Shah, Abdelaziz Alsubie, Sajid Ali, and Showkat Ahmad Lone. Electricity Spot Prices Forecasting Based on Ensemble Learning. *IEEE Access*, 9:150984–150992, 2021.

[48] Jesus Lago, Fjo De Ridder, and Bart De Schutter. Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. *Applied Energy*, 221:386–405, 2018.

[49] Ehab E. Elattar, Salah K. Elsayed, and Tamer Ahmed Farrag. Hybrid Local General Regression Neural Network and Harmony Search Algorithm for Electricity Price Forecasting. *IEEE Access*, 9:2044–2054, 2021.

[50] Muhammad Ahsan Zamee, Yeongsang Lee, and Dongjun Won. Self-Supervised Adaptive Learning Algorithm for Multi-Horizon Electricity Price Forecasting. *IEEE Access*, 12:54913–54933, 2024.

[51] Cameron Cornell, Nam Trong Dinh, and S. Ali Pourmousavi. A probabilistic forecast methodology for volatile electricity prices in the Australian National Electricity Market. *International Journal of Forecasting*, 40(4):1421–1437, 2024.

[52] Peter Bloomfield and William Steiger. Least absolute deviations curve-fitting. *SIAM Journal on Scientific and Statistical Computing*, 1(2):290–301, 1980.

[53] R.J. Hyndman and G. Athanasopoulos. *Forecasting: principles and practice*. OTexts, 2014.

[54] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[55] John A. Nelder and R. Mead. A Simplex Method for Function Minimization. *Comput. J.*, 7(4):308–313, 1965.

[56] Yi Wang, Ning Zhang, Yushi Tan, Tao Hong, Daniel S. Kirschen, and Chongqing Kang. Combining Probabilistic Load Forecasts. *IEEE Transactions on Smart Grid*, 10(4):3664–3674, 2019.

[57] Roger Koenker and Kevin F. Hallock. Quantile Regression. *Journal of Economic Perspectives*, 15(4):143–156, December 2001.

[58] Stephen Portnoy and Roger Koenker. The Gaussian hare and the Laplacian tortoise: computability of squared-error versus absolute-error estimators. *Statistical Science*, 12(4):279 – 300, 1997.

[59] David H. Wolpert. Stacked generalization. *Neural Networks*, 5(2):241–259, 1992.

[60] Pierre Geurts, Damien Ernst, and Louis Wehenkel. Extremely randomized trees. *Mach. Learn.*, 63(1):3–42, April 2006.

[61] Nicolai Meinshausen. Quantile Regression Forests. *Journal of Machine Learning Research*, 7(35):983–999, 2006.

[62] Tao Hong, Pierre Pinson, Shu Fan, Hamidreza Zareipour, Alberto Troccoli, and Rob J. Hyndman. Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. *International Journal of Forecasting*, 32(3):896–913, 2016.

[63] Raffaele Sgarlato and Florian Ziel. The Role of Weather Predictions in Electricity Price Forecasting Beyond the Day-Ahead Horizon. *IEEE Transactions on Power Systems*, 38(3):2500–2511, 2023.

[64] Francesco Moraglio and Carlo S. Ragusa. Day-Ahead Electricity Price Forecasting in the Contemporary Italian Market. *IEEE Access*, 12:72062–72078, 2024.

[65] Spyros Makridakis, Evangelos Spiliotis, and Vassilios Assimakopoulos. The M4 Competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, 36(1):54–74, 2020. M4 Competition.