CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Forecasting and trading optimisation in the day-ahead and balancing market
Authors	O'Connor, Ciaran
Publication date	2025
Original Citation	O'Connor, C. 2025. Forecasting and trading optimisation in the day-ahead and balancing market. PhD Thesis, University College Cork.
Type of publication	Doctoral thesis
Rights	© 2025, Ciaran O'Connor. - https://creativecommons.org/licenses/by/4.0/
Download date	2026-11-15 02:35:05
Item downloaded from	https://hdl.handle.net/10468/17912



UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Forecasting and Trading Optimisation in the Day-Ahead and Balancing Market

Ciaran O'Connor
MSC
119226305

Thesis submitted for the degree of Doctor of Philosophy



NATIONAL UNIVERSITY OF IRELAND, CORK

COLLEGE OF SCIENCE, ENGINEERING AND FOOD SCIENCE

SCHOOL OF COMPUTER SCIENCE & INFORMATION TECHNOLOGY
CENTRE FOR RESEARCH TRAINING IN ARTIFICIAL INTELLIGENCE

April 2025

Head of School: Professor Dirk Pesch

Supervisors: Dr. Steve Prestwich
Dr. Andrea Visentin

Contents

List of Figures . . . iii
List of Tables . . . v
Acknowledgements . . . viii
Abstract . . . x

1 Introduction **1**
1.1 Motivation and Background . . . 1
1.2 Contributions . . . 4
1.3 Publications . . . 6
1.4 Outline of the Dissertation . . . 7

2 Background **9**
2.1 Electricity Market Structure . . . 11
2.1.1 Day-Ahead Market . . . 11
2.1.2 Intra-Day Market . . . 13
2.1.3 Balancing Market . . . 14
2.1.4 Other Electricity Markets . . . 16
2.2 Predictive Models . . . 17
2.2.1 Statistical Methods . . . 18
2.2.2 Machine Learning Methods . . . 24
2.2.3 Deep Learning Methods . . . 30
2.2.4 Hybrid Models . . . 34
2.2.5 Overview . . . 36
2.3 Uncertainty Estimation in Electricity Price Forecasting . . . 36
2.3.1 Epistemic and Aleatoric Uncertainty . . . 38
2.3.2 Uncertainty Estimation Framework . . . 39
2.3.3 Quantile Regression . . . 40
2.3.4 Conformal Prediction . . . 43
2.3.5 Other Probabilistic Forecasting Methods . . . 47
2.3.6 Overview . . . 50
2.4 Trading and Strategy Optimisation . . . 52
2.4.1 Heuristic . . . 53
2.4.2 Linear Programming . . . 54
2.4.3 Mixed Integer Linear Programming . . . 55
2.4.4 Other Trading Approaches . . . 55
2.4.5 Integration of Trading Strategies with Energy Storage . . . 57
2.5 Conclusion . . . 57

3 Day-Ahead and Balancing Market Datasets **59**
3.1 Datasets . . . 60
3.1.1 Day-Ahead Market . . . 61
3.1.2 Balancing Market . . . 65
3.2 Analysis . . . 71
3.2.1 Stationarity Test . . . 72
3.2.2 Autocorrelation and Partial Autocorrelation Analysis . . . 73

3.2.3 Serial Correlation Tests . . . 76
3.2.4 Heteroscedasticity Tests . . . 78
3.2.5 Multicollinearity Tests . . . 81
3.2.6 Correlation Analysis . . . 83
3.2.7 Decomposition . . . 88
3.2.8 Conclusion . . . 91

4 Forecasting the Balancing Market Price 93
4.1 The Importance of the Balancing Market . . . 94
4.2 Methodology . . . 95
4.2.1 Experimental Configuration . . . 95
4.2.2 Balancing Market Prediction Models . . . 97
4.2.3 Point Forecasting Metrics . . . 99
4.3 Experimental Results . . . 101
4.3.1 Forecasting Accuracy . . . 101
4.3.2 Comparison with DAM . . . 106
4.3.3 Limitations . . . 111
4.4 Conclusion . . . 111

5 Dual Market Trading in the Day-Ahead and Balancing Markets with Quantile Forecasts 113
5.1 Energy Storage and Trading Strategies to Address Curtailments 114
5.2 Methodology . . . 115
5.2.1 Forecasting and Trading Framework . . . 115
5.2.2 Trading Strategies . . . 118
5.3 Experimental Results . . . 122
5.3.1 Trading Strategy Evaluation . . . 123
5.3.2 Model and Quantile Pair Performance . . . 127
5.3.3 Economic Viability - Scenario Analysis . . . 130
5.4 Conclusion . . . 135

6 Conformal Prediction for Probabilistic Electricity Price Forecasting 137
6.1 Beyond Efficiency Measures in Probabilistic Forecasting . . . 138
6.2 Methodology . . . 139
6.2.1 Electricity Price Prediction Models . . . 139
6.2.2 Uncertainty Estimation . . . 140
6.3 Probabilistic Forecasting Metrics . . . 147
6.3.1 Efficiency . . . 147
6.3.2 Validity . . . 148
6.4 Experimental Results . . . 149
6.4.1 Day-Ahead Market . . . 150
6.4.2 Balancing Market . . . 153
6.4.3 Financial Performance Analysis . . . 156
6.5 Conclusion . . . 158

7 Conclusions 160
7.1 Future Work . . . 162

List of Figures

2.1 Contrasting granularity of markets 11
2.2 The Implementation of the Day-Ahead Market 12
2.3 Basic structure of the balancing market, ordered by time of occurrence (horizontal) and by actor (vertical). Balancing Service Providers are remunerated for balancing services provided to the System Operator[VdVH16]. 14
2.4 Schematic overview of the typical sequence of existing electricity markets in the EU. Markets in dotted lines are optional [Mee20]. 16
2.5 Quantile forecast of electricity prices 40

3.1 Time Series Plots for Day-Ahead Market Price, and the Predictors. 62
3.2 Time Series Plots for Balancing Market Price, and both Historical and Forward-Looking Predictors 67
3.3 Autocorrelation Plots for the DAM Prices and predictors 74
3.4 Autocorrelation Plots for the BM Prices and predictors 74
3.5 Partial Autocorrelation Plots for the DAM Prices and predictors 75
3.6 Partial Autocorrelation Plots for the BM Prices and predictors . 76
3.7 Pearson Correlation Plots for the DAM Prices and predictors . . 84
3.8 Pearson Correlation Plots for the BM Prices and predictors . . . 85
3.9 Spearman Rank Correlation Plots for the DAM Prices and predictors 87
3.10 Spearman Rank Correlation Plots for the BM Prices and predictors 88
3.11 Seasonal and Trend Decomposition for DAM Prices 89
3.12 Seasonal and Trend Decomposition for BM Prices 90

4.1 Rolling Walk-Forward Validation process for the first three periods in our experimental configuration. 96
4.2 Comparison of the MH-RNN-DNN Model and SH-DNN Model. . 100
4.3 LEAR Forecast for BM 103
4.4 MH RNN DNN Forecasts for BM 104
4.5 Hourly breakdown of the BM prices and the models' forecast error. 107
4.6 DAM and BM prices. 108
4.7 SH DNN Forecast for DAM 109
4.8 Residual Plots for Balancing Market Forecasts 109
4.9 Computational Time Required for Training 110

5.1 Annual percentages of curtailments, constraints, and dispatch down incidents, illustrating the impact of renewable energy challenges on wind output ([Eir23]). 115
5.2 MH-RNN-DNN Model 117
5.3 TS3-Dual Strategy 123
5.4 Average profit for TS1, TS2, & TS3 in each market. Error bars highlight the range for each quantile pair. 124
5.5 TS3 Profit for DAM, BM & DUAL quantile pairs. The error bars are relative to the different forecasting models. 128

5.6 TS3 Profit for Models in the DAM, BM & Dual with Range in Quantile Pair Performance for each Model 129
5.7 TS3 Aggregate Pinball Score of Model vs Profit 130
5.8 15-year cost breakdown for the 3 MW Scottish Power and 3.9 MW Tesla capacity 132
5.9 15-year cost breakdown for Avolta Storage (C) and Tesla Mega-Packs (D) totaling 38.5 MWh 133

6.1 Quantile forecast of electricity prices 140
6.2 Construction of the Ensemble, Q-Ens. 147
6.3 Interval Width for each model in the DAM 151
6.4 DAM Coverage for 0.1-0.9 quantile pair 152
6.5 0.3-0.7 quantile pair Coverage 152
6.6 Interval Width for each model in the BM 154
6.7 BM Coverage for 0.1-0.9 quantile pair 155
6.8 BM Coverage for 0.3-0.7 quantile pair 155

List of Tables

2.1 List of Predictive Models Discussed 19
2.2 Literature on Predictive Models for EPF. The ✓is the approach presented herein 37
2.3 Literature on Predictive Models used outside of EPF or with EPF. The ✓is the approach presented herein 38
2.4 Literature on PEPF. Acronyms- Historical Simulation (HS), Monte Carlo (MC), Distributional Based (DB). The ✓is the approach presented herein. 51
2.5 Literature on other probabilistic forecasting applications. Acronyms- Historical Simulation (HS), Monte Carlo (MC), Distributional Based (DB). The ✓is the approach presented therein. 52
2.6 Literature on Trading Strategies. The ✓is the approach presented herein 56

3.1 ADF and KPSS Test Results for DAM and BM Datasets 73
3.2 Ljung-Box and Breusch-Godfrey Test Results for DAM and BM Models 78
3.3 Breusch-Pagan and White Test Results for DAM and BM 80
3.4 VIF and Condition Index Results for DAM and BM 82

4.1 Balancing Market prediction accuracy for the different training sizes. In bold, the best-performing model. MAE is measured in €/MWh. 102
4.2 DB test statistical significance. The rows represent model M_1 and the columns represent model M_2 105

5.1 Financial Profits from TS3: Bottleneck-Controlled Strategy . . . 126

6.1 APS for the DAM. Best approach for each regressor marked in bold, worst in red. 150
6.2 Winkler Scores for the DAM. Best approach for each regressor marked in bold, worst in red. 153
6.3 Winkler Scores for the DAM. Best approach for each regressor marked in bold, worst in red. 153
6.4 APS for the BM. Best approach for each regressor marked in bold, worst in red. 154
6.5 Winkler Scores for the BM. Best approach for each regressor marked in bold, worst in red. 156
6.6 Financial Performance Comparison of RF & LGBM Models. Best approach marked in bold, worst in red. 157

I, Ciaran O'Connor, certify that this thesis is my own work and has not been submitted for another degree at University College Cork or elsewhere.

Ciaran O'Connor

To my family and friends, whose constant support and encouragement have been my foundation throughout this journey.

Acknowledgements

Looking back to October 2020, when I embarked on this journey, I could not have anticipated the twists and turns that lay ahead. The challenges of recent years, along with unforeseen disruptions, shaped my research and strengthened my resilience, leaving a lasting impact on both my personal and professional life. I am profoundly grateful to the many people who supported me throughout my time at UCC, making this achievement possible.

First and foremost, my deepest thanks go to my supervisors, Steve, Nic, and especially Andrea, whose guidance has been invaluable. Steve and Nic, your early support helped lay the foundation for this journey. Andrea, your arrival at a critical phase during the pandemic proved truly transformative. Your frequent check-ins, constructive critiques, and steadfast guidance kept me focused and motivated. Your encouragement, insights, and direction kept me focused on the ambitions I set out to achieve. Thank you all for your mentorship and for making this journey a collaborative and inspiring experience.

I owe profound gratitude to my family: my parents, Liam and Noreen, for their unwavering support, and my four sisters, Aoife, Aisling, Fiona, and Cliodhna, whose encouragement has been a constant source of strength. To my closest friends, James, Sean, Connor, Cathal, Liam, Jack, David, Ali, Jason, and Lucy, thank you for being by my side long before this PhD and for staying with me through every step. I am also thankful for the friendships I made during this journey, including Challo and Giuli, with whom I shared countless memorable moments, and Hannah, for her steadfast support.

My time at UCC was enriched by my fellow PhD companions, Dhiarya, Daniella, Akasha, Andrea, and Marco. Together, we shared many laughs and memorable moments, building a network of support and camaraderie that I deeply value. I am grateful to have walked this path alongside each of you.

I would also like to acknowledge Joseph Collins, who first introduced me to electricity price forecasting during my master's project on the Day-Ahead Market. His guidance and extensive knowledge in financial and industrial markets have been invaluable throughout my PhD. Joseph's suggestion to explore the Balancing Market was pivotal, shaping the primary focus of this thesis. His involvement in the early chapters—from dataset acquisition and preprocessing to co-authoring papers—has been instrumental in my academic and professional growth. I am immensely grateful for his mentorship and support.

I extend my thanks to Roberto Rossi at the University of Edinburgh, who hosted me during my research visit. His expertise and guidance provided valuable insights that strengthened my understanding and contributed to the development of this thesis. I am also grateful to Mohamed Bahloul for his assistance with one of my papers, offering valuable advice and constructive feedback that significantly improved my work.

Additionally, I extend my thanks to the members and staff of the Centre for Research Training in Artificial Intelligence, the Insight Centre, and the Computer Science Department at UCC. Special thanks to Professor Barry O'Sullivan and Janet Choi for fostering a vibrant, collaborative environment across Irish universities, which provided invaluable opportunities for networking and professional growth. I am also grateful to Peter J. MacHale for his technical support and assistance over the years.

This work has also been supported in part by Science Foundation Ireland under Grant Nos. 18/CRT/6223, 16/RC/3918 and 12/RC/2289-P2, by the EU Horizon projects Glaciation (GA No. 101070141) and TAILOR (GA No. 952215)

Ciaran O'Connor
Cork, April 2025

Abstract

The transition to sustainable energy sources is a critical challenge of the 21st century. Over the past decade, the adoption of renewable resources such as wind and solar power has surged, driven by policy incentives and decreasing costs. While these measures are essential for decarbonising the energy sector, they introduce challenges such as price volatility, efficient energy management, and curtailments. The rapid expansion of wind and solar power also increases variability in net power demand, posing significant reliability concerns for electricity grids. Addressing these challenges requires advanced tools to predict market dynamics and manage the resulting imbalances effectively.

This dissertation responds to these needs by focusing on predicting electricity prices and modelling trading strategies in the Day-Ahead Market (DAM) and Balancing Market (BM), with a primary emphasis on the latter. The DAM is a forward market where electricity is traded a day in advance based on forecasted demand, while the BM operates closer to real-time, allowing for adjustments to address supply-demand imbalances and forecast errors. The BM, characterised by higher volatility than other markets, has become increasingly important for managing forecast errors associated with variable renewable energy. While traditional spot markets like the DAM and Intra-Day Market are well-studied, the BM remains under-explored despite its critical role in balancing real-time supply and demand, particularly in regions where large-scale energy storage solutions are not economically viable, necessitating a real-time balance between production and consumption to minimise curtailments.

The research outlines how to approach forecasting the BM, starting with the collection and preprocessing of datasets for both the Irish DAM and BM, accompanied by a detailed time series analysis. Various techniques for forecasting electricity prices are investigated, utilising statistical, Machine Learning, and Deep Learning models. The study identifies the Lasso Estimated AutoRegressive model as the most effective for accurate price prediction, closely followed by Extreme Gradient Boosting and Random Forest. Deep Learning models, while excelling in the DAM forecast, struggle to accurately forecast prices in the BM.

Expanding on these insights, the thesis explores the optimisation of trading strategies using quantile-based forecasts. Initially, a trading approach is detailed for executing trades once per day in each market, which is then extended to a heuristic optimisation strategy for higher frequency trading. The research

further evolves to a dual-market strategy, holding positions in both markets concurrently, highlighting the advantages of multi-market participation. The outcomes emphasise the pivotal role of increasing trading frequency with large scale energy storage to improve grid reliability and market liquidity.

Furthermore, to refine probabilistic price prediction, this thesis evaluates recent adaptations of Conformal Prediction techniques tailored for time series data, including Ensemble Batch Prediction Intervals and Sequential Predictive Conformal Inference. We demonstrate the efficacy of Conformal Prediction methods in achieving validity and providing more reliable prediction intervals compared to conventional approaches like Quantile Regression and Quantile Regression Averaging, which excel in efficiency. We propose an ensemble combining the efficiency of Quantile Regression with the improved coverage of Conformal Prediction techniques, delivering reliable forecasts with narrow intervals, and demonstrating improved financial returns in both markets.

This thesis advances electricity price forecasting and trading strategies by developing a transparent framework for probabilistic forecasting and trading in both the DAM and BM. For energy producers, the proposed methods enhance revenue stability by mitigating price volatility and curtailment risks. Grid operators benefit from improved forecast accuracy, enabling proactive grid management, effective scheduling, and reduced reliance on costly balancing measures. Optimised trading strategies leverage the DAM's forecast-based scheduling and the BM's real-time adjustments, addressing market-specific challenges while enhancing profitability and liquidity for traders and other market participants.

By supporting the integration of renewable energy sources and reducing curtailments, these methods align with global decarbonisation goals and the transition to a low-carbon energy system. This research provides a practical framework for improving market efficiency and resilience, equipping stakeholders with actionable tools to navigate the complexities of renewable integration.

Chapter 1

Introduction

1.1 Motivation and Background

The global energy landscape has undergone a profound transformation in recent years, driven by the urgent need to reduce greenhouse gas emissions and combat climate change. Among the foremost challenges of the 21st century is the transition to sustainable energy sources. Renewable energy technologies such as wind, solar, and hydroelectric power have seen significant adoption, spurred by policy incentives and decreasing technology costs [MR22]. Governments and regulatory bodies have implemented various support mechanisms, including feed-in tariffs, renewable energy certificates, and tax incentives, to encourage investment in renewable energy projects [KLF+21]. These initiatives not only promote sustainability but also accelerate the transition to a greener energy mix. However, the rapid growth of renewables presents challenges, particularly in maintaining grid stability and managing their inherent variability.

While renewable energy sources are crucial for reducing carbon emissions and promoting sustainable development, they lack the predictability of traditional fossil-fuel-based generation. Fossil fuels typically provide stable and consistent output, whereas renewables depend on weather conditions and other uncertain factors. This variability has significantly impacted electricity markets, particularly short-term markets, which must adapt to manage supply and demand effectively while maintaining grid stability. For instance, in Ireland, renewable energy's share of electricity generation surged from 21.7% in 2014 to 39.5% in 2022 [Eir22], intensifying supply volatility and reliability challenges [MABH16]. Curtailments, defined as the reduction in generator output due to

excess capacity, are amongst the highest in Europe in both the Republic of Ireland (11.1%) and Northern Ireland (13.5%) [YBC+22]. These dynamics underscore the need for grid-level energy storage solutions to absorb excess energy during peak production periods and release it during low production intervals, stabilising demand and mitigating market volatility.

Among the most promising solutions are renewable energy-based microgrids, which offer a means to balance supply and demand while supporting the broader integration of renewables. These microgrids not only provide affordable and clean energy [WHZ+20] but also enhance grid resilience through energy storage. Coupled with energy storage systems, microgrids can shift energy from low-demand to high-demand periods or from cheap to expensive generation times. This capability improves grid reliability and resilience, enabling better power quality management, load shifting, frequency regulation, and backup power [EC10, Gür18]. Additionally, renewable microgrids bring environmental benefits such as pollution reduction [TJH18] and economic advantages by optimising energy systems [Alb18]. The growing integration of microgrids underscores the need for advanced technologies that can further enhance their operational efficiency and economic viability.

Building on these advantages, Battery Energy Storage Systems (BESS) play a critical role in mitigating the variability of renewable sources and reducing market volatility. By storing surplus energy during low-demand periods and releasing it during peak demand, BESS act as a buffer that stabilises the grid and enables market participants to capitalise on arbitrage opportunities. This not only ensures more efficient resource utilisation but also facilitates effective energy trading through advanced forecasting models. These models optimise charge and discharge cycles to reduce costs and increase revenues, maximising the economic and operational benefits of BESS. When integrated into renewable microgrids, BESS enhance the management of renewable resources, improving grid resilience and the economic viability for market participants [AAN+23].

To fully realise these benefits, efficient electricity markets are essential for enabling trading and maintaining grid stability. Electricity markets facilitate transactions through spot markets such as the Day-Ahead Market (DAM), Intra-Day Market (IDM), and Balancing Market (BM). Each market serves a unique purpose: the DAM schedules generation and consumption based on forecasted conditions, the IDM allows for near-real-time adjustments, and the BM manages real-time imbalances. Among these, the BM is particularly critical in addressing

the volatility introduced by renewable energy sources. It operates at a higher granularity to swiftly manage fluctuations, ensuring grid stability and avoiding costly curtailments. The increasing reliance on renewable energy highlights the need for accurate Electricity Price Forecasting (EPF), responsive trading mechanisms, and strong risk management strategies to maintain grid stability by ensuring that electricity supply consistently matches demand [CDCG22].

These factors necessitate sophisticated forecasting models to manage the stochastic nature of renewable energy output and impact on electricity prices, making EPF essential for efficient resource allocation, planning, and risk mitigation. Accurate EPF enables stakeholders—from energy producers to traders and regulators—to make informed decisions that enhance grid reliability and resilience through optimised resource dispatch and effective arbitrage strategies. Trading strategies informed by EPF encourage market participation, stabilise prices, and promote liquidity. By aligning market dynamics with renewable integration, EPF and efficient trading mechanisms foster a responsive electricity market that adapts to variability and supports a stable, sustainable energy ecosystem.

This research addresses the increasing volatility and unpredictability of electricity prices in the Irish market. The growing importance of accurate forecasts of BM prices to participants is outlined in [OT19], where forecast errors of variable renewable electricity will drive demand for BM participation. Accurate BM price forecasts are essential for maintaining real-time equilibrium between supply and demand. Improved forecasting helps grid operators and participants anticipate and mitigate potential imbalances, ensuring reliable electricity supply and preventing disturbances such as blackouts. Reliable forecasts can help reduce price volatility and ensure that electricity prices reflect the true cost of production and consumption. BESS are crucial in this context, offering the flexibility required to manage the intermittency of renewable sources by aligning supply with demand, enabling arbitrage opportunities, improving liquidity, stabilising prices, and enhancing energy system sustainability and resilience.

Accurate EPF has become crucial for the efficient operation of electricity markets, yet most efforts have focused on the DAM and IDM [LMDSW21], leaving the BM relatively unexplored despite its growing importance. Traditional point forecasting methods, which predict a single expected value, fail to capture the inherent uncertainty and volatility of electricity prices. These single-point predictions, while useful, do not provide insight into potential variations or the model's confidence in the predictions. The increasing complexity of electric-

ity markets, driven by renewable energy penetration and smart grids, necessitates probabilistic forecasting methods. Probabilistic Electricity Price Forecasting (PEPF) addresses this need by offering a range of possible outcomes and quantifying associated uncertainties, thus enhancing decision-making for market participants. This approach, driven by advancements in machine learning and statistical techniques, enhances risk management and informed trading, leading to a more resilient and efficient energy market.

Most of the literature on probabilistic forecasting has focused on methods like Quantile Regression (QR) and Quantile Regression Averaging (QRA) [NW15], which predict different quantiles of the price distribution and provide insights into possible price outcomes. However, these methods often struggle with validity, as they fail to guarantee Prediction Interval (PI) coverage-the proportion of true values that fall within the PIs-significantly undermining the reliability of their forecasts. This limitation has led to the exploration of Conformal Prediction (CP) techniques, which excel in providing coverage and ensuring that PIs meet predefined confidence levels. CP methods, such as Ensemble Batch Prediction Intervals (EnbPI) and Sequential Predictive Conformal Inference (SPCI), address issues of exchangeability, providing valid and adaptive PIs, especially in non-stationary settings. These techniques account for both aleatoric and epistemic uncertainty, offering a comprehensive view of potential market scenarios. By combining multiple models, CP approaches improve the reliability of forecasts, enabling market participants to better assess risks and opportunities, leading to more robust trading strategies and improved risk management in an increasingly complex and volatile electricity market.

Given the highly regulated nature of electricity markets, the findings of this research also have policy implications, particularly in shaping market mechanisms that ease renewable integration, improve trading efficiency, and support the adoption of energy storage solutions.

1.2 Contributions

The work in this dissertation focuses on EPF and modelling trading strategies, specifically targeting the Irish BM. The main contributions of this work are outlined below, with references to the corresponding chapters of the dissertation.

Day-Ahead and Balancing Market Datasets

While several datasets for the DAM are readily available for forecasting, there is a notable absence of publicly accessible datasets for the BM, with existing papers on forecasting often suffering from vague and inconsistent preprocessing steps. This thesis addresses these gaps by introducing well-prepared datasets for both the DAM and BM, sourced from the Single Electricity Market Operator (SEMO) [SEM]. These datasets resolve common issues of inconsistency and lack of detail prevalent in the literature, offering a valuable resource for future research. Additionally, detailed documentation for both the formation and preprocessing steps (Chapter 3) ensures reproducibility and supports future research, setting a better standard for data quality and transparency in the field.

Standardised Balancing Market Forecasting Framework

To address the scarcity of research on BM forecasting, this thesis establishes a standardised framework that provides a clear and replicable modelling process (Chapter 4). This framework includes detailed instructions on hyperparameter optimisation, along with making the accompanying dataset and code publicly available, supporting reproducibility and setting benchmarks for future research. By evaluating a variety of statistical, Machine Learning (ML), and Deep Learning (DL) models, the research benchmarks their performance and assesses how each model performs with varying training data sizes. The study identifies the Lasso Estimated AutoRegressive model as the most effective for handling the more real-time nature of the BM, whereas traditional statistical models like Auto-Regressive Integrated Moving Average and DL models perform less effectively, providing insights and tools for future BM forecasting. These results highlight the distinct dynamics of the DAM and BM, emphasising the need for approaches tailored to each market's unique characteristics.

Dual-Market Quantile Trading Strategies

Building on the insights from forecasting the BM, we develop and optimise trading strategies using quantile forecasts (Chapter 5), addressing the limitations of unrealistic forecast horizons that lead to suboptimal trading and reduced market efficiency. This enhances our ability to manage curtailments and boost profitability for traders. We developed three strategies to achieve this:

a single-trade approach, a high-frequency trading approach, and an integrated dual-market approach for the DAM and BM. By leveraging BESS, our study shows the benefits of increased trading frequency and concurrent participation in both markets. Additionally, we assess the economic viability of various BESS configurations, evaluating their long-term performance and profitability over a potential lifespan of 10-15 years. Our findings underscore the importance of optimal battery sizing and the strategic advantages of dual market participation, illustrating how these elements impact costs and drive long-term profitability. Together, they contribute to advancements in energy storage solutions and more effective grid integration.

Conformal Prediction Techniques for Reliable Coverage

This thesis contributes to the field of PEPF, focusing on integrating CP methods into the forecasting framework. These methods demonstrate superior validity compared to conventional QR and QRA, achieving higher coverage guarantees and ensuring more reliable forecasts for market participants. By benchmarking statistical and ML models with probabilistic approaches, this work identifies both models and probabilistic approaches that deliver sharp and reliable forecasts across the DAM and BM. We propose an ensemble of QR and CP approaches, striking a balance between the efficiency of QR forecasts and the validity of CP forecasts, resulting in the strongest financial performance. This integration enhances the accuracy and reliability of PIs, enabling effective trading strategies and advancing the practical application of probabilistic forecasting in electricity markets.

1.3 Publications

The work described in this dissertation have been published as follows:

- Ciaran O'Connor, Joseph Collins, Andrea Visentin, and Steven Prestwich, *Electricity Price Forecasting in the Irish Balancing Market* in Energy Strategy Reviews 2024. See Chapter 4.
- Ciaran O'Connor, Joseph Collins, Andrea Visentin, and Steven Prestwich, *Optimising Quantile-based Trading Strategies in Electricity Arbitrage*, in Energy and AI 2025. See Chapter 5.
- Ciaran O'Connor, Andrea Visentin, and Steven Prestwich: *Conformal Pre-*

diction Techniques for Electricity Price Forecasting, in ECML: 9th Workshop on Advanced Analytics and Learning on Temporal Data, 2024. See Chapter 6.

- Ciaran O'Connor, Andrea Visentin, and Steven Prestwich: *Fundamentals of Time Series Analysis in Electricity Price Forecasting*, accepted in International conference on Time Series and Forecasting (ITISE 2025). See Chapter 3.
- Ciaran O'Connor, Mohamed Bahloul, Roberto Rossi, Andrea Visentin, and Steven Prestwich: *Conformal Prediction for Electricity Price Forecasting*, under review in Energy and AI. See Chapter 6.

1.4 Outline of the Dissertation

The remainder of this dissertation is organised as follows.

In Chapter 2, we provide a review of EPF methods and trading strategies, beginning with an exploration of the structure and dynamics of electricity markets. Following this, we present and review a range of predictive models, from statistical to ML and DL techniques, and assesses probabilistic approaches for capturing uncertainty with sharp, reliable PIs. The chapter concludes by examining current trading strategies in electricity markets, including optimisation and ML techniques, to exploit market volatility for arbitrage.

In Chapter 3, we discuss the datasets used for EPF in the DAM and BM, including both historical and forward-looking data sourced from SEMO. This chapter outlines preprocessing steps to ensure consistency and address common data issues. We conduct a detailed time series analysis covering stationarity, autocorrelation, heteroscedasticity, and multicollinearity, concluding with a decomposition of DAM and BM time series to support reliable forecasting.

In Chapter 4, we present a structured framework for forecasting BM prices, detailing our approach and rationale for selecting models suited to BM unique characteristics.This chapter compares a range of forecasting techniques—including statistical, ML, and DL models—assessing performance and training set size impacts on forecasting accuracy. Each model is benchmarked for accuracy and computational efficiency, establishing a reliable foundation for BM forecasting.

In Chapter 5, we build on forecasting insights to develop and optimise high-frequency trading strategies using quantile forecasts in both the DAM and BM.

Leveraging BESS, we explore the benefits of increased trading frequency and dual-market participation. Case studies and simulations explore the economic viability of these strategies under various battery configurations and costs, highlighting the benefits of higher-frequency trading and dual-market participation.

In Chapter 6, we evaluate CP, QR, and QRA techniques for PEPF. This chapter demonstrates the efficacy of CP in achieving precise and reliable PIs, comparing it with established methods such as QR. We propose an ensemble approach that balances the efficiency of QR with the coverage provided by CP. These techniques are assessed alongside statistical and ML models, showcasing the advantages of integrating CP to improve financial performance in both markets.

Chapter 7 summarises key findings and contributions to EPF and trading strategies in the DAM and BM. This chapter also discusses study limitations, and suggests future research directions, including improvements in forecasting, data integration, and trading optimisation for the evolving electricity market.

Chapter 2

Background

The electricity market is a highly complex and dynamic environment where prices are shaped by a multitude of factors, including fluctuating supply and demand, generation costs, regulatory changes, and the growing influence of renewable energy sources. For market participants—ranging from generators to consumers—accurate price forecasting and strategic trading are essential tools for optimising operations, managing risks, and securing profitability. This chapter provides an in-depth review for each of these areas related to Electricity Price Forecasting (EPF) and trading strategies, establishing a solid foundation for the forecasting and trading methodologies that will be developed throughout this thesis.

Electricity markets are the backbone of the energy sector, facilitating the trade of electricity and associated services while ensuring the stability and reliability of the power supply. The first section of this chapter 2.1 delves into the structure and functioning of the primary spot markets central to this thesis: the Day-Ahead Market (DAM), the Intra-Day Market (IDM) , and the Balancing Market (BM). These markets play a crucial role in the efficient allocation of resources by providing price signals that influence investment and operational decisions. Understanding the mechanisms of these markets is essential for developing accurate price forecasting models and effective trading strategies. This section lays the groundwork by exploring the fundamental aspects of these markets, the dynamics of price discovery, and the roles of various market participants.

Given the fundamental importance of precise price forecasting in electricity markets, the second section of this chapter, 2.2, reviews a diverse array of predictive models employed for electricity price forecasting. The selection of mod-

els spans traditional statistical methods, and more modern Machine Learning (ML) and advanced Deep Learning (DL) approaches. This review draws upon recent literature to highlight models that have consistently demonstrated strong performance in EPF across various market conditions. Models such as Lasso Estimated AutoRegressive (LEAR), Extreme Gradient Boosting (XGB), Random Forest (RF), and Deep Neural Network (DNN) are scrutinised for their strengths and limitations in different forecasting tasks. The discussion highlights the models' comparative performance, paving the way for developing forecasting techniques tailored to the challenges of the specific market.

In addition to forecasting models, we focus on quantifying the uncertainty of price predictions. Due to the inherent volatility of electricity markets, it is essential not only to forecast prices but also to assess the uncertainty surrounding these predictions. The third section of this chapter, 2.3, explores techniques for uncertainty estimation in EPF. Traditional point forecasts often fall short in capturing the full range of potential future prices, necessitating the use of probabilistic forecasting methods. This section delves into various methods such as Quantile Regression (QR), Quantile Regression Averaging (QRA), Conformal Prediction (CP), and its variations including Split Conformal Prediction (SCP), Ensemble Batch Prediction Intervals (EnbPI), and Sequential Predictive Conformal Inference (SPCI). These techniques offer dependable PIs, helping market participants manage risks and make informed decisions in volatile markets.

The final section of this chapter, 2.4, examines how forecasting models are integrated into trading strategies to optimise returns and manage risk in electricity markets. This section examines various approaches to trading and strategy optimisation, including heuristic methods, Linear Programming (LP), Mixed Integer Linear Programming (MILP), amongst others. The integration of advanced optimisation techniques with predictive models is presented as a key factor in navigating the complexities of electricity markets, ensuring that traders can achieve optimal results under varying market conditions.

This chapter serves as a review of the essential components involved in EPF and trading strategies. By exploring market-specific forecasting, predictive models, uncertainty estimation, and trading strategy optimisation, it offers a thorough understanding of the current landscape and the challenges present in electricity markets. The insights gained from this review will inform the development of innovative approaches and methodologies in the subsequent chapters of this thesis, contributing to the advancement of knowledge in electricity price fore-

casting and trading.

2.1 Electricity Market Structure

This section delves into the key electricity spot markets most relevant to this thesis: the DAM, IDM, and BM. Each market serves an important and unique function within the broader electricity market structure, collectively ensuring the grid's stability, reliability, and efficiency. By investigating the distinctive characteristics, operational intricacies, and forecasting methodologies of these markets, this section offers an in-depth analysis of how electricity is traded and managed in day-ahead (DAM), near-real-time (IDM), and real-time (BM) scenarios, as depicted in Figure 2.1. A fundamental understanding of these markets is beneficial for developing accurate forecasting models and optimising trading strategies, which are vital for participants manoeuvring through the complexities of modern electricity markets.

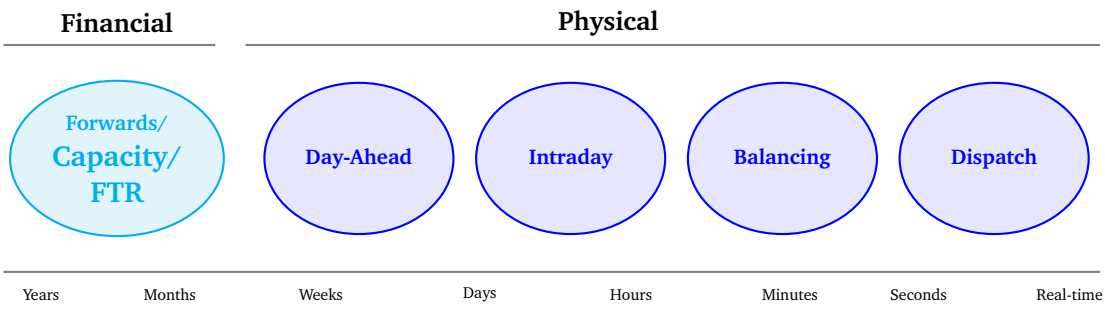


Figure 2.1: Contrasting granularity of markets

2.1.1 Day-Ahead Market

The DAM is a fundamental component of electricity trading, serving as a platform for the forward scheduling of electricity trades conducted one day prior to actual delivery. This advance scheduling is fundamental to balancing supply and demand, ensuring system reliability and efficiency. The DAM plays a vital role in the overall energy market structure by enabling market participants—including generators, suppliers, and consumers—to submit bids and offers based on their anticipations of supply and demand conditions, as detailed in Figure 2.2. A central algorithm, EUPHEMIA for the European market [EUP24], processes these submissions to determine the market-clearing price and volume, thereby facilitating price discovery and market equilibrium. European electricity markets differ from country to country; however, recent years

have seen a significant degree of convergence in the structure and operation of short-term electricity markets with several jurisdictions following a similar template ([NSV16]).

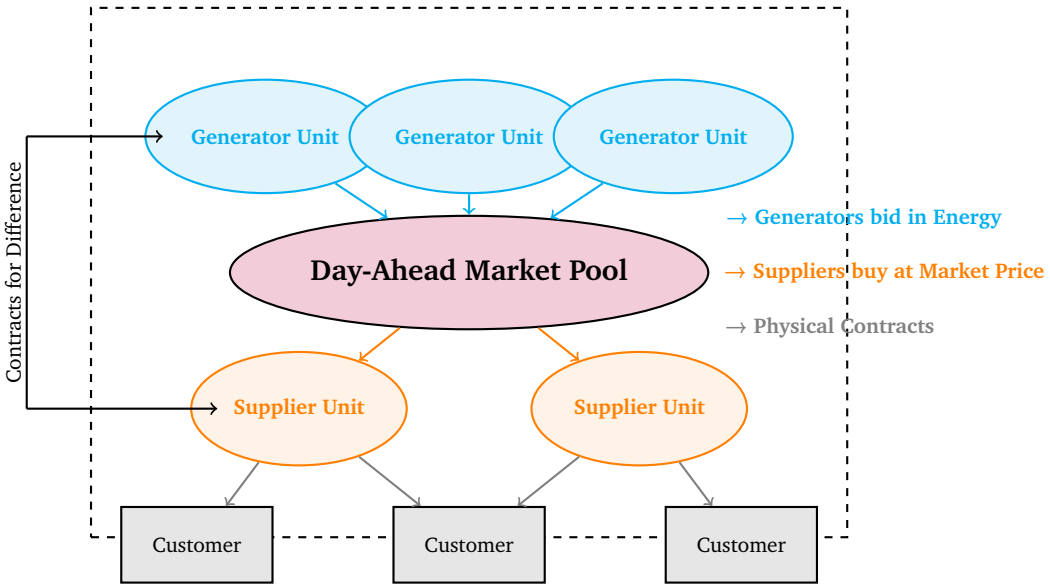


Figure 2.2: The Implementation of the Day-Ahead Market

One of the unique characteristics of the DAM is its capacity for optimising the use of available generation and interconnection capacity across different regions. This optimisation is achieved through the implicit allocation of cross-border capacity, managed via the EUPHEMIA algorithm, which integrates bids and offers to maximise economic efficiency and maintain system stability [IB+18]. The DAM's structure allows for effective cross-border electricity trading, integrating various national grids and managing the allocation of transmission capacity. This integration is critical for balancing regional disparities in supply and demand, improving the reliability of the entire system. However, the DAM also introduces specific challenges, particularly related to the management of forecast uncertainties. The increasing penetration of intermittent renewable energy sources, such as wind and solar, complicates the forecasting and scheduling processes. These renewable sources introduce variability due to factors such as weather conditions, necessitating robust forecasting models to provide accurate predictions for both demand and generation [MABH16].

The short lead time for day-ahead forecasts increases complexity, requiring advanced methodologies to handle uncertainties. Initially, traditional statistical models were the mainstay for forecasting; however, recent advancements in ML and DL have greatly improved forecasting accuracy, proving particularly effective in capturing the complex, non-linear relationships that characterise elec-

tricity markets. DL models in particular have seen impressive success in forecasting DAM prices ([UOT18, LDRDS18, CWMJ19, LB21]). Additionally, hybrid models that combine DL with other forecasting techniques have been shown to improve precision, especially in accounting for the dynamics of supply-demand interactions and varying market conditions [LMDSW21].

2.1.2 Intra-Day Market

The IDM plays a more significant role in real-time electricity trading, offering unique characteristics that distinguish it from the DAM and other markets. The IDM is designed to provide market participants with a platform for real-time trading, allowing them to adjust their positions in response to unexpected changes in supply and demand, such as those caused by plant outages or fluctuating renewable energy outputs. One of the primary distinctions between the IDM and the DAM lies in the timing and frequency of trades. While the DAM involves setting prices and trading volumes a day ahead of delivery, the IDM facilitates continuous trading up to a few hours before real-time delivery. This allows for more granular adjustments, down to 30-minute intervals (in the Irish Single Electricity Market (I-SEM)), compared to the hourly intervals seen in the DAM.

This granularity is essential for managing the variability of renewable sources like wind and solar, which can experience rapid output changes [SA19]. The IDM's real-time nature requires reliable forecasting methods to predict short-term price movements accurately. The literature underscores the importance of modelling tailored specifically for the IDM's higher frequency data. These approaches include statistical modelling [UMW19b, NZ20a], with notable success also achieved through ML and DL models [MRRFJC16, AFRB17]. These models excel at processing vast datasets and adapting to dynamic market conditions by integrating real-time data, such as weather forecasts, plant availability, and interconnector flows, significantly improving prediction accuracy [NZ20b].

Another key aspect of the IDM is its role in optimising trading strategies. Its continuous trading capability and near-real-time nature allow market participants to refine their strategies, respond to market signals, and mitigate risks associated with price volatility. Research focuses on algorithms that optimise bidding strategies in real-time, maximising profitability or minimising costs [KH19]. The IDM's focus on high-frequency data and short-term adjustments facilitates re-balancing supply and demand as more accurate information becomes avail-

able, which is crucial for integrating intermittent renewable sources. As renewable energy integration grows, real-time markets like the IDM will become even more essential for ensuring a stable and efficient electricity supply [BA24].

2.1.3 Balancing Market

The BM, unlike both the DAM and IDM, operates on much finer granularity, demanding immediate responses to fluctuations in supply and demand. This market serves as the "market of last resort," where the Transmission System Operator (TSO) must reconcile short-term deviations between scheduled market positions and actual system conditions, a task complicated by the unpredictable nature of demand changes, renewable energy generation [OT19], and unforeseen generator outages [ZHH+23]. One of the primary challenges in the BM is the accurate forecasting of these real-time imbalances. The volatility of factors such as wind generation and sudden demand shifts requires highly responsive forecasting models. Wind energy, due to its inherent variability, is particularly challenging to predict, and inaccuracies in wind forecasts can severely disrupt the TSO's ability to maintain balance, as outlined in Figure 2.3 [VdVH16]. In terms of I-SEM imbalance pricing details, the accuracy of these forecasts is crucial for determining the real-time imbalance price, which reflects the marginal cost of balancing actions in the market [BD18].

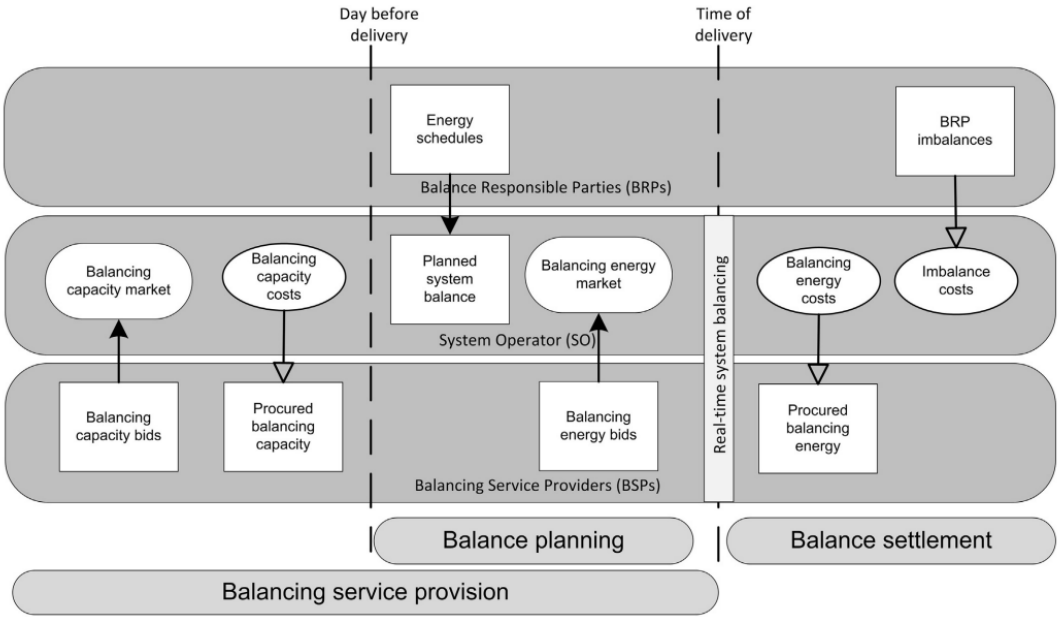


Figure 2.3: Basic structure of the balancing market, ordered by time of occurrence (horizontal) and by actor (vertical). Balancing Service Providers are remunerated for balancing services provided to the System Operator[VdVH16].

A critical feature of the Irish BM is the flagging and tagging process, which ensures that imbalance prices reflect only the cost of energy balancing actions, excluding non-energy interventions such as system security measures. In this system, actions are flagged by purpose, with only energy-related activities affecting the imbalance price, a distinction that is essential for maintaining price signals that accurately represent the marginal cost of energy imbalances without distortions from non-energy actions [BA24]. The flagging and tagging process, along with the BM's market rules, ensures that prices reflect genuine energy imbalances while avoiding excessive interventions that could undermine market flexibility. The BM's design, particularly in Ireland, balances short-term price incentives for demand response with the broader goal of grid stability, a challenge made more complex by the increasing penetration of renewable energy [BD18]. A key advancement in the BM is the integration of real-time data from sources such as weather conditions, grid status, and market activities, enabling TSO's to proactively manage grid stability through precise interventions. Moreover, regional cooperation and data sharing among TSOs across different markets improves the resilience of balancing strategies. Cross-border coordination is crucial for optimising resources and maintaining grid stability in interconnected markets, as shown by efforts to integrate European BM [PLDV20].

The BM presents distinct challenges that demand advanced forecasting techniques and precise real-time interventions to maintain system stability. While ML and DL models have become popular for forecasting in the DAM, they often struggle to match the performance of statistical models in the BM [DBdV$^+$19, Nar22], where predicting real-time deviations is critical. Statistical models consistently outperform their ML and DL counterparts, except in certain cases such as [LPKC20], where ML models performed well but still fell short in accurately predicting price peaks. Although ML and DL models excel at processing large datasets and detecting patterns, their application in the BM remains less reliable due to the market's highly volatile nature. To effectively balance the Irish BM, sophisticated forecasting models, real-time data integration, and a nuanced understanding of market dynamics are essential. Recent progress, driven by the integration of statistical and ML techniques, regional cooperation, and innovative market design, has improved both the reliability and efficiency of the BM. Ongoing research into BM price dynamics, guided by behavioural economics, regulatory frameworks, and market integration, is key to refining forecasting models and ensuring efficient market operations [VdVH16, VDVAH12, ERH21].

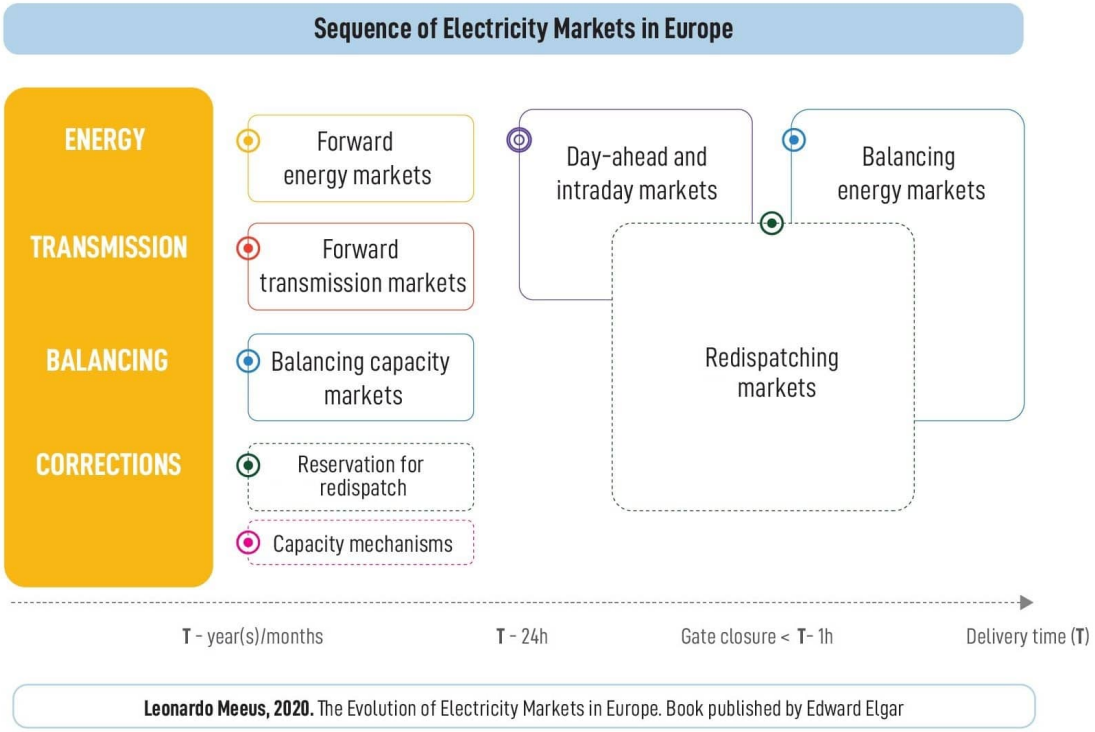


Figure 2.4: Schematic overview of the typical sequence of existing electricity markets in the EU. Markets in dotted lines are optional [Mee20].

2.1.4 Other Electricity Markets

In addition to the DAM, IDM, and BM, other markets are fundamental to the overall functioning of the electricity market. Figure 2.4 provides an overview of all these markets in and outside of this thesis, and below includes a brief overview of these markets and their unique characteristics:

- *Forward Market*: Allows participants to hedge positions in DAM, IDM, and BM through contracts-for-difference [PB21]. contracts-for-differences enable participants to lock in a strike price, providing protection against price volatility. Forecasting in this market requires predicting long-term price trends and movements.
- *Ancillary Markets*: Support the power grid's stability and reliability through services like frequency regulation, spinning reserve, voltage control, and black start capabilities. Accurate forecasting in ancillary markets ensures the availability of resources needed to maintain grid stability. This involves sophisticated models considering real-time operational data and dynamic supply and demand conditions [RRF+22].
- *Capacity Market*: Ensures sufficient generation capacity to meet peak de-

mand. Capacity providers receive payments for committing to supply electricity or reduce demand during peak periods, incentivising investment in new capacity [COS13]. Forecasting in the capacity market involves predicting peak demand periods and capacity resource availability.

- *Financial Transmission Rights Auctions*: Manage congestion costs and provide financial hedges against price differences across market zones. Financial Transmission Rights entitle holders to payments based on price differences between locations. Forecasting in this market involves predicting congestion patterns and price differentials, requiring an understanding of grid operations and transmission constraints [SK08].

Despite forward, ancillary, and capacity markets playing important roles in electricity systems, this thesis focuses on the DAM, IDM, and BM due to their direct impact on price volatility, renewable integration challenges, and real-time trading opportunities. While forward markets and financial transmission rights auctions remain relevant for long-term hedging strategies, their slower operational timescales and fundamentally distinct pricing mechanisms place them beyond the scope of this thesis.

2.2 Predictive Models

EPF is essential for maintaining market efficiency and stability, especially in markets with high renewable energy penetration. This section reviews the range of predictive models applied in the DAM, IDM, and BM. The models covered include statistical 2.2.1, ML 2.2.2, DL 2.2.3, and Hybrid approaches 2.2.4, with additional insights from load, wind, and photovoltaic forecasting to highlight their relevance to EPF, particularly in the underexplored BM. Each category is examined for strengths, limitations, and performance in forecasting prices and managing volatility. The selection process is informed by recent literature across the DAM and BM, prioritising models that demonstrate both high accuracy and computational efficiency in diverse market conditions. The following subsections provide a detailed review and comparison of these models, focusing on their roles in DAM, IDM, and BM forecasting.

2.2.1 Statistical Methods

In this section, we explore key statistical models used in EPF, including the Naive model, Auto-Regressive Integrated Moving Average (ARIMA), and LEAR.

In our price prediction approach, we define the dataset as $\{(\mathbf{x}_t, y_t)\}_{t=1}^{N}$, where $\mathbf{x}_t$ represents the predictor variables and y_t their corresponding target values. Using this dataset, we train models f to forecast future prices at each time step t. The predicted price $\hat{y}_t$ is computed as $\hat{y}_t = f(\mathbf{x}_t)$, where $\hat{y}_t$ is the model's estimate for y_t.

Naive Model

The Naive model is a simple yet effective method for EPF, ideal for capturing short-term dynamics in real-time BM prices. It operates on the principle of temporal proximity, forecasting future prices based on the most recent values. In this study, the model uses the previous 8 hours of BM prices to predict the next 8 hours, leveraging short-term temporal dependencies.

Let y_t denote the BM price at time t. The Naive model's prediction for the price at time $t + h$, where h represents the forecast horizon (with $h \leq 8$), is given by:

$$\hat{y}_{t+h} = y_{t-(8-h)} \tag{2.1}$$

Here, $\hat{y}_{t+h}$ is the forecasted price, and $y_{t-(8-h)}$ is the observed price $8 - h$ hours prior. This assumes that short-term patterns are stable enough for recent values to predict near-term prices. The Naive model's simplicity allows rapid forecasting with minimal effort, useful for real-time decisions. However, it may struggle with abrupt changes, like price spikes or shifts, and lacks the ability to account for external shocks or long-term trends.

The Naive model has proven effective in various electricity markets. In the Lithuanian Nord Pool DAM, where seasonality is pronounced, a simple Seasonal Naive approach performed well [BKM18a]. In more volatile markets like the German IDM, augmenting the Naive model with variables such as Least Absolute Shrinkage and Selection Operator (LASSO) has significantly improved accuracy, highlighting the model's limitations and the benefits of incorporating additional variables [MUW20a]. In [Nar22] a naive model is used to generate initial predictions for probabilistic forecasting, which were then improved with bootstrapping. Despite its limitations, the Naive model remains a useful base-

line for quick approximations, though less accurate than models like ARIMA or LEAR.

Table 2.1: List of Predictive Models Discussed

Acronym	Definition
Statistical Models	
AR	Auto-Regressive
ARX	Auto-Regressive with Exogenous input
ARIMA	Auto-Regressive Integrated Moving Average
LEAR	Lasso Estimated AutoRegressive
GARCH	Generalised Auto-Regressive Conditional Heteroskedasticity
TF	Transfer Function
ES	Exponential Smoothing
Machine Learning Models	
ML	Machine Learning
SVM	Support Vector Machines
SVR	Support Vector Regression
KNN	K-Nearest Neighbors
XGB	Extreme Gradient Boosting
LGBM	Light Gradient Boosting Machines
RF	Random Forest
Deep Learning Models	
DL	Deep Learning
DNN	Deep Neural Network
LSTM	Long Short-Term Memory
GRU	Gated Recurrent Units
TFT	Temporal Fusion Transformer
CNN	Convolutional Neural Networks
BiLSTM	Bi-directional Long Short-Term Memory

ARIMA

The ARIMA model is a widely used approach for both time series analysis and forecasting. It combines three key elements: Auto-Regressive (AR) components, differencing to achieve stationarity (I for Integrated), and a Moving Average

(MA) component. These components allow ARIMA to capture both the influence of past values and past forecast errors, while accounting for trends in the data.

The AR component models the relationship between the current value y_t and its previous values (lags), to capture temporal dependencies. In an AR model of order p, y_t is expressed as a linear combination of its previous p values:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t \tag{2.2}$$

The coefficients $\phi_1, \ldots, \phi_p$ represent the influence of each lagged value, intercept ϕ_0 accounts for the baseline level, and ϵ_t represents the white noise term.

Differencing, the integrated part of ARIMA, is used to achieve stationarity by removing trends. A time series is stationary when its statistical properties, like mean and variance, remain consistent over time. The series:

$$\nabla^d y_t = y_t - y_{t-1} \tag{2.3}$$

is differenced d times until the trend is removed and it becomes stationary.

The MA component models the dependency of the current value on past forecast errors. Unlike AR, which uses past values, MA relies on residuals. In an MA model of order q, y_t is expressed as a function of previous error terms:

$$y_t = \theta_0 + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} \tag{2.4}$$

Here, θ_0 is the intercept, $\theta_1, \ldots, \theta_q$ are the MA coefficients, and ϵ_t are the error terms, allowing the MA component to capture sudden shifts or shocks.

The ARIMA model unifies the AR, differencing, and MA components into a single framework, denoted as ARIMA(p, d, q), where p represents the number of AR terms, d is the number of differencing operations needed for stationarity, and q is the number of MA terms. The general form of the ARIMA model is:

$$\phi_p(B)\nabla^d y_t = \theta_q(B)\epsilon_t \tag{2.5}$$

Here, $\phi_p(B)$ and $\theta_q(B)$ are the AR and MA polynomial operators, and B is the backshift operator, defined as $B^k y_t = y_{t-k}$. The differencing operation ∇^d represented as $(1-B)^d$ ensures the series is stationary. Fitting the ARIMA model involves estimating the values of p, d, q, along with the coefficients ϕ and θ, to

accurately capture the relationships between past values and forecast errors.

ARIMA models have been effective for EPF in DAM, handling non-constant mean, variance, and multiple seasonality [CENC03], but struggles with non-linear behaviours such as mean reversion and price spikes [CHKO04] and price volatility, particularly in BM, where hybrid and ML models are more effective [PDŁ15]. To overcome these limitations, hybrid models and enhancements like wavelet transforms have improved accuracy in managing volatility [CPEM05], while adding error correction mechanisms and confidence interval estimation further improves ARIMA's effectiveness in volatile market conditions [ZYN+06]. ARIMA excels in univariate linear data but requires integration with other techniques for better accuracy in handling non-linear patterns [YCL17]. Incorporating Generalised Auto-Regressive Conditional Heteroskedasticity (GARCH)-type specifications improves ARIMA's accuracy in DAM EPF [Gir16], including density forecasts [BGDGR23]. Incorporating external variables, as in Auto-Regressive Integrated Moving Average with Exogenous input (ARIMAX) models, boosts predictive power [BGDGR23], while using seasonality in Seasonal ARIMAX models further enhances performance, particularly in markets with strong seasonal patterns like the Leipzig Power Exchange [JAS11] and Denmark-West's DAM [KX19], though these models are often outperformed by advanced methods like graph convolutional networks, which outperform it in spatial dependencies for short-term Locational Marginal Price forecasting [YTY+22], and Long Short-Term Memory (LSTM) during market volatility [LSH22, BY24]. Despite these limitations, ARIMA remains a cornerstone of time series analysis.

LEAR

The LEAR model, originally introduced as LASSO X in [UNW16], integrates LASSO with l_2 regularisation, making it particularly suitable for high-dimensional datasets. This combination improves variable selection and regularisation, improving the model's ability to handle multicollinearity and high-dimensionality effectively.

LASSO is a linear regression technique used to model the relationship between a dependent variable y_t and one or more explanatory variables $\mathbf{x}_t$. Linear regression estimates the coefficients β by minimising the sum of squared errors:

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \sum_{t=1}^{T} (y_t - \mathbf{x}_t \beta)^2 \tag{2.6}$$

However, in high-dimensional settings, ordinary linear regression often struggles with overfitting, especially when many predictors are involved. LASSO addresses this issue by applying l_1 regularisation, shrinking some coefficients to zero and effectively selecting only the most relevant predictors, thus serving as a variable selection method and overfitting control mechanism.

LEAR builds on the AR structure introduced in ARIMA (Equation 2.2), where the prices y_t is modelled as a linear function of its previous p lagged values:

$$y_t = \phi_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t \tag{2.7}$$

Here, ϕ_0 is the intercept, ϕ_i are the coefficients for the lagged prices y_{t-i}, and ϵ_t is the error term.

While LASSO efficiently selects relevant predictors by shrinking some coefficients to zero, it may struggle in the presence of correlated predictors. LEAR extends this by incorporating Elastic Net regularisation, which combines the l_1 penalty of LASSO with the l_2 penalty of Ridge regression, formulated as:

$$\hat{\phi} = \underset{\phi}{\operatorname{argmin}} \left(\sum_{t=1}^{T} \left(y_t - \sum_{i=1}^{p} \phi_i y_{t-i} \right)^2 + \lambda_1 \sum_{i=1}^{p} |\phi_i| + \lambda_2 \sum_{i=1}^{p} \phi_i^2 \right) \tag{2.8}$$

Here, λ_1 controls the l_1 penalty, promoting sparsity by shrinking some coefficients to zero, while λ_2 controls the l_2 penalty, which helps handle multicollinearity by shrinking the coefficients. This hybrid approach, combining the AR structure and both regularisation techniques enables LEAR to select correlated variable groups while balancing sparsity and stability.

In EPF, LEAR has become prominent for effectively managing numerous correlated predictors through automated variable selection and multicollinearity handling, with studies like [UNW16] showing it outperforms traditional AR models, and [LMDSW21] highlighting the crucial role of hyperparameter tuning in preventing overfitting while ensuring high accuracy. [TPPR22] underscores LEAR's superior performance in capturing distribution tails essential for risk management, while [KW23] highlights that combining LASSO with models like GARCH can further improve accuracy, especially with feature selection and exogenous variables. Moreover, [MNWZ23] demonstrate LEAR's versatility, not only in providing precise point forecasts but also in extending these into probabilistic forecasts. Despite these advantages, in the DAM, LEAR can be outperformed by DL models [OCM+23, GUO23, YGLZ24].

Other Statistical Models

Several statistical models have been applied in EPF to capture temporal patterns in electricity prices, but they are not the main focus of this thesis. Traditional models like AR, ARIMA, and GARCH are well-regarded for handling linear components of electricity price time series [Gir16, LDRDS18, OCM+23], but studies consistently show that ML models outperform them in handling complex price dynamics [LDRDS18]. Extensions like Auto-Regressive with Exogenous input (ARX) models enhance AR models by including external inputs for better forecasting [BGDGR23], while Full Auto-Regressive with Exogenous input (FARX) adds non-parametric terms to capture non-linear interactions [LS24].

The ES model is a time series forecasting method that predicts future prices by assigning exponentially decreasing weights to past observations. Holt's ES has demonstrated strong accuracy in forecasting residential electricity consumption [IOH22], with further improvements achieved by integrating Transformer architectures for enhanced long- and short-term forecasting [WLS+22]. Advanced models like TBATS extend ES by incorporating trigonometric seasonality and ARIMA errors, outperforming Artificial Neural Networks in Denmark's DAM [KX19] and excelling in forecasting photovoltaic system data [GFF+22]. However, it struggles in non-linear settings without external regressors and performs poorly against Recurrent Neural Network (RNN) [IMK+22]. Hybrid models, such as SVR and FARX, further enhance ES by combining it with trend techniques, improving long-term forecasts for energy demand and EPF [RZW+23, LS24].

The Transfer Function (TF) model has been used in EPF to effectively capture the relationship between input variables, such as demand, and price fluctuations by modelling the dynamic interactions between these factors. Studies like [NCCE02] show that incorporating demand improves predictions in markets like Spain and California, but TF struggles with high-frequency hourly prices due to linear assumptions, with comparisons highlighting the superiority of ML and DL models for non-linear approaches [GGBGDM08, LDRDS18].

GARCH extends traditional AR models by incorporating past variances to better predict time-varying volatility, making it ideal for forecasting sudden price spikes or drops in electricity markets (GARCH(p,q)). Studies like [KR05] demonstrate GARCH's superiority over ARIMA in volatile markets such as Spain and California. Extensions like the k-factor GIGARCH model further improve accuracy by incorporating long memory and stochastic volatility, as shown in the

German DAM [DGV09]. While GARCH models are effective in deregulated markets [Gir16], they struggle with non-linear characteristics in high-frequency data. Enhanced approaches like EGARCH and TGARCH improve forecasting by integrating factors such as demand and renewable energy, effectively capturing volatility clustering and asymmetry [BGDGR23]. Additionally, GARCH-t models with LASSO feature selection have proven effective at managing heteroscedasticity, outperforming ML methods in the New Zealand market [KW23].

The Copula model provides a flexible framework for capturing the dependence structure between multiple variables, without assuming a specific distribution, proving effective in capturing multivariate relationships in electricity markets, showing how renewable energy sources, especially wind, affect prices and volatility [DGRR22]. Integrating a copula model into a spatio-temporal framework improves probabilistic forecasting by effectively addressing non-linear and non-stationary patterns, outperforming baseline models [APS22].

2.2.2 Machine Learning Methods

ML models are increasingly vital in EPF due to their ability to capture complex, non-linear market dynamics without being limited by assumptions of stationarity, unlike most traditional statistical methods. The models discussed—K-Nearest Neighbors (KNN), Support Vector Regression (SVR), RF, XGB, and Light Gradient Boosting Machines (LGBM)—offer diverse and effective approaches for EPF. KNN is noted for its simplicity in capturing local patterns, while SVR excels in modelling non-linear relationships. Ensemble models like RF, XGB, and LGBM improve accuracy by aggregating multiple decision trees, with XGB standing out for its efficiency in handling large datasets. As electricity markets become more volatile, these ML models are fundamental for managing risk and enhancing prediction accuracy in EPF.

K-Nearest Neighbours

KNN is a simple, non-parametric, instance-based learning algorithm popular for classification and regression tasks. The core concept of KNN is to predict the value of a target variable y_t by finding the k closest instances (Neighbors) in the feature space based on a specified distance metric, such as Euclidean distance.

Given a query point at time t, KNN calculates the distance between the feature vector $\mathbf{x}_t$ and every other feature vector in the training set. For real-valued

features, the Euclidean distance between two points $\mathbf{x}_t$ and $\mathbf{x}_s$, is defined as:

$$d(\mathbf{x}_t, \mathbf{x}_s) = \sqrt{\sum_{i=1}^{d}(x_{t,i} - x_{s,i})^2} \tag{2.9}$$

where $\mathbf{x}_t, \mathbf{x}_s \in \mathbb{R}^d$. To ensure that all features contribute equally to the distance computation, it is standard practice to normalise features to a mean of zero and variance of one, especially when measured in different units.

Once the distances are computed, KNN selects the k nearest neighbors to the query point. For KNN regression, the predicted value $\hat{y}_t$ is computed as the average (or a weighted average) of the values of the k nearest neighbors:

$$\hat{y}_t = \frac{1}{k}\sum_{i=1}^{k} y_{(i)} \tag{2.10}$$

where $y_{(i)}$ represents the observed value of the i-th nearest neighbor. The parameter k plays an important role; a smaller k risks overfitting, while a larger k may smooth out important local patterns and lead to under-fitting.

KNN is simple to implement and requires no training phase, but it can be computationally expensive with large datasets and suffers in high-dimensional spaces. Despite these challenges, KNN has been effective in EPF for non-linear and complex markets, with the combination of KNN and techniques like SVR and Principal Component Analysis (PCA) improving both predictive accuracy and computational efficiency [FSD16]. Feature selection and extraction techniques like Decision Tree Regression and Recursive Feature Elimination [AKM+19, AJ19], and the incorporation of external factors such as meteorological data [JKG19], have significantly enhanced KNN's forecasting accuracy in both Nord-Pool and New York markets. KNN's sensitivity to neighbor selection and distance metrics often leads to under-performance in volatile markets compared to advanced models like RF and XGB [WXD+22, BCI+22, KGPA24, OCPV24b], with its accuracy in predicting extreme price peaks further constrained by dataset imbalances [PWC+24].

Support Vector Regression

SVR is an extension of SVM tailored for regression tasks, making it an effective tool for forecasting in domains where relationships between input features and target variables are complex and non-linear, such as in electricity markets. SVR

aims to find a function that predicts the target value y_t within a margin of tolerance ϵ, while balancing model complexity and prediction accuracy.

The fundamental concept of SVR is to determine a regression function $f(\mathbf{x}_t)$ that deviates from the true target values y_t by no more than ϵ for all time steps t, while also ensuring that the function is as flat as possible. This flatness is achieved by minimising the norm $\|w\|^2$ of the weight vector w, subject to constraints involving slack variables ξ_t and ξ_t^*, which account for errors beyond the ϵ-deviation. The parameter C controls the trade-off between function flatness and deviation tolerance, adjusting model complexity and generalisation.

SVR can handle both linear and non-linear relationships through kernel functions such as linear, polynomial, and Radial Basis Function (RBF) kernels. The RBF kernel is popular in EPF due to its ability to capture the intricate, non-linear patterns typical in time-series data, playing a key role in model performance.

SVR presents an attractive approach for EPF, particularly in capturing the non-linear dynamics of electricity markets, as demonstrated by [SDS03], highlighting its advantages over Neural Network (NN) with simpler model training and fewer tunable parameters. Hybrid models, like the two-stage Self-Organising Maps (SOM)-SVR approach [FMC07] and SVR-ARIMA [CW10], enhance SVR performance by managing non-stationary behaviours and combining non-linear and linear forecasting components. Despite its strong performance in short-term forecasting, SVR often struggles with extreme price spikes and is outperformed by ML and hybrid models [LDRDS18, ZFB22, HHF23], though it remains competitive in locational marginal price forecasting [CLMCC+22]. Its limitations in managing volatility, however, remain notable [RZW+23, HSW+24, OCPV24b].

Random Forest

RF is an ensemble learning method that constructs multiple decision trees during training and aggregates their predictions. This approach is particularly effective in forecasting tasks where the relationship between input features and target variables is complex and non-linear, often seen in electricity markets.

The RF algorithm begins by creating multiple bootstrapped subsets of the training data. For each subset, a decision tree is built, but unlike traditional decision tree models, RF selects a random subset of features at each node to determine the best split. This randomness in feature selection reduces the correlation

among individual trees, overall reducing the variance in the ensemble.

The prediction of a RF model is the average of the predictions from all trees:

$$\hat{y}_t = \frac{1}{N} \sum_{n=1}^{N} T(\mathbf{x}_t; \Theta_n) \tag{2.11}$$

where N is the number of trees in the forest, and $T(\mathbf{x}_t; \Theta_n)$ represents the prediction from the n-th tree, which is built using a random subset of the data and features. By averaging these predictions, the model reduces the impact of individual tree variance, resulting in a more stable and accurate overall prediction.

RF is effective for EPF in multiple markets, for real-time forecasting in New York's DAM [MHH+14], for handling concept drift in China's DAM [WXD+22], and for improving price and demand predictions in Spain through joint modelling [GMO16]. Its strength in managing complex, non-linear data is further affirmed by [LDRDS18] in a benchmark against 27 state-of-the-art methods, and by [KAU21], who highlight its superiority over traditional models in volatile markets. RF has also been effectively integrated into hybrid models [XQ23], reducing variance and improving accuracy in markets like Nord Pool [AAK+22, XCL+22]. Though it performs well in the BM, RF is often outperformed by models like LEAR and XGB [OCPV24b], but remains a consistent performer in short-term and probabilistic forecasting [RZW+23, OCPV24a].

Extreme Gradient Boosting

XGB is an advanced implementation of gradient boosting designed for both speed and performance. It builds an ensemble of decision trees in a sequential manner, where each new tree corrects the errors made by the previous trees.

XGB, like LEAR, incorporates both l_1 (Lasso) and l_2 (Ridge) regularisation to control model complexity and prevent overfitting. The objective function is:

$$\text{Obj} = \sum_{t=1}^{n} L(y_t, \hat{y}_t) + \sum_{m=1}^{M} \Omega(f_m) \tag{2.12}$$

where $L(y_t, \hat{y}_t)$ represents the loss function measuring prediction error, and $\Omega(f_m)$ is the regularisation term penalising the complexity of the m-th tree. n is the total number of data points, and M the number of trees in the ensemble.

Predictions are updated by adding the output of a new decision tree $f_m(\mathbf{x}_t)$:

$$\hat{y}_t^{(m)} = \hat{y}_t^{(m-1)} + \eta f_m(\mathbf{x}_t) \tag{2.13}$$

Here, $\hat{y}_t^{(m)}$ is the updated prediction at iteration m, $\hat{y}_t^{(m-1)}$ is the previous prediction, $f_m(\mathbf{x}_t)$ is the output of the m-th tree for input $\mathbf{x}_t$, and η is the learning rate controlling each tree's contribution to the prediction.

The final prediction is the sum of the outputs from all trees in the ensemble:

$$\hat{y}_t = \sum_{m=1}^{M} \eta f_m(\mathbf{x}_t) \tag{2.14}$$

Here, $\hat{y}_t$ is the final prediction for the input features $\mathbf{x}_t$, $\eta f_m(\mathbf{x}_t)$ is the prediction of the m-th tree, and M is the total number of trees in the ensemble.

XGB has proven effective in EPF as demonstrated by [XCL+22], where integrating neural networks with entity embedding for categorical data preprocessing, alongside Shapley Additive Explanations for feature selection, significantly improved both accuracy and interpretability in DAM forecasting. Similarly, [CXW+20] demonstrated that tailoring XGB with historical wind speed data improved wind forecasting, outperforming DNNs and Linear Regression, while [QWB+22] and [RM+21] highlighted the benefits of hyperparameter tuning and regularisation in XGB, leading to enhanced accuracy and stability in load forecasting. In the Nord Pool market, XGB outperforms models like SVM and RF due to its efficient boosting process [AAK+22]. It also excels in the BM [OCPV24b], particularly in bidding and risk management [LPKC20], though simpler models occasionally show better performance [KW23]. When integrated with DL models like DNNs, XGB further enhances EPF accuracy in hybrid frameworks [HSW+24].

Light Gradient Boosting Machines

The LGBM is a gradient-boosting framework designed for high efficiency and scalability, particularly with large-scale, high-dimensional datasets. LGBM builds trees sequentially, with each iteration correcting the residuals of the previous trees. However, it optimises performance by using a leaf-wise strategy, splitting the leaf that maximises loss reduction.

In each iteration m, LGBM minimises the objective function:

$$\text{Obj} = \sum_{t=1}^{n} L(y_t, \hat{y}_t^{(m)}) + \sum_{k=1}^{K} \Omega(f_k) \tag{2.15}$$

where $L(y_t, \hat{y}_t^{(m)})$ represents the loss function, $\hat{y}_t^{(m)}$ is the updated prediction at iteration m, and $\Omega(f_k)$ penalises the complexity of the tree f_k.

LGBM's unique feature is its leaf-wise growth strategy, allowing it to grow deeper trees by splitting the leaf that provides the greatest reduction in the loss function. This approach requires careful tuning to avoid overfitting, addressed by Gradient-based One-Side Sampling (GOSS), which prioritises larger gradient data points to reduce computational cost while retaining predictive power, and Exclusive Feature Bundling (EFB), which reduces dimensionality by bundling mutually exclusive features.

LGBM also applies both l_1 and l_2 regularisation to control overfitting and improve generalisation, with the final prediction as the sum of all tree outputs:

$$\hat{y}_t = \sum_{m=1}^{M} \eta f_m(\mathbf{x}_t) \tag{2.16}$$

where $\eta f_m(\mathbf{x}_t)$ represents the prediction of the m-th tree, the same as XGB.

LGBM consistently delivers fast and accurate predictions in time series forecasting. [SPYW22] and [Mal23] demonstrated LGBM's effectiveness in both energy disaggregation for smart manufacturing and wind turbine power generation forecasting, where its key features—leaf-wise growth, GOSS, and EFB—enabled efficient handling of large, complex datasets, outperforming traditional models and optimising energy management. Building on these strengths, both XGB and LGBM have consistently outperformed top ML and DL models in DAM and BM forecasting [OCPV24b, OCPV24a].

Other Machine Learning Models

In addition to the popular ML models used in this thesis, various other ML models, mainly hybrid approaches, have been explored in EPF. For example, [LGT10] combines the RBF Network with Orthogonal Experimental Design to optimize parameters and reduce overfitting, achieving superior accuracy in predicting price spikes. Similarly, [ZFB22] integrates CatBoost for feature selection with Bi-directional Long Short-Term Memory (BiLSTM) for capturing temporal

dependencies, improving forecasting accuracy over models like Multi-Layer Perceptron and SVR, albeit with higher computational costs.

2.2.3 Deep Learning Methods

Over the past two decades, Deep Learning (DL) has emerged as a powerful extension of neural networks, driven by innovations such as Hinton's efficient training of deep belief networks through greedy layer-wise pretraining [HOT06]. This breakthrough enabled the training of deeper networks beyond traditional models like Multi-Layer Perceptron, leading to significant improvements in generalisation [GBC16]. DL has since been widely applied across fields like image recognition [KSH12], speech recognition [HDY+12], and machine translation [BCB14], with notable success in energy-related applications, particularly wind power forecasting [WWL+16, CCL+17]. In EPF, DL models such as DNNs, LSTMs, Gated Recurrent Units (GRU), Convolutional Neural Networks (CNN), and Temporal Fusion Transformer (TFT) have proven successful at capturing complex temporal dependencies and improving probabilistic forecasts. However, many studies lack comprehensive comparisons with state-of-the-art statistical and ML methods, limiting generalisability [LDRDS18]. While recent work has addressed this by comparing multiple DL models, the need for robust evaluations across diverse markets remains [UOT18, ZAJ+19].

Dense Neural Networks

Dense NNs are a class of ML models composed of multiple layers of interconnected nodes, or neurons, that capture complex, non-linear relationships in data. Each neuron applies a non-linear transformation and passes the result to the next layer, enabling Dense NNs to model intricate patterns beyond statistical methods.

A Dense NN with L layers processes an input vector $\mathbf{x}_t$. In the first layer, $\mathbf{x}_t$ is passed as the input $\mathbf{Z}^{(0)} = \mathbf{x}_t$. Each subsequent layer l transforms the output from the previous layer, producing output $\mathbf{Z}^{(l)}$ for the l-th layer as follows:

$$\mathbf{Z}^{(l)} = \sigma\left(\mathbf{W}^{(l)}\mathbf{Z}^{(l-1)} + \mathbf{b}^{(l)}\right), \quad l = 1, \ldots, L \tag{2.17}$$

Here, $\mathbf{W}^{(l)}$ represents the weight matrix that defines the connections between the neurons in layer $l-1$ and layer l. The bias vector $\mathbf{b}^{(l)}$ allows the model to shift the activation function output. The activation function $\sigma(\cdot)$, applied

element-wise to $\mathbf{W}^{(l)}\mathbf{Z}^{(l-1)} + \mathbf{b}^{(l)}$, captures the non-linear relationships.

For regression tasks like EPF, the output layer often consists of a single neuron with a linear activation function, yielding the final prediction, $\hat{y}_t$, as:

$$\hat{y}_t = \mathbf{W}^{(L)}\mathbf{Z}^{(L-1)} + \mathbf{b}^{(L)} \tag{2.18}$$

where $\mathbf{W}^{(L)}$ and $\mathbf{b}^{(L)}$ are the weights and biases of the final layer. The weights $\mathbf{W}^{(l)}$ and biases $\mathbf{b}^{(l)}$ are adjusted using optimisation techniques like stochastic gradient descent or the Adam optimiser to minimise the loss function. To prevent overfitting, Dense NNs often use regularisation techniques such as l_2-norm regularisation or dropout. l_2-regularisation penalises large weights, while dropout randomly ignores neurons during training, reducing dependency on specific neurons. Both techniques improve generalisation to unseen data.

Research on DL methods for EPF has grown significantly, with [LDRDS18] showing that DNNs and hybrid models like LSTM-DNN improve in capturing non-linear relationships, outperforming traditional statistical and ML models in the DAM. However, subsequent studies have encountered limitations, due to a limited number of benchmark models and short test periods [AE19], insufficient hyperparameter tuning for new datasets [LW19, MJ19], and overfitting issues [SW19], which have hindered generalisability. Building on this, more recent research, such as [LDRVDS18], has improved forecasting precision by incorporating inter-market features into DNNs to better capture electricity market interconnections. While DNNs excel at capturing non-linear dynamics, simpler models sometimes outperform them, especially in volatile markets like the BM [JLMW22, OCPV24b]. Nonetheless, in hybrid frameworks, DNNs have proven highly effective in multi-market settings [YGLZ24], and integrating DNNs with XGB for feature selection further improves performance [HSW$^+$24].

Long Short-Term Memory

Originally introduced by [Hoc97], LSTMs are particularly useful in time series forecasting tasks. LSTM is a type of RNN designed to effectively capture long-term dependencies in sequential data, addressing the vanishing gradient problem often encountered in traditional RNNs.

An LSTM unit consists of three key components: the cell state, the hidden state, and gates (input, forget, and output gates). The cell state $\mathbf{C}_t$ acts as a memory, retaining information over long periods. The hidden state $\mathbf{h}_t$ represents the

output at each time step. The gates control the flow of information into and out of the cell. At each time step t, the LSTM receives an input vector $\mathbf{x}_t$ and updates its cell state $\mathbf{C}_t$ and hidden state $\mathbf{h}_t$ using the following equations:

$$\mathbf{f}_t = \sigma\left(\mathbf{W}_f \mathbf{x}_t + \mathbf{U}_f \mathbf{h}_{t-1} + \mathbf{b}_f\right) \quad \text{(Forget Gate)} \tag{2.19}$$

$$\mathbf{i}_t = \sigma\left(\mathbf{W}_i \mathbf{x}_t + \mathbf{U}_i \mathbf{h}_{t-1} + \mathbf{b}_i\right) \quad \text{(Input Gate)} \tag{2.20}$$

$$\tilde{\mathbf{C}}_t = \tanh\left(\mathbf{W}_C \mathbf{x}_t + \mathbf{U}_C \mathbf{h}_{t-1} + \mathbf{b}_C\right) \quad \text{(Cell Candidate)} \tag{2.21}$$

$$\mathbf{C}_t = \mathbf{f}_t \odot \mathbf{C}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{C}}_t \quad \text{(Cell State Update)} \tag{2.22}$$

$$\mathbf{o}_t = \sigma\left(\mathbf{W}_o \mathbf{x}_t + \mathbf{U}_o \mathbf{h}_{t-1} + \mathbf{b}_o\right) \quad \text{(Output Gate)} \tag{2.23}$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh\left(\mathbf{C}_t\right) \quad \text{(Hidden State Update)} \tag{2.24}$$

Here, $\sigma(\cdot)$ represents the sigmoid activation function, $\tanh(\cdot)$ is the hyperbolic tangent function, $\mathbf{W}$ and $\mathbf{U}$ are weight matrices for the input $\mathbf{x}_t$ and the previous hidden state $\mathbf{h}_{t-1}$, and bias terms $\mathbf{b}$. The forget gate $\mathbf{f}_t$ controls how much of the previous cell state $\mathbf{C}_{t-1}$ is retained, the input gate $\mathbf{i}_t$ determines how much of the new candidate cell state $\tilde{\mathbf{C}}_t$ is added, with the output gate $\mathbf{o}_t$ governing how much of the updated cell state $\mathbf{C}_t$ is used to compute the hidden state $\mathbf{h}_t$.

The LSTM processes a sequence of input features $\mathbf{x}_t$ over all time steps to generate the final prediction $\hat{y}_t$ based on the hidden state for final time step:

$$\hat{y}_t = \mathbf{W}_y \mathbf{h}_t + \mathbf{b}_y \tag{2.25}$$

where $\mathbf{W}_y$ and $\mathbf{b}_y$ are the weight and bias for the output layer. To improve generalisation and prevent overfitting, LSTM models, like DNNs, often use regularisation techniques such as dropout, where a random subset of neurons is ignored during training, and l_2-norm regularisation, which penalises large weights.

Since 2018, research on LSTMs in EPF has gained momentum, showcasing their ability to capture complex temporal dependencies and non-linear patterns, especially when integrated into hybrid models like CNN-LSTMs [CPH+18, MJA+19]. Early studies, however, faced limitations such as small datasets and limited model comparisons [JH18, MJA+19], along with short test periods and insufficient evaluations [ZDL+18, ZZM+19]. Despite these challenges, LSTMs demonstrated clear advantages over GRUs and traditional statistical methods, excelling in the DAM [KH18, UOT18]. To further improve their performance, optimisation techniques like the Adam algorithm were applied in the New South

Wales DAM [CZC18], while combining LSTMs with wavelet transforms helped stabilise time series and improve accuracy across multiple markets [CZC19]. More recently, hybrid models combining wavelet transforms, feature selection, and LSTMs have improved forecasting accuracy in markets like the US and Spain [MK21], while empirical wavelet transforms and attention mechanisms have been tailored for high-renewable markets such as Denmark [MWZ+22]. Further advancements, including adaptive feature selection with Bayesian optimisation [XQ23], have refined LSTM's ability to manage non-linear volatility in markets like Denmark's DAM and IDM [KNT24, AAK+22], though GRUs can offer better computational efficiency in some cases [RRE22].

Other Deep Learning Models

Although DNNs and LSTMs are among the most successful models for forecasting in the DAM and IDM, several other advanced DL techniques have also shown considerable promise in EPF.

CNNs are NNs designed to efficiently process structured grid data, such as time series or spatial data, by using convolutional layers to automatically detect patterns like trends and local dependencies. CNNs have proven effective in EPF, excelling in feature extraction and handling non-linear market dynamics [KFA+20, YTY+22]. CNNs combined with PCA reduce overfitting and computational complexity [KFA+20], while CNN autoencoders outperform GRUs in EPF, efficiently managing high-dimensional data [YS22]. CNNs capture complex patterns and seasonality in electricity prices, though slightly less accurate than DNNs and GRUs in some cases [LDRDS18]. Hybrid approaches, like combining CNNs with self-attention mechanisms, further improve accuracy in short-term EPF by capturing both local and long-range dependencies [PMM+24].

GRUs are a simplified version of LSTMs that capture sequential dependencies in time series data using gating mechanisms. They offer a computationally lighter and easier-to-train alternative, though they are less effective with long-term dependencies. In the Turkish DAM, GRU outperformed other ML and RNN models [UOT18], while also proving effective in short-term EPF [LDRDS18]. GRU's integration with Bagged Regression Trees improved mid-term accuracy in handling non-linear fluctuations [IMK+22]. While CNN-based models outperformed GRU in real-time EPF [YS22], GRU showed superiority in long-term forecasts when combined with stacked autoencoders [RRE22]. Though slightly less accurate than LSTM in hybrid frameworks [XQ23], GRU performs well with

proper hyperparameter tuning [KW23]. Recent innovations, such as embedding time as a feature, GRU's ability to capture periodic patterns and handle both synchronous and asynchronous time-series data [ZRB24]. Moreover, hybrid models integrating GRU have improved performance EPF in the DAM [SW24].

TFTs are a NN architecture designed for time series forecasting, integrating attention mechanisms, recurrent components, and static covariates to model dynamic relationships and improve interpretability, making them particularly suitable for EPF [Kot22]. TFT outperforms ARIMA, LSTM, and XGB in photovoltaic power forecasting due to its interpretability [LSGSEC+22], and performs well in energy consumption forecasting with lower error rates compared to LSTM and Temporal CNN [NSSK23]. It improves Probabilistic Electricity Price Forecasting (PEPF) accuracy in markets like Nord Pool and Poland [JPDW24], though it may overfit smaller datasets [SW24]. TFT also excels in real-time extreme EPF, achieving superior performance compared to LSTM variant and other temporal models [PWC+24].

DeepAR is a probabilistic forecasting model based on RNNs that models individual time series while sharing parameters across related time series. DeepAR effectively captures seasonality and trends in DAM EPF, outperforming traditional methods [SFGJ20, Kot22] but falling short against models that incorporate heteroscedasticity [SW24].

Prophet captures seasonality, trend changes, and holiday effects by modelling non-linear trends, while effectively handling missing data and outliers. In the Italian DAM, [SSMC23] applies Prophet with Seasonal and Trend (STL) to improve forecast accuracy by isolating key patterns and reducing noise. Similarly, [PWC+24] demonstrates its effectiveness in univariate forecasting of exogenous variables like weather, further enhancing real-time EPF through its ability to model both periodic and non-periodic behaviours.

2.2.4 Hybrid Models

Hybrid models, combining statistical, ML, and DL methods, have proven effective in handling non-linearities, volatility, and different time horizons, offering a flexible framework for capturing complex patterns.

Statistical Hybrid Models

Hybrid models like ARIMA-GARCH capture linear autocorrelations and non-linear volatility [Gir16], while FARX-elastic net and FARX-LASSO enhance accuracy by combining AR exogenous inputs with Elastic Net and Lasso to reduce overfitting [LDRDS18]. Statistical-ML hybrids, like Wavelet-ARIMA with RBF networks and SOM with SVR, enhance performance by combining linear decomposition with non-linear modelling, while statistical-DL hybrids, such as the DNN-LEAR model, improve accuracy through DL-based hyperparameter optimisation [LMDSW21].

Machine Learning Hybrid Models

[AAK+22] introduced a hybrid model that combines Automatic Relevance Determination with Extra Tree Regression to effectively manage trend, seasonality, and volatility. Similarly, [ZTW+22] used variational and ensemble empirical modal decomposition with an extreme learning machine optimized by differential evolution to capture non-linearities in markets like Australia and Spain. For long-term forecasting, models like those by [GWBS22] integrate data-driven methods with market-based predictions, leveraging Fourier analysis to address both short- and long-term dynamics.

Deep Learning Hybrid Models

DL hybrid models have significantly improved EPF by combining advanced feature extraction and dependency management techniques. For example, CNN-SVR hybrids enhance accuracy through CNN's feature extraction capabilities [ZAJ+19], while CNN-LSTM models effectively capture both local and sequential dependencies, boosting accuracy in markets like Iran's DAM [HHF23]. LSTM-based models, such as the LSTM-Neural Prophet hybrid, improve short- and long-term forecasting by leveraging temporal dependencies [SFF22], with further optimization through Back Propagation NN outperforming standalone models [BHTL22]. A two-stage DNN-LSTM model also improves prediction accuracy by combining neural networks for feature extraction and sequential dependency management [GDCPSS24], while GRU-LSTM hybrids are highly effective in capturing both sequential and non-linear patterns in the DAM [AAK+22]. LSTM's robustness in long-term forecasting is enhanced by advanced optimisation techniques that address the vanishing gradient problem, as seen in CatBoost-BiLSTM hybrids for feature selection and long-term dependency capture [ZFB22],

and BiLSTM-attention hybrids for improved spike detection [GWA23]. Finally, integrating DNN with XGB further enhances performance by combining powerful feature selection with DL's ability to model complex relationships [HSW+24].

While advanced models (e.g., ML & DL hybrids) may achieve high accuracy, their computational complexity can hinder real-time deployment in volatile markets like the BM, necessitating trade-offs between precision and practicality.

2.2.5 Overview

The trends in Tables 2.2 and 2.3 show a progression from traditional statistical methods, predominantly used in earlier studies such as [KEF15] and [Gir16], to more sophisticated approaches as data complexity increased. These statistical methods were primarily applied to the DAM and, to a lesser extent, the IDM and BM. However, from 2017 onward, there has been a marked shift toward ML and DL models. Studies such as [YCL17], [XB19], and [LDRVDS18] highlight the growing use of ML techniques, including RF and SVM, often combined with DL models to improve accuracy in handling complex market dynamics.

In recent years, DL models like LSTM, CNN, and BiLSTM have become increasingly popular, excelling in capturing sequential dependencies and non-linear patterns in markets like DAM and IDM, as noted in [ZAJ+19] and [HHF23]. The prevalence of hybrid models, combining ML and DL, has further enhanced forecasting accuracy, with studies such as [ZFB22] and [GWA23] demonstrating their success in improving spike detection and performance across markets. Table 2.3 extends this trend beyond EPF to include applications in wind power and load forecasting, with models like CatBoost-BiLSTM and attention mechanisms showing strong adaptability and success in managing high-dimensional datasets [ZRB24].

2.3 Uncertainty Estimation in Electricity Price Forecasting

Probabilistic forecasting methods have become increasingly important in EPF due to the inherent uncertainty and volatility of energy markets. Traditional point forecasts, while straightforward, often fall short in capturing the full range of potential future prices, particularly in markets characterised by non-linear and dynamic behaviour [NW18, ZS18]. This section delves into the advanced

Table 2.2: Literature on Predictive Models for EPF. The ✓is the approach presented herein

Reference	Statistical	ML	DL	Hybrid	DAM	IDM	BM	Other
Klaeboe et al. 2015 [KEF15]	✓				✓		✓	
Poplawski et al. 2015 [PDŁ15]	✓	✓					✓	
Uniejewski et al. 2016 [UNW16]	✓				✓			
Girish et al. 2016 [Gir16]	✓				✓			
Feijoo et al. 2016 [FSD16]	✓	✓			✓			
Gonzalez et al. 2016 [GMO16]	✓	✓			✓			
Yang et al. 2017 [YCL17]	✓	✓		✓	✓			
Lago et al. 2018 [LDRDS18]	✓	✓	✓	✓	✓			
Rita et al. 2018 [BKM18a]	✓		✓		✓			
Ugurlu et al. 2018 [UOT18]	✓		✓		✓			
Kuo et al. 2018 [KH18]		✓	✓	✓	✓			
Chinnathambi et al. 2018 [CPH+18]			✓		✓			
Jiang et al. 2018 [JH18]	✓		✓	✓	✓			
Chang et al. 2018 [CZC18]	✓		✓	✓	✓			
Zhang et al. 2018 [ZCS18]	✓			✓	✓			
Johannesen et al. 2019 [JKG19]		✓			✓			
Zhou et al. 2019 [ZZM+19]		✓	✓	✓	✓			
Xu et al. 2019 [XB19]	✓		✓	✓	✓			
Zahid et al. 2019 [ZAJ+19]		✓	✓	✓	✓			
Chang et al. 2019 [CZC19]			✓	✓	✓			
Zahid et al. 2019 [ZAJ+19]		✓	✓	✓	✓			
Atef et al. 2019 [AE19]		✓	✓		✓			
Luo et al. 2019 [LW19]	✓		✓		✓			
Lago et al. 2019 [LDRVDS18]			✓	✓	✓			
Karabiber et al. 2019 [KX19]	✓		✓		✓			
Johannesen et al. 2019 [JKG19]		✓			✓			
Mujeeb et al. 2019 [MJ19]	✓	✓	✓	✓	✓			✓
Ashfaq et al. 2019 [AJ19]		✓	✓	✓	✓			✓
Schnurch et al. 2019 [SW19]	✓	✓	✓		✓			
Marcjasz et al. 2020 [MUW20a]	✓					✓		
Lucas et al. 2020 [LPKC20]		✓					✓	
Khan et al. 2020 [KFA+20]			✓		✓			
Li et al. 2021 [LB21]	✓	✓	✓	✓	✓			
Lago et al. 2021 [LMDSW21]	✓		✓	✓				
Kara et al. 2021 [KAU21]		✓			✓			
Wang et al. 2022 [WXD+22]		✓			✓			
Lehna et al. 2022 [LSH22]	✓		✓	✓	✓			
Narajewski et al. 2022 [Nar22]	✓		✓	✓			✓	
Tschora et al. 2022 [TPPR22]	✓	✓	✓		✓			
Xie et al. 2022 [XCL+22]		✓	✓		✓			
Zhang et al. 2022 [ZFB22]	✓	✓	✓	✓	✓			
Meng et al. 2022 [MWZ+22]		✓	✓	✓	✓			
Yang et al. 2022 [YTY+22]	✓		✓	✓	✓			
Kotsias et al. 2022 [Kot22]	✓	✓	✓			✓		
Iqbal et al. 2022 [IMK+22]	✓	✓	✓	✓	✓			
Beltran et al. 2022 [BCI+22]	✓	✓	✓	✓	✓			
Jkedrzejewski et al. 2022 [JLMW22]			✓		✓			
Alkawaz et al. 2022 [AAK+22]	✓	✓		✓	✓			
Rezaei et al. 2022 [RRE22]			✓	✓	✓			
Yang et al. 2022 [YS22]	✓	✓	✓	✓			✓	
Heidarpanah et al. 2023 [HHF23]		✓	✓	✓	✓			
Bille et al. 2023 [BGDGR23]	✓			✓	✓			
Xiong et al. 2023 [XQ23]			✓	✓	✓			
Kapoor et al. 2023 [KW23]	✓	✓	✓	✓	✓			
Marcjasz et al. 2023 [MNWZ23]	✓		✓	✓	✓			
Olivares et al. 2023 [OCM+23]	✓		✓	✓	✓			
Gunduz et al. 2023 [GUO23]	✓		✓	✓	✓			
Stefenon et al. 2023 [SSMC23]			✓		✓			
Yang et al. 2024 [YGLZ24]	✓		✓	✓	✓			
Kilicc et al. 2024 [KNT24]	✓	✓	✓	✓		✓		
Ghimire et al. 2024 [GDCPSS24]		✓	✓	✓		✓		
Bozlak et al. 2024 [BY24]	✓		✓	✓	✓			
Lisi et al. 2024 [LS24]	✓			✓	✓			
Kitsatoglou et al. 2024 [KGPA24]	✓	✓	✓	✓	✓			
Peng et al. 2024 [PWC+24]		✓	✓				✓	
Huang et al. 2024 [HSW+24]		✓	✓	✓	✓			
Pourdaryaei et al. 2024 [PMM+24]			✓		✓			
Shi et al. 2024 [SW24]		✓	✓	✓	✓			

Table 2.3: Literature on Predictive Models used outside of EPF or with EPF. The ✓is the approach presented herein

Reference	Statistical	ML	DL	Hybrid	DAM	IDM	BM	Other
Zhu et al. 2018 [ZDL+18]		✓	✓					✓
Ashfaq et al. 2019 [AJ19]		✓	✓	✓	✓			✓
Mujeeb et al. 2019 [MJ19]	✓	✓	✓	✓	✓			✓
Ali et al. 2019 [AKM+19]		✓			✓			✓
Cai et al. 2020 [CXW+20]	✓	✓	✓					✓
Salinas et al. 2020 [SFGJ20]	✓	✓	✓		✓			✓
Rafi et al. 2021 [RM+21]	✓	✓	✓					✓
Memarzadeh et al. 2021 [MK21]			✓	✓	✓			✓
Gellert et al. 2022 [GFF+22]	✓							✓
Lopez et al. 2022 [LSGSEC+22]	✓	✓	✓					✓
Woo et al. 2022 [WLS+22]	✓		✓	✓				✓
Durante et al. 2022 [DGRR22]	✓				✓			✓
Arrieta et al. 2022 [APS22]	✓			✓				✓
Shohan et al. 2022 [SFF22]		✓	✓	✓				✓
Bashir et al. 2022 [BHTL22]	✓		✓	✓				✓
Suvarna et al. 2022 [SPYW22]		✓		✓				✓
Ishak et al. 2022 [IOH22]	✓							✓
Qinghe et al. 2022 [QWB+22]		✓		✓				✓
Cantillo et al. 2022 [CLMCC+22]		✓						✓
Nazir et al. 2023 [NSSK23]			✓					✓
Rao et al. 2023 [RZW+23]	✓	✓	✓	✓				✓
Seyed et al. 2023 [Mal23]		✓						✓
Gomez et al. 2023 [GWA23]	✓	✓	✓	✓	✓			✓
Zhang et al. 2024 [ZRB24]			✓					✓

techniques for uncertainty estimation in EPF, focusing on methods that achieve sharp and reliable probabilistic forecasts.

The shift towards PEPF has been emphasised by initiatives such as the Global Energy Forecasting Competition in 2014, which highlighted the importance of metrics like the pinball score for evaluating forecast performance [MNW16]. Key methodologies explored in this section include QR [KBJ78], QRA [NW15], CP [GVV98], and its variations such as SCP [VGS05], EnbPI [XX21], and SPCI [XX23]. These approaches not only improve the accuracy of forecasts but also generate PIs that reliably capture market variability and uncertainty.

Recent reviews have underscored the importance of probabilistic forecasting in addressing uncertainties related to smart grids, supply-demand dynamics, and price variations [KL22]. By integrating both well-established methods and recent innovations surrounding CP, this chapter provides a review of the state-of-the-art in uncertainty estimation, highlighting both the mathematical foundations and practical applications of these techniques.

2.3.1 Epistemic and Aleatoric Uncertainty

Uncertainty in EPF can be broadly categorised into two types: epistemic and aleatoric. Aleatoric uncertainty, often referred to as statistical uncertainty, arises

from the inherent randomness or noise within the data itself. This type of uncertainty is typically considered irreducible as it is intrinsic to the process being modelled. For instance, the variability in electricity prices caused by unpredictable factors like sudden changes in demand or supply, such as weather fluctuations or power plant outages, introduces aleatoric uncertainty. While this randomness cannot be eliminated, probabilistic techniques can model it to reflect natural variability. In contrast, epistemic uncertainty stems from a lack of knowledge or data about the process being modelled. Unlike aleatoric uncertainty, epistemic uncertainty is considered reducible and can be minimised through model optimisation or by gathering more data [KG17]. This may arise from incomplete market information, economic factors, or model limitations.

Although aleatoric uncertainty is often viewed as irreducible, its classification can shift depending on the context or the inclusion of more precise modelling or additional variables [DKD09]. This fluidity blurs the lines between aleatoric and epistemic uncertainties, showing that these categories are not absolute but context-dependent. Frequentist perspectives align with aleatoric uncertainty as the stable frequencies of random events, while Bayesian perspectives align with epistemic uncertainty as degrees of belief based on incomplete knowledge [Hac06]. Epistemic uncertainty is often further subdivided into model uncertainty—related to whether the model structure is correctly specified—and parameter uncertainty, which addresses the accuracy of parameter estimates [Sul15]. In ML models, particularly NNs, epistemic uncertainty is often tied to parameter estimation, as the model space is usually large enough to capture ground truth. Thus, this uncertainty can be reduced by acquiring more data [HW21]. However, in practice, the distinction between aleatoric and epistemic uncertainty often becomes blurred as model complexity and data limitations intertwine [Hor96]. This dynamic boundary highlights the need for flexibility in categorising uncertainties, as what appears irreducible may become reducible with improved models and understanding.

2.3.2 Uncertainty Estimation Framework

Building on the price prediction framework described in Section 2.2.1, we extend the regression task to include probabilistic forecasting. The dataset remains defined as $\mathcal{Z} = \{(\mathbf{x}_t, y_t)\}_{t=1}^{N}$, where $\mathbf{x}_t$ represents the covariates and y_t the target variable. As before, a model f is trained to predict y_t based on $\mathbf{x}_t$, with the point prediction given by $\hat{y}_t = f(\mathbf{x}_t)$.

In the probabilistic setting, however, the goal is not only to estimate a single point prediction but also to quantify the uncertainty around it. This involves estimating conditional quantiles $Q_{y_t}(\alpha \mid \mathbf{x}_t)$, which predict the value below which a certain percentage α of the future price y_t is expected to fall. The lower quantile at level α is denoted as $\hat{q}_t^{\alpha}$, and the upper quantile at level $1-\alpha$ as $\hat{q}_t^{1-\alpha}$.

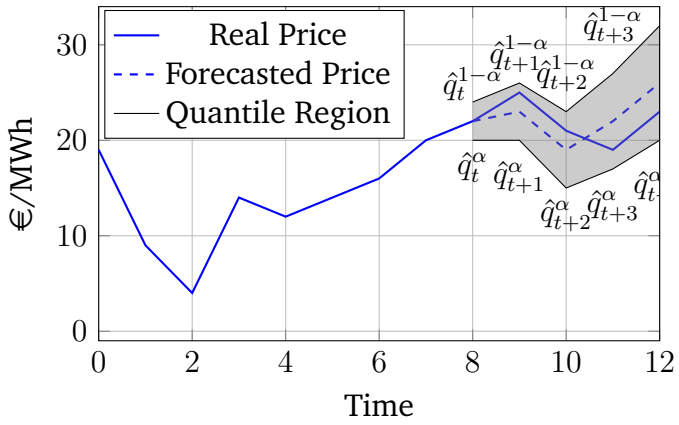


Figure 2.5: Quantile forecast of electricity prices

In the probabilistic prediction setting, as shown in Figure 2.5, the model f generates a PI defined by these quantiles:

$$\Gamma_{1-2\alpha}(\mathbf{x}_t) = [\hat{q}_t^{\alpha}, \hat{q}_t^{1-\alpha}] \tag{2.26}$$

This interval indicates that the future price y_t is expected to fall within the range $[\hat{q}_t^{\alpha}, \hat{q}_t^{1-\alpha}]$ with a confidence level of $1-2\alpha$. For example, if $\alpha = 0.1$, the PI corresponds to an 80% confidence level ($1-2\alpha = 0.8$), implying that there is an 80% probability that y_t will lie within the interval.

2.3.3 Quantile Regression

QR, introduced by [KBJ78], is a probabilistic forecasting method that extends the Ordinary Least Squares framework by estimating conditional quantiles of the response variable. Unlike traditional mean-focused methods, QR provides insights into the uncertainty associated with different parts of the distribution.

Consider the linear model:

$$y_t = \mathbf{x}_t^{\top}\beta + \epsilon_t, \tag{2.27}$$

where y_t represents the dependent variable at time t, $\mathbf{x}_t$ is a K-dimensional vector of predictors, β is a K-dimensional vector of coefficients, and ϵ_t denotes

the error term. The objective of QR is to estimate the conditional quantile function, which represents the value below which a certain percentage α of the response variable y_t falls, given the predictor variable $\mathbf{x_t}$, where $\alpha \in (0, 1)$. This allows QR to model different parts of the price distribution, capturing the influence of predictors at various points across the distribution of y_t.

For a given α quantile level, the QR estimator $\hat{\beta}_\alpha$, is defined as the solution to the following optimisation problem:

$$\hat{\beta}_\alpha = \arg\min_{\beta} \sum_{t=1}^{N} \rho_\alpha \left(y_t - \mathbf{x}_t^\top \beta \right), \tag{2.28}$$

where $\rho_\alpha(u)$ is the check function given by:

$$\rho_\alpha(u) = u \left[\alpha - \mathbf{1}_{u<0} \right], \tag{2.29}$$

and $\mathbf{1}_{u<0}$ is an indicator function that equals 1 if $u < 0$ and 0 otherwise. This check function can also be expressed as:

$$\rho_\alpha(u) = \begin{cases} \alpha u & \text{if } u \geq 0, \\ (\alpha - 1)u & \text{if } u < 0. \end{cases} \tag{2.30}$$

which asymmetrically penalises positive and negative residuals, with weights α and $(1 - \alpha)$, respectively. This asymmetric penalty allows QR to estimate different quantiles, providing flexibility to capture the variability in y_t under various conditions, such as electricity price spikes during peak demand periods.

The QR estimator $\hat{\beta}\alpha$ has key asymptotic properties. It is *consistent*, meaning that as the sample size N increases, $\hat{\beta}\alpha$ converges to the true parameter vector $\beta\alpha$. Additionally, $\hat{\beta}\alpha$ is *asymptotically normal*, meaning that as N grows, the scaled difference between the estimated and true parameters converges to a multivariate normal distribution. Specifically, $\sqrt{N}(\hat{\beta}\alpha - \beta\alpha) \xrightarrow{d} \mathcal{N}(0, \mathbf{\Sigma}\alpha)$, where $\mathbf{\Sigma}\alpha$ is the asymptotic covariance matrix, expressed as:

$$\mathbf{\Sigma}_\alpha = \alpha(1 - \alpha) \left[f_{y_t|\mathbf{x}_t}(\hat{q}_t^\alpha)^2 \cdot (X'X)^{-1} \right] \tag{2.31}$$

with $\alpha(1 - \alpha)$ accounting for the variability in estimating the quantile level α, $f_{y_t|\mathbf{x}_t}(\hat{q}_t^\alpha)$ is the conditional density of y_t given predictors $\mathbf{x_t}$, and $(\mathbf{X'X})^{-1}$ is the inverse of the product of the design matrix with its transpose. The design matrix X contains the values of all predictor variables for each observation

$X = \{(\mathbf{x}_t)\}_{t=1}^{N}$, and its inverse accounts for the scaling and correlations.

QR is widely used in PEPF [Dud16], with [JPMN14] introducing a time-adaptive QR method for DAM prices that improve tail accuracy using exponential distribution fits, outperforming models like GARCH. Building on QR's versatility, [HBK$^+$16] applies QR to UK electricity prices, modelling sensitivities to factors like gas and coal prices across different quantiles for risk management. Further advancing QR's capabilities, [CS20] combines QR with a ResNet model to produce asymmetrical PIs, showing improved performance under price spikes. Expanding on this integration, [JBA22] introduces Ensemble CQR, which combines QR with CP to improve the reliability and sharpness of probabilistic forecasts, while [MN24] integrates QR with multiple SCP to generate robust PEPF that adapts to market uncertainties, and [LSTL23] integrates QR with LSTNet and SHAP feature selection, refining distributions through kernel density estimation for sharper probabilistic forecasts. Similarly, [ZF23] employed QR to capture non-linear relationships in price forecasts, optimising the reliability and sharpness of PIs without assuming a specific distribution. Finally, [MAM24] demonstrates QR's effectiveness in capturing price uncertainty in pay-as-bid markets under regional and climatic variations, while [XLWD24] integrated QR with LSTM, using multi-objective tuna swarm optimisation to offer accurate probabilistic forecasts under varying market conditions.

Quantile Regression Averaging

QRA, introduced by [NW15], utilises the combined outputs from M point forecasting models, $[f^1(\mathbf{x}_t), f^2(\mathbf{x}_t), \ldots, f^M(\mathbf{x}_t)]$, applying QR to to estimate quantiles $\hat{q}_t^{\alpha}$ and $\hat{q}_t^{1-\alpha}$, defining the lower and upper bounds of the expected price distribution. By leveraging the entire set of model outputs, QRA is capable of capturing the impact of predictors across different regions of the price distribution. This method captures both central tendencies and tail behaviours, allowing it to adapt to market fluctuations and provide robust forecasts. The predictors $\mathbf{x}_t$ can also include explanatory variables beyond point forecasts from these individual models. As such, QRA has gained prominence in PEPF literature due to its ability to generate both accurate and more reliable PIs.

QRA has been a leading approach in PEPF since its introduction by [NW15], where it combines point forecasts from multiple models and uses quantile regression to estimate conditional quantiles, consistently outperforming individual models. [MNW16] extends QRA with Factor QRA, using PCA to reduce

model complexity and improve accuracy. Further, [MSU23] and [MSU24] demonstrate Factor QRA's ability to capture uncertainty and enhance forecast reliability in electricity markets by leveraging latent factors. QRA's effectiveness in capturing price uncertainty is demonstrated by its ability to outperform traditional approaches and consistently provide more accurate PIs than simpler methods like bootstrapping [MUW20b], and historical simulation [UMW19a, MUW22].

While [SUW19] compares QRA with QR Machines and finds QR Machines to be more computationally efficient, QRA offers more reliable PIs. On the other hand, [CS20] benchmarks QRA against ResNet-QR, noting that ResNet-QR produces narrower PIs with higher coverage. Building on prior work, [UW21] introduces Lasso-Regularized QRA to improve forecast accuracy by selecting relevant predictors and reducing overfitting, while [Uni23b], [Uni23a], and [ZSMM24] demonstrate that variants like Smoothing QRA and Factor QRA further improve the sharpness and reliability of probabilistic forecasts through dimensionality reduction and smoothing techniques.

Exploring the integration of QRA with CP, [KZ21] demonstrates that while QRA effectively captures variability, CP improves the reliability of PIs. [Fér24] explores further integration of QRA and CP with methods like Quantile RF and Conformalised Quantile Regression (CQR), offering more refined PIs for PEPF. [MNWZ23] highlights how QRA enhances probabilistic forecasts by leveraging ensemble forecasts from LASSO and DNN models, though Distributional DNNs show improved accuracy with flexible distributions. Finally, [Jan24] compares QRA with Expectile Regression Averaging, showing that both methods improve forecast accuracy, with QRA outperforming Expectile Regression Averaging.

2.3.4 Conformal Prediction

CP offers a distribution-free framework for uncertainty quantification, providing a compelling alternative to QR and QRA. Originally introduced by [GVV98] and later extended by [VGS05] and [SV08], CP constructs PIs with valid coverage guarantees, regardless of the underlying data distribution. At the core of CP is the concept of a *nonconformity measure*, which quantifies how unusual or atypical a new data point $(\mathbf{x}_t, y_t)$ appears in relation to a dataset $\mathcal{Z}$. For regression tasks, a common nonconformity score is the absolute prediction error $A(\mathbf{x}_t, y_t) = |y_t - f(\mathbf{x}_t)|$, where $f(\mathbf{x}_t)$ is the predicted value. The nonconformity score measures how far the predicted value deviates from the true value. Other measures may be used depending on the specific model or application.

After computing nonconformity scores for all examples in $\mathcal{Z}$, CP constructs a PI for a new observation $\mathbf{x}_{\text{new}}$. For a candidate label y_{new}, the corresponding p-value is calculated as:

$$p(y_{\text{new}}) = \frac{|\{(\mathbf{x}_t, y_t) \in \mathcal{Z} : A(\mathbf{x}_t, y_t) \geq A(\mathbf{x}_{\text{new}}, y_{\text{new}})\}| + 1}{N + 1}.$$

This p-value reflects the proportion of past nonconformity scores that are greater or equal to the score for y_{new}. The PI for $\mathbf{x}_{\text{new}}$ is constructed by including all candidate labels y_{new} whose p-values exceed a predefined significance level α:

$$\Gamma_{1-2\alpha}(\mathbf{x}_{\text{new}}) = \{y_{\text{new}} : p(y_{\text{new}}) > \alpha\}.$$

CP provides a rigorous validity guarantee, ensuring that with probability at least $1-2\alpha$, the true response y_{new} will lie within the PI, $\Gamma_{1-2\alpha}(\mathbf{x}_t)$. This validity property holds regardless of the underlying data distribution, making CP effective when the distribution is unknown or difficult to model.

The PI width is determined by the quantile λ_t of the nonconformity scores where $\lambda_t = \text{Quantile}_\alpha\left(A_1, A_2, \ldots, A_N\right)$, for the α quantile level, with λ_t corresponding to the $(1-\alpha)(N+1)$-th smallest nonconformity score in the sorted list.

Although CP is a more recent application in time series compared to QR and QRA, it has shown strong performance in recent years. [TFBCR19] extends the CP framework to handle covariate shift through Weighted CP, ensuring valid PIs under distributional changes. Building on this, [JBA22] combines QR with CP in Ensemble CQR to sharpen PIs for probabilistic forecasts, while [CBL+23] offers a model-agnostic framework that integrates SCP and CQR for dynamic time series. Similarly, [BBGL24] enhances CP with NN ensembles and CQR to provide robust PIs for PEPF. Further innovations include [WFH+23], who integrates CP with generative transformers to produce reliable, narrow PIs for wind power forecasting, and [DBW23], who highlights CP's flexibility in transforming point predictions into valid probabilistic forecasts. Lastly, [LUW24] applies CP Averaging to refine PIs for EPF in the DAM.

Split Conformal Prediction

SCP is an inductive variant of CP that improves computational efficiency by dividing the dataset $\mathcal{Z}$ into a training set of size n_{train} and a calibration set of size n_{cal}, where $N = n_{\text{train}} + n_{\text{cal}}$. By restricting nonconformity score calculations to the calibration set, SCP reduces the need to recalculate scores across the

entire dataset, significantly cutting computational time and making it ideal for larger datasets. First, a model is trained on the training set to estimate the conditional expectation $f(\mathbf{x}_t)$. Nonconformity scores for the calibration set are then computed using the defined nonconformity score $A(\mathbf{x}_t, y_t) = |y_t - f(\mathbf{x}_t)|$.

To construct PIs for a new input $\mathbf{x}_{\text{new}}$, SCP selects λ_t corresponding to the $(1 - \alpha)(N + 1)$-th smallest nonconformity score computed from the calibration set, such that the PI for a new input $\mathbf{x}_t$ is given by:

$$\Gamma_{1-2\alpha}(\mathbf{x}_t) = [f(\mathbf{x}_t) - \lambda_t, f(\mathbf{x}_t) + \lambda_t]$$

providing a probabilistic guarantee that y_t lies within this range with a confidence level of $1 - 2\alpha$, assuming data exchangeability. In cases where the exchangeability assumption is violated, such as in time-series data like electricity prices, adaptations EnbPI and SPCI can better handle temporal dependencies.

[KZ21] applies SCP to generate PEPF for the DAM and IDM, demonstrating sharp and valid PIs, particularly in volatile markets. [HLT+22] combines SCP with QR for wind power interval forecasting, ensuring reliable PIs through Temporal Convolutional Networks. [CBL+23] introduces the MAPIE library, emphasising SCP's efficiency in generating probabilistic forecasts by splitting data into training and calibration sets, suitable for large datasets. [DZFG24] propose Online Sequential SCP-horizon, an adaptive SCP method tailored for non-stationary electricity prices, offering dynamically adjusted PIs over time. [Fér24] highlights SCP's ability to construct reliable PIs with guaranteed coverage, exploring extensions like online sequential split CQR for improved time-series adaptability. [MN24] extends SCP through a multiple-split approach to create multidimensional PIs, improving the models ability to capture market uncertainties. [JAVWVH24] introduces Split Conformal Distribution Regression Forests to improve SCP's efficiency and accuracy in wind power forecasting. [RBA24] applies SCP to day-ahead photovoltaic power forecasts, providing reliable PIs that improve forecast accuracy and minimise energy imbalances.

Ensemble Batch Predictive Interval

EnbPI, introduced by [XX21], builds on the CP framework, providing PIs with finite-sample coverage guarantees. Unlike traditional CP methods, EnbPI leverages ensemble learning and bootstrap sampling to generate reliable PIs without the need for separate calibration sets, making it particularly effective in time-

series forecasting applications with complex dependencies.

Utilising the dataset $\mathcal{Z}$, EnbPI generates B bootstrap samples $\mathcal{Z}^{(b)}$, where each sample trains a predictive model $f^{(b)}(\mathbf{x}_t)$, resulting in multiple separate models $\{f^{(1)}(\mathbf{x}_t), f^{(2)}(\mathbf{x}_t), \ldots, f^{(B)}(\mathbf{x}_t)\}$, with residuals calculated as $e_t^{(b)} = |y_t - f^{(b)}(\mathbf{x}_t)|$. Thus the interval $\lambda_t^{(b)}$ from these residuals, $\lambda_t^{(b)} = \text{Quantile}_\alpha(e_1^{(b)}, e_2^{(b)}, \ldots, e_N^{(b)})$, together with the average of the ensemble predictions $f^{(b)}(\mathbf{x}_t)$, forms the PI:

$$\Gamma_{1-2\alpha}(\mathbf{x}_t) = \left[\frac{1}{B}\sum_{b=1}^{B} f^{(b)}(\mathbf{x}_t) - \frac{1}{B}\sum_{b=1}^{B}\lambda_t^{(b)}, \frac{1}{B}\sum_{b=1}^{B} f^{(b)}(\mathbf{x}_t) + \frac{1}{B}\sum_{b=1}^{B}\lambda_t^{(b)}\right],$$

This construction allows EnbPI to offer valid coverage without retraining models, and by leveraging leave-one-out residuals, ensuring minimal computation.

EnbPI's scalability, reliable coverage, and ability to handle non-exchangeable time-series data make it ideal for high-uncertainty environments like PEPF, though its potential in this field remains underexplored. [JBA22] introduces EnbPI for generating probabilistic forecasts by combining ensemble learning with CP, offering reliable PIs without assuming a specific data distribution. [ZFG+22] applies EnbPI to EPF, constructing PIs by aggregating ensemble model predictions and recalibrating based on recent data, handling seasonality and trends effectively. [CBL+23] incorporates EnbPI into the MAPIE library, highlighting its use of block bootstrap resampling for dynamically updating prediction intervals in non-exchangeable time-series datasets. [DlTRM+24] applies EnbPI to the Mexican wholesale electricity market, using bootstrap ensemble estimators to capture market volatility and improve forecast accuracy, achieving a target coverage of 95%.

Sequential Predictive Conformal Inference

SPCI, introduced by [XX23], improves on EnbPI by dynamically adjusting PIs to accommodate new observations in time-series data. SPCI uses a sequential updating mechanism that adapts to evolving data distributions, ensuring reliable uncertainty quantification even in non-stationary environments like EPF.

Unlike static CP approaches, which recalibrate PIs using a fixed set of past data, SPCI updates non-conformity scores in real-time, leveraging the information from new residuals. For each new observation $(\mathbf{x}_t, y_t)$, SPCI adjusts the PIs based on the residuals ϵ_t. This feedback loop enables SPCI to refine its predictions continuously, ensuring that the intervals remain accurate and informative.

SPCI provides theoretical guarantees of asymptotic conditional coverage, ensuring that the probability of the true value y_{t+1} lying within the constructed interval converges to the desired confidence level $1-\alpha$ over time, such that:

$$\lim_{t\to\infty} \mathbb{P}\left(y_{t+1} \in \Gamma_{1-\alpha}(\mathbf{x}_{t+1}) \mid \mathbf{X}_t\right) \geq 1-2\alpha,$$

where $\mathbf{X}_t$ represents all information available up to time t.

The construction of the PI for a new input $\mathbf{x}_t$ is given by:

$$\Gamma_{1-2\alpha}(\mathbf{x}_t) = \left[f(\mathbf{x}_t) - \lambda_t, f(\mathbf{x}_t) + \lambda_t\right],$$

where λ_t is the quantile of the residuals up to time t. By using a sliding window of recent residuals and modelling their quantiles through QR Forests, SPCI efficiently captures changes in data behaviour, advancing time-series forecasting with theoretical coverage guarantees and producing narrower, more responsive PIs for reliable uncertainty quantification.

Much like EnbPI, despite its demonstrated usefulness to EPF, research on SPCI remains limited and underexplored. [XJX24] introduces MultiDimSPCI, extending SPCI to multivariate time series by constructing ellipsoidal prediction regions, ensuring valid coverage in high-dimensional, non-exchangeable time series. [ACT24] enhances SPCI with Conformal PID Control, which adapts PIs in response to distribution shifts and trends, improving long-term robustness and accuracy. [LXX24] presents SPCI-T, integrating the Transformer decoder to capture long-range dependencies, improving PI accuracy and coverage, especially for non-stationary and multi-step time series.

2.3.5 Other Probabilistic Forecasting Methods

Given the need to capture uncertainty in volatile markets, various probabilistic forecasting methods, including Bayesian approaches, Monte Carlo simulations, Distribution-Based forecasts, Monte Carlo, Bootstrap, and Historical Simulation are used to provide reliable PIs. [NW18] reviews these methods, assessing their accuracy and reliability for PEPF.

Bayesian approaches for probabilistic forecasting combine prior distributions with data to produce posterior distributions. [PS08] introduces a Bayesian approach for PEPF using multivariate skew t-distributions to capture time series features like skewness and heavy tails. [VJ10] applies Bayesian Neural

Network (BNN) for DAM EPF, providing full predictive distributions and reducing overfitting. [MS19] explores Bayesian inference for continuous time Auto-Regressive Moving Average (ARMA) models, improving probabilistic forecasts by accounting for volatility and seasonality. [CDZD19] uses Bayesian optimisation to enhance hybrid models, optimising hyperparameters for improved PEPF accuracy. [KK19] applies Bayesian stochastic volatility models with jumps, producing more accurate PIs. [GAG+19] combines data clustering with a Bayesian Recurrent Neural Network, improving forecast reliability in PEPF. [BMPV19] integrates Bayesian learning into DNNs for robust probabilistic forecasts in the DAM. [BMR+20] compares BNNs with Mixture Density RNNs, highlighting Bayesian inference for risk assessment. [BBC+20] applies Bayesian bootstrap QR for probabilistic forecasting of photovoltaic power, improving accuracy through resampled data. [BMVT21] discusses Bayesian Model Averaging, which improves forecast reliability by combining multiple model predictions. [YYT+23] and [NM24] present hybrid approaches, combining BNNs with Monte Carlo simulation and hierarchical frameworks for forecasting, improving accuracy by capturing causal relationships and uncertainties.

Bootstrap methods generate probabilistic predictions by resampling data or forecast residuals to create multiple PIs. [CK07] introduces bias-corrected bootstrap methods to construct PIs for AR models, addressing small-sample biases and improving forecast accuracy. [AGMRJS11] applies a bootstrap approach in seasonal dynamic factor models to enhance PEPF by resampling residuals. [WXW+13] integrates Bootstrap with Extreme Learning Machine to generate PI for PEPF, capturing forecast uncertainty. [KNC13] uses Bootstrap to construct PIs in NNs, offering narrower, more informative PIs but sometimes underperforming during high volatility. [TSEEK17] employs Bootstrap in a hybrid ensemble framework to generate PEPF, improving prediction stability. [UMW19a] explores residual-based bootstrapping to create more accurate PIs by resampling forecast residuals. [BBC+20] applies Bayesian bootstrap QR to improve probabilistic photovoltaic power forecasts by resampling data and enhancing quantile estimation. [BMVT21] uses bootstrap aggregation to reduce forecast variance in ensemble models for PEPF. [NZ22] employs bootstrapped errors to simulate multiple price trajectories in electricity auctions, enhancing the robustness of forecasts. [Nar22] uses Bootstrap for PEPF in the German BM, improving empirical coverage under market volatility. [ZF23] compares bootstrap methods with QR, showing that while both capture non-linear relationships, QR provides sharper PIs.

Distribution-Based Probabilistic Forecasts assume a specific probability distribution, such as Gaussian or Poisson. [WM08] and [JW10] compares parametric and semiparametric time series models, showing that kernel density estimators provide smoother and more reliable distribution-based probabilistic forecasts than traditional parametric models. [Ser11] uses GAMLSS to dynamically model the full distribution of electricity prices, offering flexible and accurate interval forecasts. [ZS18] extends the X-Model by incorporating supply and demand curves, simulating full time-dependent distributions and capturing extreme price events. [Dud18] introduces a Nadaraya-Watson estimator, generating asymmetric predictive densities that outperform traditional models in capturing price spikes. [MRRFJR18] applies Nadaraya-Watson-Kernel Density Estimator to estimate Beta distribution Probability Distribution Functions for DAM PEPF, offering a comprehensive view of forecast uncertainty. [DBdV$^+$19] presents a Two-Step Probabilistic Approach, providing PEPF for the BM with confidence intervals based on transition probabilities. [KSN23] employs echo state networks and copula processes to model price distributions, delivering well-calibrated PIs, especially in volatile markets. [MNWZ23] introduces Distributional DNNs that model the full distribution of prices, improving forecasting performance and risk management. [XLWD24] combines QR-LSTM with Kernel Density Estimator to estimate full probability distributions from quantile forecasts, enhancing forecast robustness. [LUW24] uses Isotonic Distributional Regression for postprocessing, significantly improving accuracy and outperforming traditional methods like QRA and CP in PEPF.

Monte Carlo approaches for probabilistic forecasting generate random samples to simulate possible outcomes. [WS10] employs Monte Carlo simulations to generate probabilistic forecasts for Auto-Regressive Moving Average with Exogenous input (ARMAX), GARCH, and Adaptive Wavelet NN models, improving accuracy through variance reduction techniques. [BRMD16] uses this method to model uncertainty in demand, fuel prices, and renewable output, generating multiple scenarios for medium-term price forecasts. [BRM16] extends this approach by simulating key risk factors to create a probabilistic risk profile for medium-term EPF. [YXS$^+$17] applies Monte Carlo within NNs to adjust weights through random sampling, producing reliable demand forecasts. [Tat18] uses a mean-reverting model to simulate price trends and spikes in Lithuania's electricity market, improving short-term volatility forecasts. [OLSK$^+$18] incorporates this technique in a hybrid model to simulate price outcomes and manage uncertainties in PEPF. [GWBS22] applies it to model uncertainty in drivers

like solar generation and gas prices. [NZ22] simulates multiple price trajectories for the DAM, optimising bidding strategies by accounting for uncertainty. [YKG+22] employs Sequential Markov Chain Monte Carlo to forecast supply curves, enhancing real-time adaptability and volatility management. [YYT+23] enhances the approach through a Bayesian network, capturing interdependencies and simulating electricity price distributions to improve forecast reliability in volatile markets.

Historical Simulation is a model-independent approach widely used for PEPF by constructing empirical PIs from past forecast errors. [WM08] applied this method to spot electricity prices by calculating sample quantiles from one-step-ahead prediction errors, effectively capturing market uncertainty without relying on parametric assumptions. Similarly, [NW15] used historical simulation for PEPF by constructing empirical PIs from ARX and Switched Non-linear ARX residuals. [UMW19a] emphasised its simplicity compared to bootstrapping while achieving similar performance in PEPF. [MUW20b] further applied residual-based historical simulation to model price volatility by leveraging the empirical distribution of past residuals. [Jan24] highlighted the method's robustness in capturing volatility and price spikes without parametric assumptions. Finally, [MN24] demonstrated its use in multidimensional forecasting, constructing PIs by adding historical forecast errors to point forecasts.

2.3.6 Overview

The evolution of uncertainty quantification methods for PEPF has shifted from traditional approaches such as Bootstrap, Historical Simulation, Monte Carlo, and Distributional-Based techniques, which remain essential for stress-testing and risk management, particularly in hybrid frameworks combining statistical and ML models. Bayesian methods were an early step forward, but since its introduction in 2015 by [NW15] and subsequent review in 2018 [NW18], QRA has dominated the field, as shown in Table 2.4. By using point forecasts from individual models as independent variables and observed spot prices as the dependent variable, QRA generates PIs without directly combining individual model intervals, avoiding compounded errors. This allows it to address aleatoric uncertainty by capturing cross-model variability, and epistemic uncertainty by mitigating biases inherent to individual models. Its ability to generate sharp and reliable PIs has made it a standard baseline for PEPF, forming the foundation for many enhanced techniques [MNW16].

Table 2.4: Literature on PEPF. Acronyms- Historical Simulation (HS), Monte Carlo (MC), Distributional Based (DB). The ✓is the approach presented herein.

Reference	**QR**	**QRA**	**CP**	**Bayesian**	**Bootstrap**	**DB**	**MC**	**HS**	**PEPF**	**Other**
Wan et al. 2013 [WXW+13]					✓				✓	
Khosravi et al. 2013 [KNC13]					✓				✓	
Jonsson et al. 2014 [JPMN14]	✓					✓			✓	
Nowotarski et al. 2015 [NW15]	✓	✓						✓	✓	
Maciejowska et al. 2016 [MNW16]		✓							✓	
Hagfors et al. 2016 [HBK+16]	✓								✓	
Dudek et al. 2016 [Dud16]	✓								✓	
Bello et al. 2016 [BRM16]							✓		✓	
Bello et al. 2016 [BRMD16]							✓		✓	
Tahmasebifar et al. 2017 [TSEEK17]					✓				✓	
Tat et al. 2018 [Tat18]							✓		✓	
Nowotarski et al. 2018 [NW18]	✓	✓		✓	✓		✓	✓	✓	
Ziel et al. 2018 [ZS18]	✓					✓			✓	
Dudek et al. 2018 [Dud18]						✓			✓	
Monteiro et al. 2018 [MRRFJR18]						✓			✓	
Osorio et al. 2018 [OLSK+18]							✓		✓	
Dumas et al. 2019 [DBdV+19]						✓			✓	
Kostrzewski et al. 2019 [KK19]				✓					✓	
Brusaferri et al. 2019 [BMPV19]				✓					✓	
Cheng et al. 2019 [CDZD19]				✓					✓	
Uniejewski et al. 2019 [UMW19a]		✓			✓			✓	✓	
Serafin et al. 2019 [SUW19]		✓				✓			✓	
Muller et al. 2019 [MS19]				✓					✓	
Ghayekhloo et al. 2019 [GAG+19]				✓					✓	
Marcjasz et al. 2020 [MUW20b]	✓							✓	✓	
Chaweewat et al. 2020 [CS20]		✓							✓	
Brusaferri et al. 2020 [BMR+20]				✓					✓	
Uniejewski et al. 2021 [UW21]		✓							✓	
Kath et al. 2021 [KZ21]		✓	✓						✓	
Bhatia et al. 2021 [BMVT21]				✓	✓				✓	
Narajewski et al. 2022 [Nar22]					✓	✓			✓	
Narajewski et al. 2022 [NZ22]					✓		✓		✓	
Maciejowska et al. 2022 [MUW22]		✓							✓	
Gabrielli et al. 2022 [GWBS22]							✓		✓	
Yildirim et al. 2022 [YKG+22]							✓		✓	
Yuanchen et al. 2023 [YYT+23]				✓			✓		✓	
Klein et al. 2023 [KSN23]				✓		✓	✓		✓	
Uniejewski et al. 2023 [Uni23b]		✓							✓	
Uniejewski et al. 20 [Uni23a]		✓							✓	
Marcjasz et al. 2023 [MNWZ23]		✓				✓			✓	
Liu et al. 2023 [LSTL23]	✓								✓	
Zhang et al. 2023 [ZF23]	✓				✓				✓	
Maciejowska et al. 2023 [MSU23]		✓							✓	
Nickelsen et al. 2024 [NM24]				✓					✓	
Monjazeb et al. 2024 [MAM24]	✓								✓	
De et al. 2024 [DlTRM+24]			✓						✓	
Xu et al. 20 [XLWD24]	✓					✓			✓	
Maciejowska et al. 2024 [MN24]	✓		✓					✓	✓	
Maciejowska et al. 2024 [MSU24]		✓	✓					✓	✓	
Zakrzewski et al. 2024 [ZSMM24]		✓							✓	
Feron et al. 2024 [Fér24]		✓	✓			✓			✓	
Janczura et al. 2024 [Jan24]		✓						✓	✓	
Brusaferri et al. 2024 [BBGL24]		✓	✓						✓	
Lipiecki et al. 2024 [LUW24]		✓	✓			✓			✓	
Dutot et al. 2024 [DZFG24]	✓		✓						✓	

Table 2.5: Literature on other probabilistic forecasting applications. Acronyms-Historical Simulation (HS), Monte Carlo (MC), Distributional Based (DB). The ✓is the approach presented therein.

Reference	QR	QRA	CP	Bayesian	Bootstrap	DB	MC	HS	PEPF	Other
Clements et al. 2007 [CK07]					✓		✓			✓
Yong et al. 2017 [YXS+17]							✓			✓
Tibshirani et al. 2019 [TFBCR19]			✓	✓						✓
Bozorg et al. 2020 [BBC+20]				✓	✓					✓
Xu et al. 2021 [XX21]			✓							✓
Jensen et al. 2022 [JBA22]	✓		✓							✓
Hu et al. 2022 [HLT+22]	✓		✓							✓
Foygel et al. 2022 [FBCRT22]			✓							✓
Zaffran et al. 2022 [ZFG+22]			✓							✓
Cordier et al. 2023 [CBL+23]			✓							✓
Xu et al. 2023 [XX23]			✓							✓
Wang et al. 2023 [WFH+23]			✓							✓
Subhankar et al. 2023 [GSB+23]			✓							✓
Dewolf et al. 2023 [DBW23]			✓	✓						✓
Xu et al. 2024 [XJX24]			✓							✓
Angelopoulos et al. 2024 [ACT24]			✓							✓
Lee et al. 2024 [LXX24]			✓							✓
Jonkers et al. 2024 [JAVWVH24]	✓		✓							✓
Renkema et al. 2024 [RBA24]	✓		✓							✓

More recently, CP has gained attention for its ability to provide valid PIs with minimal assumptions. Time-series adaptations such as EnbPI and SPCI have emerged, offering narrower intervals and robust uncertainty estimates. While traditional SCP ensures valid PIs under the assumption of data exchangeability, this assumption limits its applicability to time series due to temporal dependencies. EnbPI, by integrating ensemble learning with bootstrap sampling and eliminating the need for a separate calibration set, effectively manages both aleatoric and epistemic uncertainties. Its leave-one-out strategy refines uncertainty estimation, generating reliable PIs. SPCI extends these capabilities by dynamically updating residuals to reflect evolving data, capturing temporal dependencies and heteroskedasticity. With adaptive PIs and asymptotic conditional coverage, SPCI has proven especially valuable for PEPF in dynamic markets [XJX24]. However, CP remains in the early stages of application, with studies utilising CP being focused on non EPF areas, as shown in Table 2.5.

2.4 Trading and Strategy Optimisation

Efficient trading strategies are essential for integrating renewables and capitalising on arbitrage opportunities in volatile electricity markets, where real-time pricing and market uncertainties pose significant challenges. This section re-

views key optimisation techniques that increase financial performance in energy trading, focusing on heuristic methods (Section 2.4.1), Linear Programming (LP) (Section 2.4.2), Mixed Integer Linear Programming (MILP) (Section 2.4.3), and the integration of forecasts into trading methodologies (Section 2.4.5). When combined with accurate forecasting models, these methodologies enable traders to optimise bids, manage energy storage, and navigate market complexities with precision. By examining the strengths and limitations of these strategies, this section underscores their role in maximising profitability and aiding the stability of electricity markets.

2.4.1 Heuristic

In the domain of EPF and trading, heuristic methods have become a widely used approach for navigating the complexities of the DAM [ZCBD16, WBKK18] and BM [BCT18, LWB+19], as well as dual market trading [CLF+11]. These methods are valued for delivering good solutions within reasonable computational times, even if they do not guarantee optimal outcomes. In the fast-paced environment of electricity markets, where dynamics are constantly evolving, the ability to make rapid decisions is essential for maintaining competitiveness.

Designed to prioritise flexibility and speed, heuristic trading strategies are ideal for real-time applications where quick responses to market fluctuations are fundamental to optimising profitability. Their computational efficiency reduces the risk of missed opportunities in volatile markets, while their inherent flexibility enables them to incorporate new data and adjust to changing conditions. The simplicity of heuristic methods, with fewer parameters and simpler algorithms, also makes them accessible and easier to implement.

Probabilistic forecasting further improves heuristic strategies providing more flexible risk management. Instead of relying on single-point estimates, it provides a range of potential price outcomes, accommodating market uncertainty and volatility. This approach helps traders better understand market movements and tailor decisions, such as in quantile trading, where buy and sell signals are based on forecasted price quantiles [UW21]. These techniques optimise returns while mitigating exposure to adverse price swings [Uni23b].

Heuristic methods are often combined with ML models for improved accuracy, with ML forecasts feeding into heuristic optimisation to identify effective trading strategies [AB21]. This hybrid approach improves performance and adapt-

ability in dynamic market conditions. However, heuristic methods do not guarantee optimal outcomes, which may lead to missed profits. Additionally, they can be sensitive to initial conditions and parameter settings, and lack the rigorous theoretical guarantees of exact methods [MSU23], which can be a disadvantage in highly regulated or risk-averse environments.

2.4.2 Linear Programming

LP is a widely used optimisation tool in energy market trading, offering exact solutions to challenges like volatile prices, fluctuating demand, and operational constraints [PdSN11, Spi14]. LP efficiently solves large-scale optimisation problems, particularly where relationships between variables can be approximated as linear. It is commonly used for unit commitment, economic dispatch, and optimal power flow, optimising generation schedules to meet demand at the lowest cost while respecting capacity, ramp rates, and minimum up/down times, while also maintaining network constraints.

LP is most effective when both perfect and imperfect foresight are considered, with accurate forecasts improving profitability [SDJW09]. A key application of LP in electricity trading is maximising arbitrage revenue through Energy Storage Systems (ESS). By storing energy when prices are low and selling when prices are high, traders can profit from price differentials. LP models effectively manage storage operations by incorporating constraints on charging, discharging, and capacity limits, ensuring optimal timing to maximise profits [BPPE14]. LP models incorporate operational constraints and efficiency factors, such as the round-trip efficiency of storage devices, making them suitable for real-world applications. Even with imperfections, LP optimises trading strategies by focusing on peak price events [MFS15]. In ESS valuation and optimisation, LP balances charging and discharging to maximise profits while adhering to regulations and technical constraints, ensuring safe operation and improved efficiency [BLZ16].

While traditional LP provides exact solutions for optimal scheduling and cost minimisation, its effectiveness can be challenged by uncertainties like renewable energy and price fluctuations. Stochastic LP addresses this by integrating probabilistic scenarios, creating a more robust framework to maximise profits and mitigate risks from volatility. By leveraging historical data and probabilistic forecasts, this approach optimises trading strategies under uncertainty [YF17].

Multistage stochastic programming builds on LP by modelling future price curves

and energy storage states across multiple periods, enabling continuous adjustments to trading strategies. It accounts for operational constraints, degradation costs, and market fluctuations, allowing for dynamic decision-making as new information arises [SPE22]. While effective, LP struggles with non-linearities and large datasets, where ML models can improve scalability and performance.

2.4.3 Mixed Integer Linear Programming

MILP extends LP by incorporating integer variables, enabling the modelling of both continuous and discrete decision-making processes. This is particularly valuable for optimising trading strategies and energy management, where binary decisions (e.g., on/off status of generators) and fixed costs are common. MILP's ability to model both discrete and continuous elements allows for more accurate modelling of operational decisions under uncertainty, especially when decisions in the DAM impact subsequent actions in the BM.

This makes MILP highly effective in optimising price and quantity bids across both DAM and BM. While this may increase exposure to risk, its dynamic nature ensures efficient cycling of storage units, supporting agile trading strategies [KUZ+17]. To address computational challenges in non-convex problems, stepwise approximation methods convert mixed-integer quadratic programs into MILP formulations, balancing accuracy and efficiency [BGDJB19].

Incorporating network constraints and renewable energy sources further improve MILP's flexibility. Stochastic MILP frameworks can account for uncertainties in renewable generation and demand, improving reliability and profitability under volatile conditions [TG19]. In microgrid operations, MILP optimises energy management under uncertainties in sources, prices, and demand, using scenario generation and reduction for efficiency [GMM21].

MILP offers precise solutions to complex market constraints, enabling traders to maximise profits and minimise costs in changing market environments [CSMZ21]. It plays a key role in integrating renewable energy and storage systems, improving both efficiency and grid stability [MODJ23]. This makes MILP indispensable for sustainable electricity trading and risk management.

2.4.4 Other Trading Approaches

Recent advancements in trading strategies tackle the challenges posed by renewable energy and market volatility, with a summary of these approaches

Table 2.6: Literature on Trading Strategies. The ✓is the approach presented herein

Reference	Heuristic	LP	MILP	DP	RL	DAM	IDM	BM	Multiple
Sioshansi et al. 2009 [SDJW09]		✓						✓	
Pinto et al. 2011 [PdSN11]		✓				✓			✓
Connolly et al. 2011 [CLF+11]	✓								✓
Spisto et al. 2014 [Spi14]		✓				✓			
Bradbury et al. 2014 [BPPE14]		✓						✓	
Mcconnell et al. 2015 [MFS15]		✓						✓	
Jiang et al. 2015 [JP15]				✓				✓	
Berrada et al. 2016 [BLZ16]		✓				✓		✓	✓
Zafirakis et al. 2016 [ZCBD16]	✓					✓			
krishnamurthy et al. 2017 [KUZ+17]			✓			✓		✓	✓
Yu et al. 2017 [YF17]		✓				✓			
Wang et al. 2018 [WZ18]					✓			✓	
Wilson et al. 2018 [WBKK18]	✓					✓			
Bassett et al. 2018 [BCT18]	✓							✓	
Kin et al. 2019 [LWB+19]	✓							✓	
Brijs et al. 2019 [BGDJB19]			✓			✓		✓	✓
Tohidi et al. 2019 [TG19]			✓			✓		✓	✓
Gomes et al. 2021 [GMM21]			✓			✓		✓	
Cosic et al. 2021 [CSMZ21]			✓						
Uniejewski et al. 2021 [UW21]	✓					✓			
Abramova et al. 2021 [AB21]	✓					✓			
Silva et al. 2022 [SPE22]		✓				✓	✓	✓	✓
Uniejewski et al. 2023 [Uni23b]	✓					✓			
Maciejowska et al. 2023 [MSU23]	✓					✓	✓		
Kraft et al. 2023 [KRKB23]		✓				✓	✓		✓
Mercier et al. 2023 [MODJ23]			✓			✓			
Alghumayjan et al. 2024 [AHZ+24]				✓	✓	✓		✓	✓

provided in Table 2.6. Stochastic optimisation, often using multi-stage stochastic MILP, models uncertainties in prices and renewable generation to optimise trading decisions across the DAM and IDM. This approach assesses profitability and risk, balancing risk-neutral IDM trading for higher profits, while risk-averse strategies mitigate risks through DAM trading [KRKB23].

Dynamic Programming (DP) is also utilised to optimise sequential decision-making processes in electricity trading, breaking the problem into smaller sub-problems for different time periods. DP effectively incorporates uncertainties like fluctuating prices and demand, into a unified decision-making framework, helping develop optimal trading strategies amid market volatility [JP15].

Reinforcement Learning (RL) is gaining traction for real-time price arbitrage, excelling at adapting to non-stationary prices. By refining a value function through charging and discharging actions, RL captures immediate profits and historical insights, often outperforming traditional methods [WZ18]. The integration of RL with DP further improves decision-making by continuously updating strategies with real-time data, enhancing the efficiency and profitability of ESS [AHZ+24]. These methods highlight the growing sophistication of trading strategies, enabling optimised decisions across market conditions and risks.

2.4.5 Integration of Trading Strategies with Energy Storage

Integrating forecasting models into trading strategies is fundamental to optimising decision-making in electricity markets. Accurate EPF models predict future prices, enabling traders to refine strategies, optimise buy-sell decisions, and improve profitability. Combined with techniques like LP and MILP, these models improve trading performance. For example, quantile-based forecasts can help traders manage risks and increase profits, ultimately improving market efficiency by reducing volatility and increasing liquidity.

Integrating ESS with trading strategies improves energy dispatch by leveraging arbitrage opportunities, storing energy at low prices and discharging when prices rise. Advanced optimisation techniques maximize profitability by optimising charge-discharge schedules, relying on accurate price forecasts, especially in volatile renewable-driven markets. Beyond trading, ESS plays a key role in ancillary services like frequency regulation and voltage support, offering additional revenue streams and improving grid stability [WTGB17].

ESS, combined with renewables, balances supply-demand by storing excess energy during high production and discharging during low production, maximizing renewable energy utilization and providing essential services for grid reliability [SR16]. Optimised trading strategies have broader implications, such as influencing market prices, stabilising the grid, and improving the reliability of power systems. Case studies (e.g., [CLF+11], [Spi14], [BLZ16], [TG19], [MODJ23]) demonstrate the real-world effectiveness of integrating forecasting models with trading strategies, leading to significant economic benefits.

Despite challenges like battery degradation and high capital costs, advancements are making ESS more economically viable, with incentives like tax credits enhancing returns [LS22]. Effective degradation management through predictive maintenance further improves ESS lifespan and operational costs, boosting long-term profitability [GMI+22].

2.5 Conclusion

This chapter has provided a review of the key components involved in both EPF and trading strategies, with a focus on the DAM and BM. The analysis of electricity markets highlighted the complexity and volatility inherent in these environments, driven by factors such as fluctuating supply and demand, the

growing influence of renewable energy, and market-specific structures.

The review of predictive models illustrated the evolving landscape of forecasting methodologies, from traditional statistical approaches like ARIMA and GARCH to more sophisticated ML and DL models, such as RF, XGB, DNNs, and LSTM networks. While ML and DL models excel in capturing complex non-linear relationships and perform well in the more stable DAM, they struggle in the more volatile BM, where the need for adaptable models is greater.

Uncertainty quantification techniques such as QR, QRA, CP, EnbPI, SPCI, and ensemble methods have become vital tools for improving both the accuracy and reliability of forecasts amid growing market volatility. By offering not just point estimates but also probabilistic forecasts, these approaches help market participants manage risk more effectively and make informed decisions.

Finally, the integration of forecasting models with trading strategies, including heuristic methods, LP, and MILP, demonstrates the importance of optimising trading decisions to maximise profitability while managing risks. The inclusion of ESS and advanced optimisation techniques further emphasises the need for effective trading strategies to navigate the complexities of electricity markets.

In sum, this chapter lays the foundation for understanding the interplay between accurate forecasting, risk management, and optimised trading strategies, setting the stage for the methodological developments in the subsequent chapters. As renewable energy integration accelerates, the insights gained here will be fundamental in improving the markets' stability and efficiency.

Chapter 3

Day-Ahead and Balancing Market Datasets

This chapter introduces the datasets used for EPF in the DAM and BM, alongside a time series analysis of these datasets. The first section focuses on the structure and variables of each dataset, covering historical and forward-looking data. For the DAM, the dataset includes variables such as market prices, and the price predictors, wind and demand forecasts, providing insights into the short-term pricing mechanisms of electricity markets. The BM dataset emphasises real-time adjustments, with data capturing real-time imbalances, demand corrections, and the impact of renewable energy sources. These datasets are fundamental to developing forecasting models tailored to the unique operational environments of both markets.

In the second section, we conduct a comprehensive time series analysis on both datasets. This includes stationarity tests to determine the data's suitability for modelling, and autocorrelation and partial autocorrelation analyses to identify underlying patterns in the series. Additionally, serial correlation, heteroscedasticity, and multicollinearity diagnostics are performed to ensure the robustness of the data structure. Correlation analyses between variables help uncover significant relationships, while decomposition analysis breaks the series into trend, seasonal, and residual components. Finally, residual analysis is used to validate model performance, ensuring that residuals are free from autocorrelation and heteroscedasticity. These analyses provide a solid foundation for building accurate and reliable forecasting models capable of handling the complexities of both the DAM and BM.

3.1 Datasets

The integration of renewable energy sources like wind and solar has significantly transformed electricity markets. Accurate EPF is now crucial for maintaining grid stability, optimising trading strategies, and managing financial risks. In the I-SEM, which covers both Ireland and Northern Ireland, the increasing share of variable renewable energy has led to greater price volatility and forecasting challenges. To address these issues, the European electricity market, including the I-SEM, has been structurally harmonised through policy initiatives aimed at integrating renewable energy sources [VDVDH17]. This harmonisation enables market participants to engage in both forward and spot markets, providing them with the tools necessary to manage price risks and balance supply and demand in real time. These reforms support cross-border electricity trading, foster greater renewable integration, and enhance market efficiency across Europe [CMS22].

The datasets used in this research are sourced from the *Single Electricity Market Operator (SEMO[1])* platform, which operates the electricity market across the island of Ireland, and SEMOpx[2], the power exchange for the Day-Ahead and Intraday markets in Ireland. These datasets encompass historical and forward-looking data for the I-SEM from 2019 to 2022, a period marked by substantial renewable energy integration and heightened market volatility driven by the intermittency of renewables and evolving market structures. Key variables include historical market prices, demand forecasts, generation predictions, interconnector flows, and actual generation outputs. These variables are vital for understanding price fluctuations driven by both market fundamentals and the intermittent nature of renewable energy sources, allowing for a deep analysis of the dynamics in the DAM and BM.

For the DAM, prices are forecasted for the next 24 hours using hourly data, aligning with the market's daily settlement process. In the BM, where real-time imbalances are managed, the forecasting horizon is set at 8 hours with half-hourly granularity, aligning with the horizon of SEMO-published forecast data for key predictors (including renewable and demand forecasts). This distinction reflects the different operational needs of each market: the DAM focuses on pre-scheduling generation and consumption, while the BM corrects real-time deviations. The following provides a detailed breakdown of the dataset

[1]https://www.sem-o.com/
[2]https://www.semopx.com/market-data/

variables, preprocessing methods, and the reasoning behind the modelling approach. These processes ensure that the resulting forecasting models are both accurate and robust enough to handle the inherent uncertainties and volatility in modern electricity markets.

3.1.1 Day-Ahead Market

The DAM is a vital platform where electricity market participants—generators, suppliers, and large consumers—schedule their production or consumption one day ahead. In this auction-based system, participants submit bids for the next 24 hours, aiming to balance forecasted supply and demand. This is especially vital in markets with high renewable energy penetration, such as the I-SEM, where accurate scheduling minimises the need for costly real-time adjustments caused by the volatility of renewable generation. The DAM auction occurs daily at noon CET, determining prices for the next 24 hours based on submitted bids and system constraints.

The clearing process relies on the EUPHEMIA algorithm [EUP24], which matches supply with demand while accounting for complex bid types (e.g., block orders) and transmission constraints across interconnected regions. This results in market prices that reflect the marginal cost of meeting forecasted demand. As the integration of renewable sources like wind and solar increases, forecasting DAM prices has become a growing priority in both academic and industry settings. The unpredictability of renewable generation, particularly wind, complicates bidding strategies and market price formation, requiring advanced models that can account for the uncertainties linked to weather-dependent supply and fluctuating demand.

In this study, the primary objective is to forecast DAM prices for the next 24 hourly periods, using a combination of historical and forward-looking data to capture the key drivers of price volatility, as shown in Figure 3.1. These forecasts help market participants develop more informed bidding strategies, allowing them to better anticipate price fluctuations and optimise their positions. Bids for the DAM must be submitted by 11:00 AM on the day before delivery (D-1), and prices are set for the 24 hours of the following day (D), starting from midnight. Therefore, the forecast horizon begins from the time of bid submission and targets the 24 settlement periods for the next day.

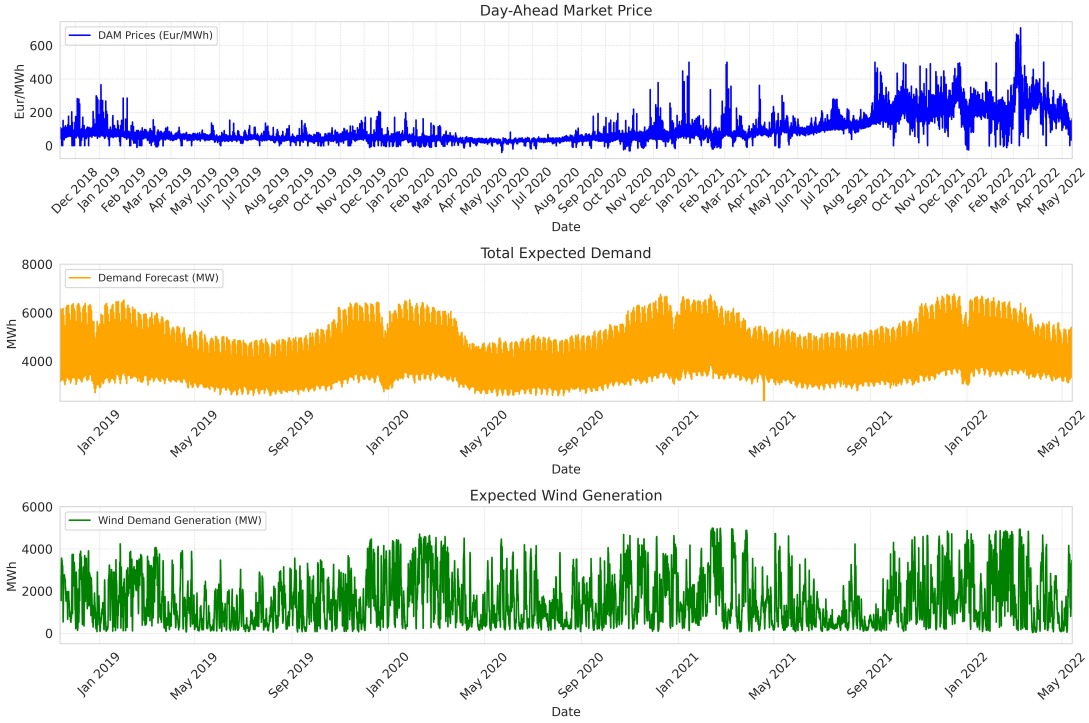


Figure 3.1: Time Series Plots for Day-Ahead Market Price, and the Predictors.

Forward-Looking Data

Forward-looking data is fundamental to predicting future conditions that will directly affect prices in the DAM. Incorporating forecasts enables the model to anticipate how wind generation and demand fluctuations will impact price outcomes. The forward-looking variables in the dataset include:

- *Demand forecast* (DF) in MWh: Demand fluctuations, driven by daily and seasonal patterns, are key determinants of price changes in the DAM. Accurate 24-hour demand forecasts from the TSO are essential for balancing supply and demand and guiding bidding strategies. These forecasts help anticipate consumption trends, directly influencing market prices. This is represented as: $[DF_t, ..., DF_{t+23}]$.
- *Wind Generation forecast* (WF) in MWh: The variability of wind energy significantly contributes to price volatility, especially in renewable-heavy markets like the I-SEM. We use 24-hour wind generation forecasts provided by the TSO to estimate future renewable supply. These forecasts allow the model to anticipate how changes in wind generation will influence market prices. The wind forecast is represented as: $[WF_t, ..., WF_{t+23}]$.

The notation (DF_t to DF_{t+23} and WF_t to WF_{t+23}) represents consecutive hourly

forecasts from the first hour (t) through 23 hours ahead ($t + 23$). Incorporating forward-looking wind and demand forecasts enables the model to capture the impact of renewable variability and demand fluctuations on future prices, aligning with the DAM's reliance on predicted supply and demand for bidding.

Historical Data

Historical data is fundamental to analysing past market behaviour and identifying trends that may influence future price movements. The selected historical variables reflect key market dynamics and align directly with the operation of the DAM. The dataset includes the following variables:

- *DAM Prices (DAMP)* in €/MWh: We use the DAM prices from the previous 7 days (168 hours) to capture short-term trends and shifts influenced by past market conditions, helping reveal patterns that may extend into future price movements. The historical price input is represented as:$[DAMP_{t-167}, ..., DAMP_{t-1}]$.
- *Past Wind Generation Forecasts (PWF)* in MWh: We use the past 7-days of the TSO's wind generation forecasts to capture the relationship between fluctuating wind power and price behaviour, to improve the model's ability to predict how wind variability affects prices. The wind generation data is represented as:$[PWF_{t-167}, ..., PWF_{t-1}]$.
- *Past Demand Forecasts (PDF)* in MWh: Again, we use the past 7-days of the TSO's demand forecasts to capture consumption patterns that drive price changes, improving the model's ability to capture how demand variability impacts prices. This is represented as:$[PDF_{t-167}, ..., PDF_{t-1}]$.

By incorporating historical data on DAM prices, wind generation, and electricity demand, the model is equipped to capture important market behaviours and trends. This approach reflects the operational realities of the DAM, where past performance is a strong indicator of future price dynamics.

Data Preprocessing

Preprocessing the DAM dataset is key to ensuring that the input data is consistent, accurate, and suitable for training models, and generating forecasts. Given the multiple data sources, varying time granularities, and the importance of capturing both historical and forward-looking trends, the following preprocessing steps were implemented:

- **Handling missing data** Missing values in the DAM dataset were filled using two approaches, depending on the type of variable. For predictors, the missing values were filled with the average of the last 6 periods on the same day, ensuring data gaps do not introduce bias or inconsistency. For price data, which rarely had missing values, gaps were filled using values from the same period one week earlier.
- **Time Zone and Daylight Savings Adjustments**: Since the DAM spans multiple regions with different time zones, all time points were standardised to a common time zone for consistency. Daylight savings adjustments were also applied to ensure uniformity in the temporal structure of the data, avoiding discrepancies that could impact forecasting accuracy.
- **Erroneous data detection and management**: The DAM is sensitive to price spikes and sudden shifts in demand caused by unforeseen events. Erroneous data in the predictors was either replaced with the mean of the previous 6 periods, similar to the method used for missing data or removed depending on the severity of the error.
- **Data Alignment**: The dataset is constructed to include only the data that would have been available at the time of prediction, preventing any future information from being used in training. Predictors like wind and demand forecasts are aligned with the DAM settlement periods based on their actual availability, ensuring that the model uses only the data accessible at the time of each prediction. This approach ensures realistic and robust forecasting.
- **Lag Feature Creation**: To capture the influence of recent trends on future prices, a rolling window was applied to variables such as DAM prices, wind generation, and electricity demand. For instance, historical DAM prices from $[DAM_{t-167}, ..., DAM_t]$ were used to model temporal dependencies, which are essential for accurate EPF.
- **Scaling/Normalisation of Data**: Min-Max scaling was applied to standardize the range of predictor variables. This ensures that no individual feature, such as DAM prices or demand and wind forecasts, disproportionately affects the model due to differences in scale.
- **Handling Seasonality**: Dummy variables for hour-of-day, day-of-week, and month-of-year were created to capture recurring patterns in electricity price, demand, and generation. This helps the model account for

cyclical variations, leading to more accurate forecasts that reflect seasonal trends.

These preprocessing steps are designed to align the DAM data with the operational realities of the market. By ensuring that the dataset is clean, synchronized, and representative of typical market behaviour, the forecasting model can generate more accurate and actionable price predictions.

3.1.2 Balancing Market

The BM is essential for maintaining the stability of the electricity grid in real-time by allowing the TSO to correct imbalances between supply and demand. Unlike the DAM, which schedules electricity a day in advance, the BM operates on a much shorter timescale, addressing fluctuations in generation and demand as they occur, typically every 5 minutes. This market is particularly critical in markets such as I-SEM given the high levels of variable renewable energy, such as wind and solar, where forecasting errors and sudden changes in output are common.

The BM's primary function is to resolve imbalances that arise from forecast errors, sudden demand shifts, or unplanned generation outages [OT19]. Market participants, including generators, demand response providers, and storage operators, submit bids to either increase or decrease their output, or adjust their consumption, to assist in balancing the system. These bids are structured as price-quantity pairs, offering flexibility in real-time. The TSO uses these bids to manage the grid's balance while minimising costs and maintaining system security.

Key features of the BM include:

- *Price-Quantity Pairs*: Participants submit bids detailing the price at which they are willing to adjust their generation or consumption. The TSO accepts these bids as necessary to balance the grid, compensating participants for deviations from their pre-notified schedules (Physical Notifications, PNs). This mechanism ensures that the market can quickly respond to real-time fluctuations in supply and demand.
- *Flagging and Tagging*: To ensure that the BM price accurately reflects the cost of balancing energy, the TSO uses a flagging and tagging process. This classification system distinguishes between balancing actions that correct supply-demand imbalances and those taken for system security reasons.

Only flagged energy-related actions influence the BM price, preventing non-energy tagged actions from distorting price signals.

- *De Minimis Acceptance Threshold*: A minimum threshold for energy actions, the De Minimis Acceptance Threshold ensures that only significant transactions impact the BM price. Actions below this threshold are excluded to prevent minor deviations from disproportionately affecting the market price.

The BM operates with high temporal resolution, determining prices based on five-minute imbalance pricing periods, which are then aggregated over the I-SEM's 30-minute settlement periods. This granularity allows the market to respond swiftly to real-time changes, making it indispensable for maintaining system balance in a renewable-heavy grid. The accuracy and timeliness of BM data are critical, as delays or inaccuracies can lead to inefficient balancing actions and increased system costs. For a more detailed discussion of the BM, including the intricacies of the flagging and tagging process, see [BD18].

In our problem setting, the target is the BM price (BMP) for the proceeding, open, 16 settlement periods.

$$y_t = [BMP_{t+2}, ..., BMP_{t+17}]. \tag{3.1}$$

The forecast horizon begins at $t + 2$ instead of t, as at time t when a forecast is being generated, BM periods t and $t + 1$ are closed. If a participant wants to adjust its BM commercial order data, it can only do so for BM periods $t + 2$ onwards, hence our focus on the $[t + 2, ..., t + 17]$ horizon.

The explanatory variables for forecasting BMP are categorised into historical and forward-looking data. Historical data captures past market behaviour and includes a rolling window of the 48 most recent BM prices, along with key features like BM volume, forecast deviations, interconnector flows, and physical notifications. These variables provide critical insights into recent trends and patterns that influence short-term price movements. Forward-looking data, such as forecasts for renewable generation, electricity demand, the Net Interconnector Schedule, and future DAM prices, given they are known in advance, are crucial for anticipating future BM prices. These forecasted values align with the BM settlement periods the model seeks to predict. Some variables, like DAM prices, are present in both historical and forward-looking datasets due to their significant impact on market dynamics. This integration of historical

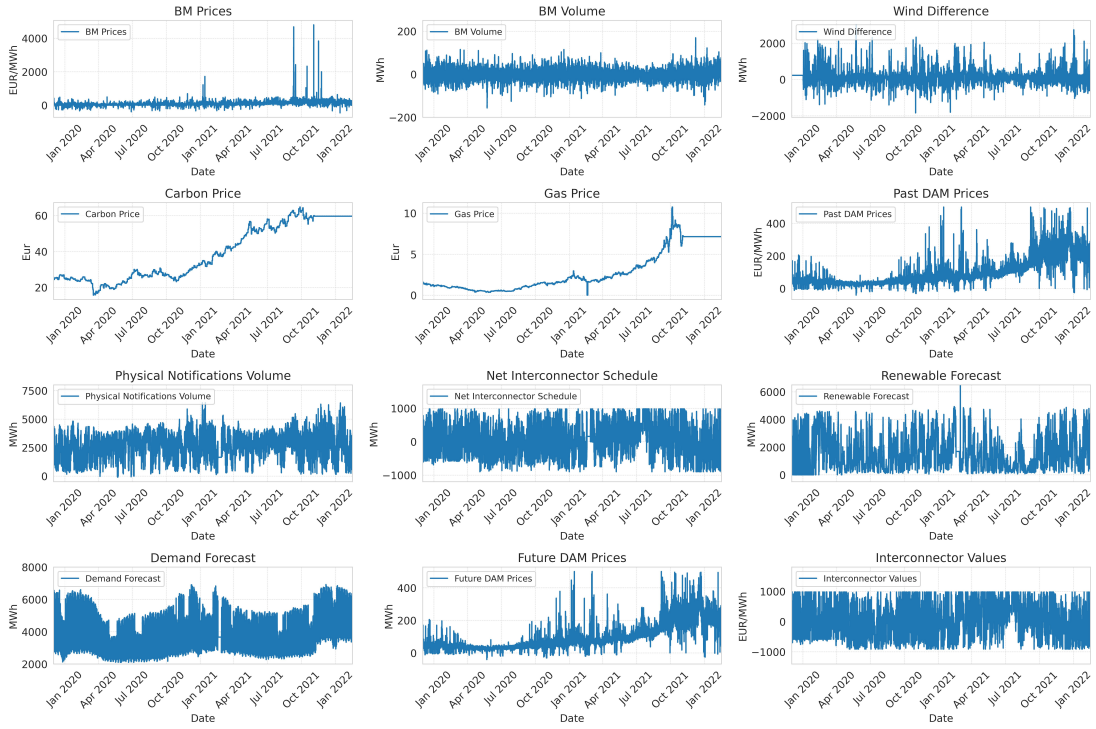


Figure 3.2: Time Series Plots for Balancing Market Price, and both Historical and Forward-Looking Predictors

and forward-looking data ensures a comprehensive approach to capturing the complex factors that drive BM prices. As illustrated in Figure 3.2, forecasting BM prices requires a comprehensive set of variables that capture both historical trends and future expectations, which are critical for accurate and reliable price predictions.

Forward-Looking Data

Forward-looking data includes forecast variables that provide an anticipatory view of market conditions. These forecasts are critical for predicting BM prices as they help account for expected changes in supply and demand that could influence real-time market dynamics.

- *Physical Notifications Volume* (PHPN) in MWh: This variable represents the sum of physical notifications—planned electricity generation or consumption adjustments—for the forecast horizon. It reflects anticipated deviations from scheduled market positions, influencing balancing actions: $[PHPN_{t+2}, ..., PHPN_{t+17}]$.
- *Net Interconnector Schedule* (PHI) in MWh: This is the projected interconnector schedule over the forecast horizon, showing expected cross-border

flows of electricity. Variations in these flows can impact BM prices by altering the available supply and demand balance: $[PHI_{t+2}, ..., PHI_{t+17}]$.

- *Renewable Forecast* (PHFW) in MWh: This variable provides the TSO's forecast of renewable energy output (e.g., wind and solar) for the forecast horizon. Accurate forecasts of renewable generation are crucial, as these sources are highly variable and directly impact the balancing needs: $[PHFW_{t+2}, ..., PHFW_{t+17}]$.
- *Demand Forecast* (PHFD) in MWh: The demand forecast indicates expected electricity consumption over the forecast horizon. It helps in predicting periods of market tightness or surplus, guiding balancing actions: $[PHFD_{t+2}, ..., PHFD_{t+17}]$.
- *DAM Prices* (DAM) in €/MWh: Forward DAM prices for the next 8 hours provide an indication of expected market conditions and can influence BM prices by signalling future supply-demand dynamics:$[DAM_{t+1}, ..., DAM_{t+16}]$.

Historical Data

Historical data is essential for understanding past market behaviour and identifying patterns that inform future price movements. The dataset includes a rolling window of historical data, incorporating key variables such as BM prices, volumes, and deviations between forecasted and actual conditions. The historical data considered includes:

- *Balancing Market Prices* (BMP) in €/MWh: Historical BM prices from the most recent 24 hours are used to capture recent market trends and price volatility. This data provides a basis for understanding how prices have responded to recent supply-demand imbalances: $[BMP_{t-51}, ..., BMP_{t-3}]$.
- *Balancing Market Volume* (BMV) in MWh: The BM volume data reflects the quantity of energy traded to balance the system over the past 48 periods. High volumes often indicate significant system imbalances, which can drive price volatility: $[BMV_{t-51}, ..., BMV_{t-3}]$.
- *Forecast Wind - Actual Wind* (WDiff) in MWh: This variable captures the difference between forecasted and actual wind generation over the most recent 48 settlement periods. Large deviations can cause significant price spikes in the BM, as unexpected shortfalls or surpluses in wind generation require immediate corrective actions: $[WDiff_{t-51}, ..., WDiff_{t-3}]$.

- *Interconnector Flows* (I) in MW: Interconnector flows represent electricity transmission between regions or countries, influencing market prices based on supply-demand dynamics. This data captures the cross-border electricity movements, potentially impacting BM prices: $[I_{t-50}, ..., I_{t-2}]$.
- *DAM Prices* (DAM) in €/MWh: Recent DAM prices provide a forward signal that influences the BM. High DAM prices may signal tight market conditions, which can affect real-time balancing needs. This variable includes DAM prices from the previous 24 hours, published at hourly granularity, adapted to half-hour settlement periods: $[DAM_{t-48}, ..., DAM_t]$.

Preprocessing

Preprocessing the BM dataset is essential to ensure the data is consistent, accurate, and suitable for training forecasting models. Given the real-time nature of BM operations and the complexities of multiple data sources, the following preprocessing steps were implemented:

- **Handling missing data**: In the BM dataset, missing values were handled according to the variable type. For BM prices, missing values were replaced with values from the same period one week earlier, preserving market cycles. For predictors such as imbalance volumes and demand, the average of the previous 6 periods on the same day was used, ensuring continuity without distorting short-term patterns.
- **Real-Time Alignment and Availability**: Since the BM operates in real-time, aligning variables to the precise time they become available is vital. Predictors like imbalance volumes and interconnector values were synchronised with BM settlement periods to avoid any look-ahead bias, ensuring that only data available at the time of prediction was used.
- **Data Aggregation**: The TSO provides aggregated data from 5-minute to 30-minute intervals. This aggregation was verified within the dataset to ensure consistency and prevent discrepancies due to different aggregation methods.
- **Outlier Detection and Treatment**: BM prices are prone to volatility due to imbalances and sudden operational changes. Instead of removing price spikes, they were retained to capture significant market behaviour. However, extreme values in predictors, such as demand and renewable forecasts, were capped or replaced using the median of the previous 3 periods

to minimise the impact of erroneous data while preserving genuine market signals.

- **Lag Feature Creation**: Lagged features for all BM predictors, such as BM prices, BM volumes, wind difference etc. were generated to capture short-term dependencies crucial for BM forecasting. Given the high frequency of BM fluctuations, lag windows were set to up to 48 half-hour intervals (one day) to account for recent trends influencing the current period.
- **Data Resampling for Consistency**: All variables were resampled to a consistent 30-minute interval, matching BM settlement periods. This was essential as various predictors update at different intervals. Resampling ensured uniform alignment with BM prices and prevented mismatches in the data.
- **Scaling/Normalization**: Min-Max scaling was applied to BM prices and predictors to standardize their range. This mitigated the effect of extreme values while maintaining the relative importance of each predictor, ensuring that price spikes did not disproportionately impact the model.
- **Handling Seasonality and Intraday Patterns**: BM prices and imbalances display strong intraday patterns, particularly during peak demand periods. Dummy variables were created for half hourly period, day-of-week, and month of the year to account for these recurring patterns, helping the model capture the cyclical nature of the BM.
- **Time Zone and Daylight Savings Adjustments**: To avoid discrepancies due to time shifts, all data points were aligned to a common time zone. Daylight savings adjustments were applied to ensure consistency across the dataset.
- **Lead/Lag Indicators**: DAM prices are unique as they serve both as historical variables and known future indicators. Historical DAM prices from the past 24 hours, published at hourly granularity, are adapted to the BM's half-hourly intervals. This allows the model to leverage recent trends while incorporating forward-looking DAM prices for the next 8 hours, improving predictions of BM prices.
- **Feature Engineering**: New features, such as the difference between forecasted and *WDiff*, were engineered to capture more complex relationships in the data and improve model accuracy.

These preprocessing steps are designed to align the BM data with the real-time operational demands of the market. By ensuring that the dataset is clean, synchronised, and representative of typical market behaviour, the forecasting model can generate more reliable and actionable price predictions.

3.2 Analysis

In the realm of EPF, the preparation and integrity of the underlying data is fundamental towards accurate forecasting. The accuracy of forecasting models depends not only on sophisticated algorithms but also on ensuring the data meets key statistical properties. This chapter explores the methodologies used to analyse and preprocess the DAM and BM datasets.

The analysis begins by addressing the stationarity of the datasets, an essential prerequisite for most time series models, using stationarity tests such as the Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) to verify that the time series data exhibits constant statistical properties over time, thus mitigating the risk of unreliable forecasts due to trends or volatility shifts. Following stationarity verification, the temporal dependencies within the data are scrutinised through Autocorrelation Function (ACF) and Partial Autocorrelation Functions (PACF) functions, which are helpful in identifying patterns and lags in the data, informing the selection of appropriate modelling techniques.

Beyond foundational checks, the analysis addresses additional issues like serial correlation, heteroscedasticity, and multicollinearity using diagnostic tools such as the *Ljung-Box* and *Breusch-Godfrey* tests for serial correlation, the *Breusch-Pagan* and *White* tests for heteroscedasticity, and the Variance Inflation Factor (VIF) and *Condition Index* for multicollinearity, along with *Pearson Correlation Coefficient* and *Spearman Rank Correlation* for correlation analysis. These methods ensure that datasets meet the assumptions needed for reliable forecasting. The chapter also employs STL, breaking down time series into trend, seasonality, and residual components to reveal deeper insights into market behaviour. Lastly, residual analysis ensures the models capture key data patterns, ensuring forecast robustness.

3.2.1 Stationarity Test

In time series analysis, stationarity is a fundamental property where the statistical characteristics of a series, such as mean, variance, and autocorrelation, remain constant over time. Non-stationary data, which may exhibit trends, seasonality, or varying volatility, can lead to inaccurate and misleading results in forecasting models, particularly in the more volatile BM. Ensuring stationarity in these datasets is fundamental to producing reliable forecasts, as forecasting techniques often assume that the underlying data are stationary. To assess and address potential non-stationarity in the DAM and BM datasets, two widely used tests were applied: the ADF, introduced in [DF81], and the KPSS Test, introduced in [KPSS92].

Augmented Dickey-Fuller Test

The ADF test is used to assess whether a time series contains a unit root, indicating non-stationarity. The ADF test evaluates the null hypothesis (H_0) that it is stationary, using a model that incorporates differenced terms, a trend component, and lagged values to account for autocorrelation. The test statistic is then compared against critical values at standard significance levels; if the statistic is more negative than the critical value, the null hypothesis of non-stationarity is rejected. Applying the ADF test to both DAM and BM datasets ensures that the data is suitable for further statistical modelling by either confirming stationarity or guiding the need for differencing, reducing the risk of spurious results due to non-stationarity.

Kwiatkowski-Phillips-Schmidt-Shin Test

The KPSS test complements the ADF test by reversing the null hypothesis, assuming stationarity unless proven otherwise. While the ADF test checks for unit root non-stationarity, the KPSS test is particularly useful for assessing trend stationarity, which the ADF may not fully capture. The KPSS model decomposes the time series into a deterministic trend, a random walk, and a stationary error term, expressed as $y_t = \beta_0 + \beta_1 t + r_t + \epsilon_t$, where y_t is the observed series, β_0 is the intercept, $\beta_1 t$ represents the trend, and $r_t = r_{t-1} + u_t$ captures the non-stationary behaviour. The null hypothesis (H_0) assumes the variance of the random walk component is zero, implying stationarity around a trend; rejecting H_0 suggests non-stationarity. If the KPSS statistic exceeds critical values, it indicates non-stationarity; values below critical thresholds support stationarity.

By applying both ADF and KPSS tests, a comprehensive evaluation of unit root and trend stationarity is achieved.

Table 3.1: ADF and KPSS Test Results for DAM and BM Datasets

Dataset	Test	Statistic	p-value	Conclusion
DAM	ADF	-4.4239	0.00027	Stationary
	KPSS (Initial)	18.5197	0.01	Non-stationary
	KPSS (Differenced)	0.0095	0.1	Stationary
BM	ADF	-10.5306	0.000917	Stationary
	KPSS (Initial)	24.1871	0.01	Non-stationary
	KPSS (Differenced)	0.0079	0.1	Stationary

The test results, as shown in Table 3.1, provide a detailed assessment of the stationarity for both the DAM and BM datasets. The initial ADF tests indicated that both datasets were stationary without differencing, with highly negative test statistics and low p-values. However, the KPSS test identified non-stationarity, suggesting the presence of a trend. After differencing the data, the KPSS test confirmed that both the DAM and BM series became stationary. This divergence highlights the importance of using both tests to capture different aspects of stationarity, ensuring that unit root and trend stationarity are addressed before further modelling.

3.2.2 Autocorrelation and Partial Autocorrelation Analysis

Autocorrelation analysis is a fundamental tool in time series analysis, used to identify dependencies within data, revealing how past prices and predictors influence future observations. Two key methods used are the ACF [Yul27] and PACF [Ner71], which help in understanding price movements, seasonal patterns, and determining appropriate lag structures for forecasting models.

Autocorrelation Function

The ACF measures the correlation between a time series y_t and its past values y_{t-k} at different time lags k. This function is particularly useful for identifying repetitive patterns, such as seasonality, by indicating how much past values influence future observations. The ACF at lag k is defined as:

$$\rho_k = \frac{\sum_{t=k+1}^{T}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{T}(y_t - \bar{y})^2} \tag{3.2}$$

where y_t represents the value of the series at time t, $\bar{y}$ is the mean of the series, and T is the total number of observations. The ACF measures how much current values of a time series are influenced by past values, revealing repeating trends or cycles in electricity demand or prices.

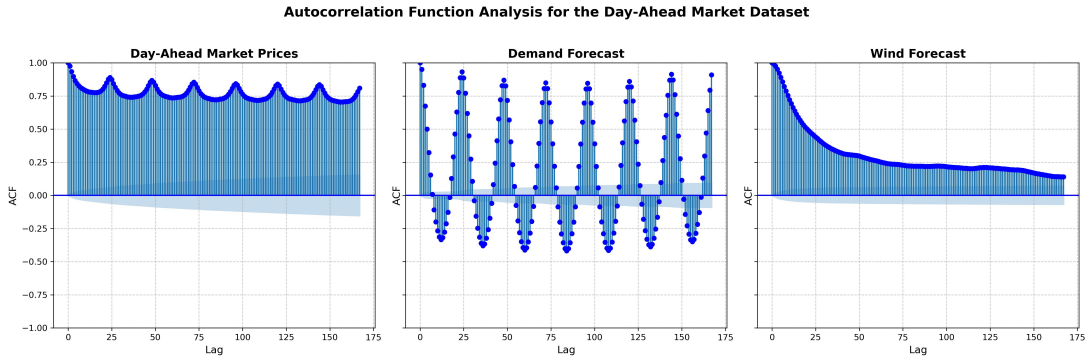


Figure 3.3: Autocorrelation Plots for the DAM Prices and predictors

The ACF was computed for DAM and BM prices, along with their respective predictors. In the DAM dataset, a lag of 167 periods (hourly data) captures nearly a full week of dependencies, while the half-hourly BM data uses 48 lags to capture daily cycles. The ACF for DAM prices (Figure 3.3) reveals strong periodicity driven by regular demand and supply fluctuations. Demand forecasts show similar periodic spikes, with faster decay, indicating shorter-term dependencies compared to prices. In contrast, wind forecasts show weaker autocorrelation, reflecting their volatility and lower predictability.

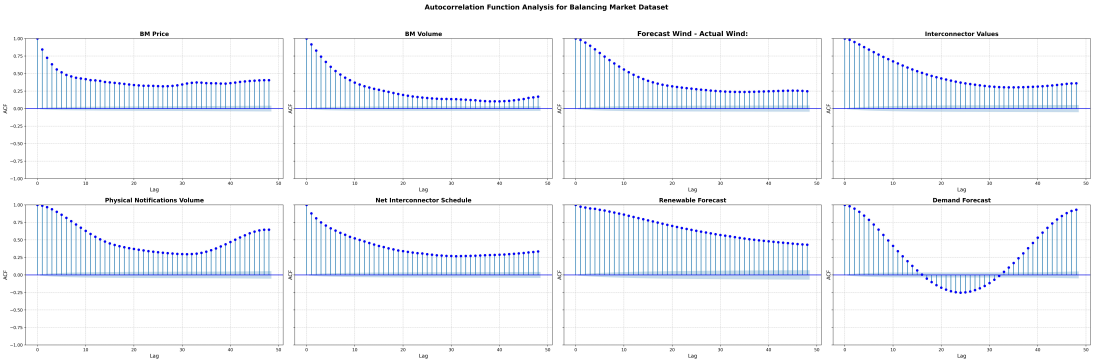


Figure 3.4: Autocorrelation Plots for the BM Prices and predictors

For the BM dataset (Figure 3.4), most predictors show rapid decay in autocorrelation, reflecting the BM's real-time responsiveness. The Demand Forecast exhibits seasonal patterns, while Physical Notifications Volume decays more slowly, indicating stronger long-term dependencies. This contrast between short-term volatility and longer-term patterns highlights the complexity of BM mar-

ket dynamics. Renewable forecasts exhibit a consistent decline, indicating their fluctuating nature, while wind forecasts also show rapid decline, underscoring the high variability of wind generation and the need for adaptive models.

Partial Autocorrelation Function

The PACF provides additional insights by measuring the correlation between the time series y_t and its lagged values y_{t-k}, controlling for the effects of intermediate lags. Unlike the ACF, which captures both direct and indirect effects of lagged values, the PACF isolates the direct impact of a particular lag on the current value of the series. This is particularly useful for identifying the specific lags that should be included in AR models.

The PACF at lag k is obtained by fitting an AR model of order k to the series and then examining the coefficient β_k, which represents the partial autocorrelation at lag k. Mathematically, for a time series y_t, the PACF is computed as the coefficient β_k in the autoregressive model:

$$y_t = \lambda + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_k y_{t-k} + \epsilon_t \tag{3.3}$$

where λ is a constant and ϵ_t represents the residual. The PACF is particularly useful in determining the number of lags to include in an AR model, as significant PACF values at certain lags indicate the presence of a direct effect.

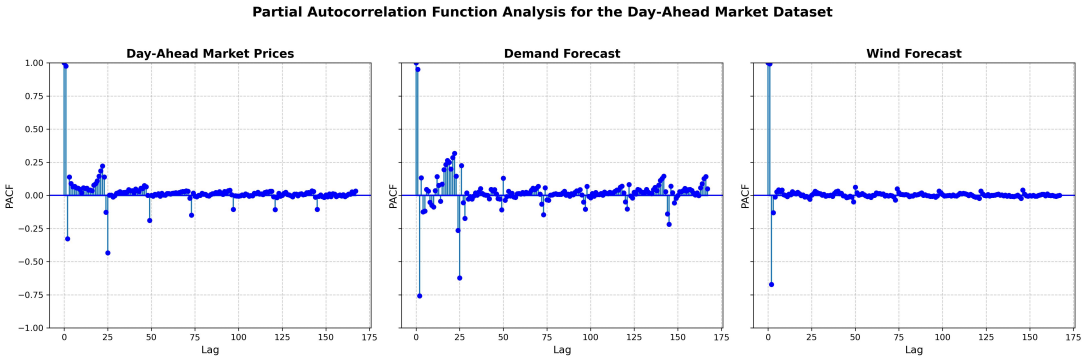


Figure 3.5: Partial Autocorrelation Plots for the DAM Prices and predictors

The PACF was calculated for the same variables to complement the ACF analysis. For DAM prices (Figure 3.5), the spike at lag 1 indicates strong short-term autocorrelation, with moderate peaks at lags 24, 48, and 72, reflecting daily cycles. The Demand Forecast shows a similar spike at lag 1 but diminishes quickly, highlighting the influence of recent hours. Wind forecasts follow a similar pattern, with strong dependence on the most recent data, and minimal influence

beyond lag 1.

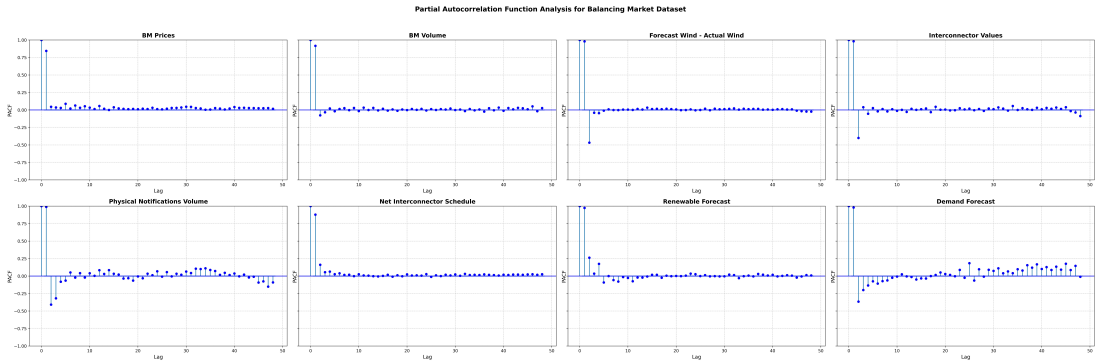


Figure 3.6: Partial Autocorrelation Plots for the BM Prices and predictors

In the BM dataset (Figure 3.6), PACF results confirm strong short-term dependencies, especially at lag 1 for price-related variables. Recurring patterns at half-hour intervals suggest cyclical behaviour in the market. Predictors like Physical Notifications Volume and Demand Forecast show significant short-term dependencies, reinforcing the need to focus on recent data for BM forecasting. These observations emphasize the importance of focusing on shorter lags in forecasting models, as the influence of past values diminishes quickly over time.

The ACF and PACF results highlight the importance of recent observations, guiding the use of short-term AR models. Seasonal patterns, such as daily and weekly cycles, suggest incorporating periodic components into models, while volatile predictors like wind forecasts, where autocorrelation declines rapidly, require adaptive or rolling-window models. This analysis provides critical insights into the temporal dependencies driving DAM and BM market dynamics, ensuring the development of forecasting models that are responsive to market conditions and capable of more accurate predictions.

3.2.3 Serial Correlation Tests

Serial correlation, or autocorrelation, occurs when the residuals in a time series model are correlated with each other over time, which can lead to inefficient estimates, incorrect inferences, and unreliable models for EPF in the DAM and BM. To address this, we employ the LASSO model, which is well-suited for high-dimensional data. To test for serial correlation in the residuals of the LASSO model, we use the *Ljung-Box Test*, introduced in [LB78], and the *Breusch-Godfrey Test*, introduced in [Bre78] and extended by [God78]. These tests help diagnose whether any autocorrelation remains after accounting for the model's

structure, ensuring the adequacy of the forecasting models.

Ljung-Box Test

The *Ljung-Box Test* assesses whether the residuals of a model are uncorrelated across multiple lags, indicating that the model has captured the time-dependent patterns. The Ljung-Box test statistic Q is calculated as:

$$Q = T(T+2)\sum_{k=1}^{m}\frac{\hat{\rho}_k^2}{T-k} \tag{3.4}$$

where T is the total number of observations, $\hat{\rho}_k$ represents the sample autocorrelation at lag k, and m is the number of lags being tested. The null hypothesis H_0 posits that there is no significant autocorrelation at any lag up to m, i.e. $H_0 : \hat{\rho}_1 = \hat{\rho}_2 = \cdots = \hat{\rho}_m = 0$. If the p-value from the test is small, we reject the null hypothesis, indicating that significant autocorrelation exists in the residuals, and the model has not fully captured the time-dependent patterns.

Breusch-Godfrey Test

The *Breusch-Godfrey Test* detects higher-order serial correlation in residuals, which may extend across multiple lags. Unlike tests that only assess first-order autocorrelation, the Breusch-Godfrey Test allows for testing of higher-order serial correlation, making it particularly useful in time series data where dependencies may not be limited to just one previous period. The test involves running an auxiliary regression in which the residuals ϵ_t from the original model are regressed on their own lagged values up to a specified lag m, along with the independent variables from the original model. This auxiliary regression takes the form:

$$\epsilon_t = \alpha + \beta_1\epsilon_{t-1} + \beta_2\epsilon_{t-2} + \cdots + \beta_m\epsilon_{t-m} + \mu_t \tag{3.5}$$

where the coefficients $\beta_1, \beta_2, \ldots, \beta_m$ represent the influence of past residuals on the current residual ϵ_t, with error term μ_t.

The null hypothesis of the test states that there is no serial correlation in the residuals up to lag m, which is expressed as $H_0 : \beta_1 = \beta_2 = \cdots = \beta_m = 0$. If the test returns a low p-value, we reject the null hypothesis, indicating the presence of serial correlation. The test statistic follows a chi-square distribution with m degrees of freedom, and if the residuals show significant autocorrelation, it may suggest model misspecification, meaning the model may need further

refinement to account for the serial dependencies in the data.

The Ljung-Box test results (Table 3.2) for the DAM model at lags 24, 48, 96, and 167 all show high p-values, well above 0.05, indicating no significant autocorrelation in the residuals. This suggests that the model effectively captures the temporal dependencies, and the residuals behave independently across the tested time lags. Similarly, the Breusch-Godfrey test shows no evidence of serial correlation at any of the tested lags, with all p-values above the conventional threshold. These results confirm that the DAM model adequately accounts for serial dependencies, resulting in robust and independent residuals.

Table 3.2: Ljung-Box and Breusch-Godfrey Test Results for DAM and BM Models

Dataset	Lag	Ljung-Box (p-value)	Breusch-Godfrey (p-value)
DAM	24	0.8994	0.1762
	48	0.6690	0.4045
	96	0.7131	0.1982
	167	0.8109	0.1574
BM	16	0.0748	0.9725
	32	0.1583	0.8760
	48	0.4339	0.5509
	96	0.7571	0.9052

In the BM model, the Ljung-Box test shows marginal evidence of autocorrelation at lag 16 (p-value = 0.0748), suggesting some short-term dependencies may remain. However, at higher lags (32, 48, and 96), the p-values rise significantly, indicating that the model performs well in capturing dependencies over longer periods, with no significant autocorrelation detected. The Breusch-Godfrey test results further support this conclusion, with consistently high p-values across all lags, indicating no significant serial correlation in the residuals. This suggests that, despite the slight short-term autocorrelation detected in the Ljung-Box test, the BM model proves capable in accounting for serial dependencies across various lags.

3.2.4 Heteroscedasticity Tests

Heteroscedasticity occurs when the variance of the error terms in a regression model is not constant across all levels of the independent variables. In time series data heteroscedasticity often emerges due to market volatility, shifts in demand, or fluctuations in renewable energy outputs, leading to periods of high and low variability in prices. Addressing heteroscedasticity is important,

as it violates the assumption of constant variance, a key foundation of many statistical models. If left uncorrected, heteroscedasticity can result in inefficient parameter estimates and unreliable hypothesis testing. To diagnose the presence of heteroscedasticity in the residuals of the LASSO model, we employ two tests: the *Breusch-Pagan Test* [BP79] and the *White Test* [Whi80].

Breusch-Pagan Test

The *Breusch-Pagan Test* is used to detect heteroscedasticity in the residuals of a regression model, which occurs when the variance of the residuals depends on the independent variables. The test involves running an auxiliary regression where the squared residuals $\hat{\epsilon}_t^2$ from the original model are regressed on the independent variable vector $\mathbf{x}_t$. This auxiliary regression is expressed as:

$$\hat{\epsilon}_t^2 = \alpha + \boldsymbol{\beta}\mathbf{x}_t + u_t \tag{3.6}$$

where the coefficient vector $\boldsymbol{\beta}$ measures the relationship between the variance of the residuals and the independent variables, and u_t represents the error term in the auxiliary regression.

The null hypothesis of the test states that the variance of the residuals is constant, meaning that none of the independent variables influence the residual variance. This is expressed as $H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$. If the test produces a low p-value, we reject the null hypothesis, indicating that heteroscedasticity is present. The test statistic, known as the Lagrange Multiplier (LM) statistic, follows a chi-square distribution with k degrees of freedom, where k is the number of independent variables. If heteroscedasticity is detected, it suggests that the model may need refinement to handle non-constant variance in the residuals.

White Test

The *White Test* is a versatile method for detecting heteroscedasticity in a regression model, without assuming any specific form of the relationship between the residual variance and the independent variables. Unlike the Breusch-Pagan Test, which tests for a linear relationship, the White Test can detect both linear and non-linear forms of heteroscedasticity. The auxiliary regression includes the independent variables, their squared terms, and interaction terms to cap-

ture potential non-linear effects:

$$\hat{\epsilon}_t^2 = \alpha + \boldsymbol{\beta}\mathbf{x}_t + \gamma_1\mathbf{x}_t^2 + \gamma_2(\mathbf{x}_i\mathbf{x}_j) + u_t$$

where $\hat{\epsilon}_t^2$ represents the squared residuals, $\mathbf{x}_t$ are the independent variables, $\mathbf{x}_t^2$ are their squared terms, and $\mathbf{x}_i\mathbf{x}_j$ are interaction terms capturing non-linear relationships. This allows the test to identify more complex forms of heteroscedasticity.

The null hypothesis of the White Test states that the variance of the residuals is constant, or homoscedastic, H_0 : The variance of the residuals is homoscedastic (constant). If the test statistic, which follows a chi-square distribution, is significant, the null hypothesis is rejected, indicating the presence of heteroscedasticity. This suggests that the model's residual variance depends on the independent variables, requiring further adjustments or refinements in the model.

Table 3.3: Breusch-Pagan and White Test Results for DAM and BM

Dataset	Test	LM Statistic	p-value	F-statistic	p-value	Result
DAM	Breusch-Pagan	1855.85	0.0	398.29	0.0	Heteroscedastic
	White	2366.72	0.0	-	-	Heteroscedastic
BM	Breusch-Pagan	1320.28	0.0	114.31	0.0	Heteroscedastic
	White	4986.68	0.0	-	-	Heteroscedastic

For the DAM, the *Breusch-Pagan Test* (Table 3.3) yielded an LM statistic of 1855.85 with a p-value of 0.0, strongly rejecting the null hypothesis of homoscedasticity. This suggests that the residual variance is systematically related to one or more independent variables, indicating significant heteroscedasticity. The F-statistic of 398.29 further corroborates these findings, implying that heteroscedasticity in the DAM model could lead to inefficient estimates and distorted inference, ultimately affecting the model's reliability in price forecasting.

Similarly, for the BM, the Breusch-Pagan Test yielded an LM statistic of 1320.28 and a p-value of 0, again rejecting the null hypothesis of homoscedasticity. The F-statistic of 114.31 with an extremely low p-value supports this conclusion, suggesting that heteroscedasticity is a significant issue in the BM model. This heteroscedasticity is likely due to the BM's high volatility, driven by dynamic supply-demand imbalances. Correcting for heteroscedasticity using robust standard errors or generalized least squares could improve model accuracy in both markets.

The *White Test* results for both DAM and BM confirm the presence of significant

heteroscedasticity, with test statistics of 2366.72 and 4986.68 respectively, and p-values of 0.0. These results align with the findings of the Breusch-Pagan test, further confirming that the residual variance is not constant. For DAM, the heteroscedasticity likely arises from market sensitivity to demand fluctuations, historical price movements, and renewable generation variability. In the BM, the high test statistic reflects the market's inherent volatility, driven by factors such as wind forecast errors, load imbalances, and interconnector flows.

Addressing heteroscedasticity by incorporating advanced techniques like QR or ML models can significantly improve forecast reliability and performance. QR allows for a more detailed understanding of different parts of the price distribution, making it particularly useful in volatile markets like DAM and BM. Additionally, transformations on the data such as a logistic or square root transformation of the dependent variable, or leveraging ML models such as XGB or RF, which are less sensitive to assumptions about residuals, enables the capture of complex, non-linear relationships in the data. These approaches, tailored to handle heteroscedasticity, ensure more reliable and accurate EPF.

3.2.5 Multicollinearity Tests

Multicollinearity occurs when two or more independent variables in a regression model exhibit a high degree of correlation, which can result in inflated standard errors, making coefficient estimates unreliable. In the context of EPF for the DAM and BM, this can obscure relationships between predictors and the dependent variable, leading to unstable models. Therefore, addressing multicollinearity is crucial for producing accurate, interpretable, and reliable models by improving coefficient precision, enhancing predictor significance, and reducing overfitting. To assess multicollinearity in the DAM and BM datasets, two diagnostic tools were employed: the VIF and the Condition Index, as introduced in [Sne83].

Variance Inflation Factor

The VIF measures how much the variance of a regression coefficient is inflated due to multicollinearity. It assesses the linear relationship between a predictor x_t and the other predictors, and is calculated as:

$$\text{VIF}_t = \frac{1}{1 - R_t^2} \tag{3.7}$$

where R_t^2 represents the proportion of variance in x_t explained by the other predictors. A VIF above 10 generally signals problematic multicollinearity, which may require the removal or combination of predictors.

Condition Index

The Condition Index detects multicollinearity stemming from interactions among multiple variables. While the VIF focuses on individual predictors, the Condition Index evaluates overall multicollinearity using the eigenvalues of the design matrix $X'X$, calculated as:

$$\kappa_t = \frac{\lambda_{\text{max}}^{1/2}}{\lambda_t^{1/2}} \tag{3.8}$$

where λ_{max} is the largest eigenvalue, and λ_t is the t-th eigenvalue. A Condition Index above 30 typically indicates severe multicollinearity [BKW05], suggesting that some variables may need to be removed or combined to improve model stability.

Table 3.4: VIF and Condition Index Results for DAM and BM

Dataset	Variable	VIF	Condition Index
DAM	Past Prices	1.16	1.50
	Wind Forecast	1.02	1.00
	Demand Forecast	1.18	1.17
BM	Past Prices	1.92	1.00
	BM Volume	1.22	1.36
	Forecast Wind - Actual Wind	1.29	1.93
	Carbon Price	5.32	2.02
	Gas Price	8.13	2.20
	DAM Prices	21.39	2.85
	Physical Notifications Volume	4.95	3.22
	Net Interconnector Schedule	1.73	3.64
	Renewable Forecast	5.98	5.07
	Demand Forecast	2.91	13.18
	Interconnector	2.61	7.58

The VIF and Condition Index results (Table 3.4) for the DAM indicate no significant multicollinearity issues. The variables *Past Prices*, *Wind Forecast*, and *Demand Forecast* all have VIF values close to 1, suggesting minimal correlation with other predictors. These results are corroborated by the low Condition Index values, all of which are well below the threshold of 30, indicating that the

model is well-conditioned and capable of providing stable and reliable coefficient estimates.

In contrast, the BM shows some signs of multicollinearity. The VIF values for DAM Prices (21.39) and Gas Prices (8.13) suggest strong multicollinearity, particularly for DAM Prices, which greatly exceeds the typical VIF threshold of 10. Additionally, Carbon Price and Renewable Forecast exhibit moderate VIF values of 5.32 and 5.98, respectively, indicating some degree of multicollinearity but within acceptable limits. The Condition Index values further emphasize potential concerns, particularly with Demand Forecast (13.18), Interconnector (7.58), and Renewable Forecast (5.07), which show higher shared variance among variables. Despite these moderate multicollinearity levels, these variables are retained in the model to capture important market dynamics, particularly in the BM market where interconnected factors like demand, renewable energy, and external prices are essential to model accuracy.

Although these levels of multicollinearity do not pose an immediate threat to the model's performance, they are monitored during model tuning and further analysis to ensure they do not adversely affect the predictive capabilities and interpretability of the models utilised.

3.2.6 Correlation Analysis

Correlation analysis is fundamental to understanding the relationships between variables within the DAM and BM datasets. By analysing the connections between each market's predictors, correlation analysis aids in the development of accurate prediction models and supports informed decision-making in market operations. This study utilises two key methods: the Pearson Correlation Coefficient [Pea95], which measures linear relationships, and the Spearman Rank Correlation [Spe61], which assesses monotonic relationships. Together, these methods offer a comprehensive view of the associations present in the data.

Pearson Correlation Coefficient

The Pearson Correlation Coefficient, r, measures the strength and direction of the linear relationship between two continuous variables, defined as:

$$r = \frac{\sum_{t=1}^{n}(\mathbf{x}_t - \bar{\mathbf{x}})(y_t - \bar{y})}{\sqrt{\sum_{t=1}^{n}(\mathbf{x}_t - \bar{\mathbf{x}})^2}\sqrt{\sum_{t=1}^{n}(y_t - \bar{y})^2}} \tag{3.9}$$

where $\mathbf{x}_t$ and y_t are individual data points at time t, $\bar{x}$ and $\bar{y}$ are the means of $\mathbf{x}$ and y, and n is the number of observations. The coefficient r ranges from -1 to 1, with $r = 1$ indicating a perfect positive linear relationship, $r = -1$ indicating a perfect negative linear relationship, and $r = 0$ indicating no linear relationship. Pearson's correlation is ideal for assessing linear relationships, which are common in market price dynamics.

The Pearson Correlation Coefficient is calculated for pairs of key variables in the dataset, including DAM prices, BM prices, and their predictors. Both historical and forward-looking data are analysed to understand how past and predicted values relate to each other. A correlation matrix is generated to visualise the strength and direction of linear relationships across all variables. The significance of the correlations is assessed using p-values to determine which correlations are statistically significant. The results are interpreted to identify strong positive or negative correlations that might indicate key drivers of price movements or potential issues with multicollinearity.

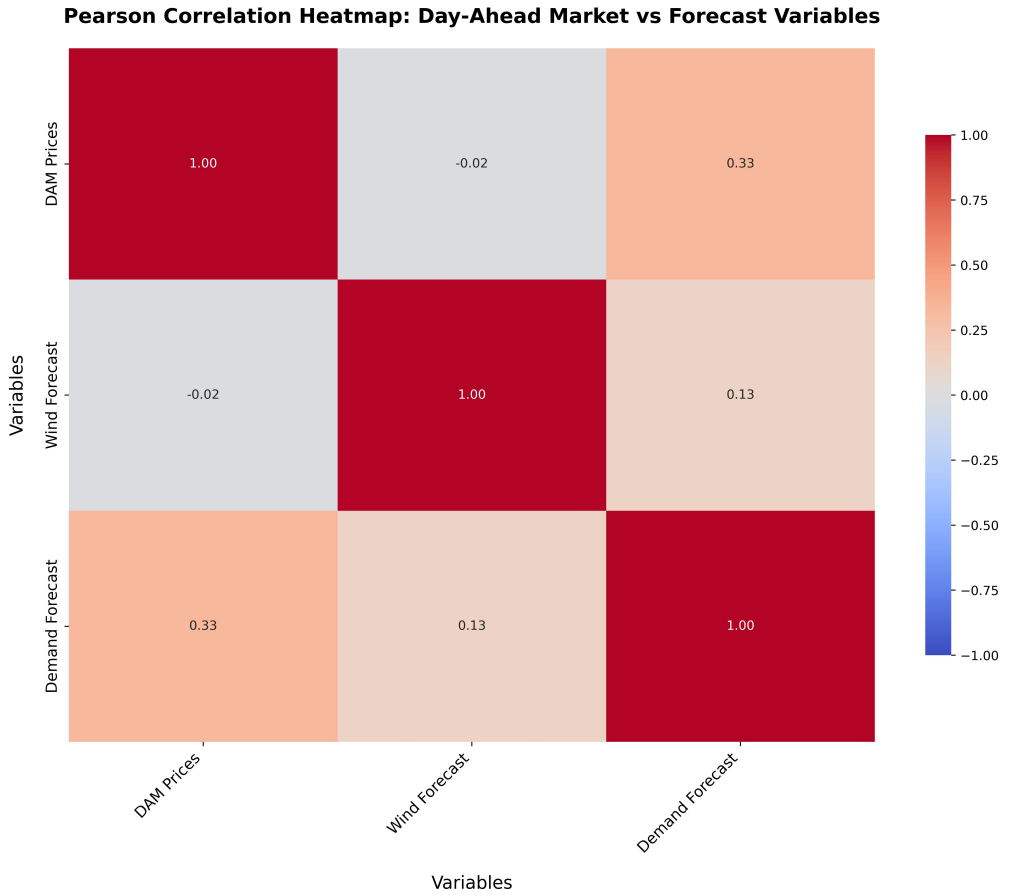


Figure 3.7: Pearson Correlation Plots for the DAM Prices and predictors

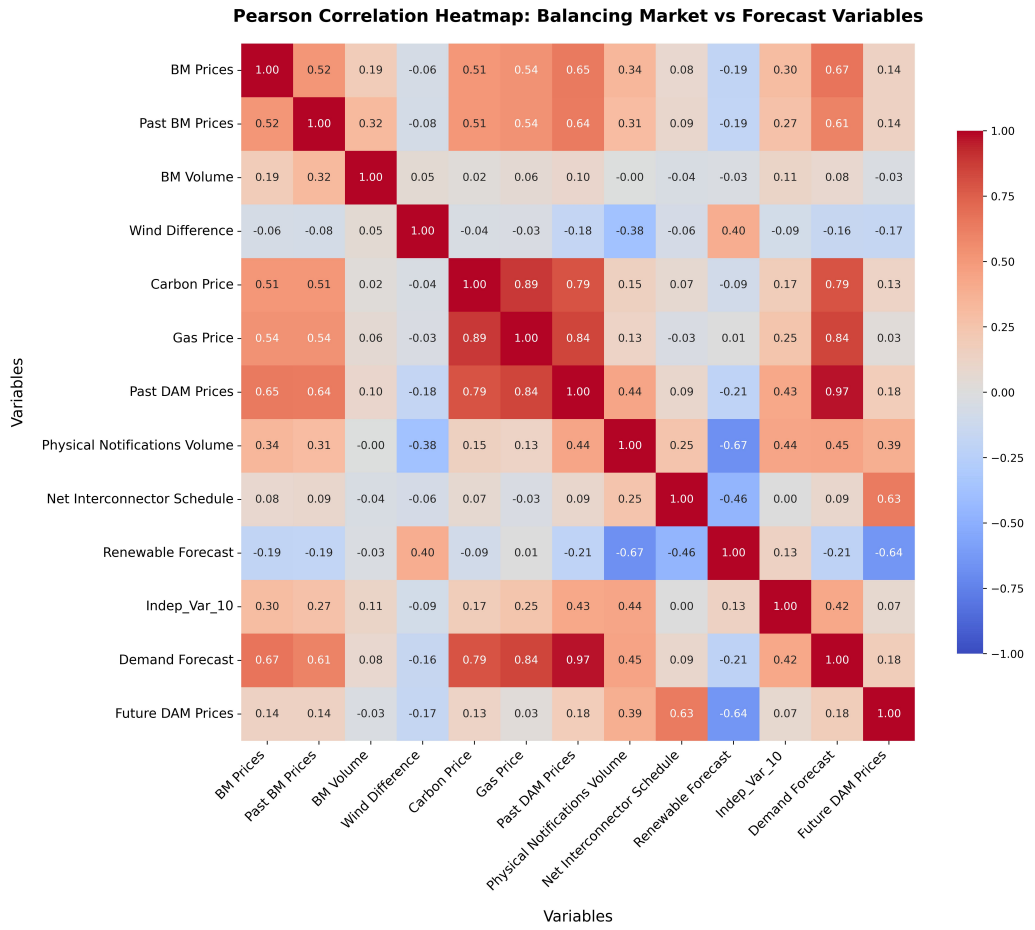


Figure 3.8: Pearson Correlation Plots for the BM Prices and predictors

The Pearson correlation matrix for the DAM predictors, as shown in Figure 3.7, reveals several notable relationships. A moderate positive correlation (0.33) exists between *DAM Prices* and *Demand Forecasts*, indicating that higher demand forecasts tend to drive up prices. However, there is a very weak negative correlation (-0.02) between *DAM Prices* and *Wind Forecasts*, suggesting an insignificant relationship between these variables, as wind energy may not strongly influence DAM pricing. In contrast, the BM (Figure 3.7), displays more complex dynamics, with *BM Prices* showing a moderate positive correlation (0.52) with *Past BM Prices*, indicating that past market prices help predict future prices. *BM Prices* are also positively correlated with *Gas Prices* (0.54) and *Carbon Prices* (0.51), reflecting the impact of fuel and emissions costs on the BM. Notably, *BM Prices* have a strong correlation (0.67) with *Demand Forecasts*, highlighting the importance of real-time demand expectations in BM pricing. Additionally, a weak negative correlation (-0.19) between *BM Prices* and *Renewable Forecast* indicates that higher renewable forecasts may reduce the need for

expensive balancing actions, thus lowering prices. *Demand Forecasts* also correlates strongly with *Past DAM Prices* (0.97) and with both *Gas Prices* (0.84) and *Carbon Prices* (0.79), emphasising the influence of fuel costs on market prices. The relationship between *Physical Notifications Volume* and *Renewable Forecasts* shows a significant negative correlation (-0.67), suggesting that higher renewable forecasts reduce the volume of physical trades in the BM. These findings demonstrate the strong interplay between demand, fuel prices, and renewable forecasts in shaping both DAM and BM prices.

Spearman Rank Correlation

The Spearman Rank Correlation, denoted by ρ, is a non-parametric measure that evaluates the strength and direction of the monotonic relationship between two variables. Unlike the Pearson correlation, which assesses only linear relationships, Spearman's correlation determines how well the relationship between two variables can be described by a monotonic function. This makes it particularly useful for data that may not exhibit a linear relationship or does not follow a normal distribution. Spearman Rank Correlation is calculated as:

$$\rho = 1 - \frac{6 \sum_{t=1}^{n} d_t^2}{n(n^2 - 1)} \tag{3.10}$$

where d_t represents the difference between the ranks of each pair of observations at time t, and n is the number of observations. Like r in 3.9, ρ ranges from -1 to 1, with values near 1 indicating a strong positive monotonic relationship, and values near -1 indicating a strong negative monotonic relationship.

The Spearman Rank Correlation was calculated for the same variables as in the Pearson analysis, including *DAM prices*, *BM prices*, and their predictors, to capture potential non-linear relationships. Comparing Spearman and Pearson coefficients reveals discrepancies that highlight non-linear dynamics. P-values assess the significance of these correlations, refining the understanding of variable relationships where Pearson showed weak or no linear correlation. In the DAM, the correlation between *DAM prices* and *demand forecasts* is moderately strong and positive ($\rho = 0.39$), indicating that as *demand forecasts* increase, *DAM prices* tend to rise as well, as shown in Figure 3.9. Conversely, the weak negative correlation between *DAM prices* and *wind forecast* ($\rho = -0.07$) suggests a minimal inverse relationship, likely due to the variability and intermittency of wind power. The *demand forecast* also shows a weak positive correlation

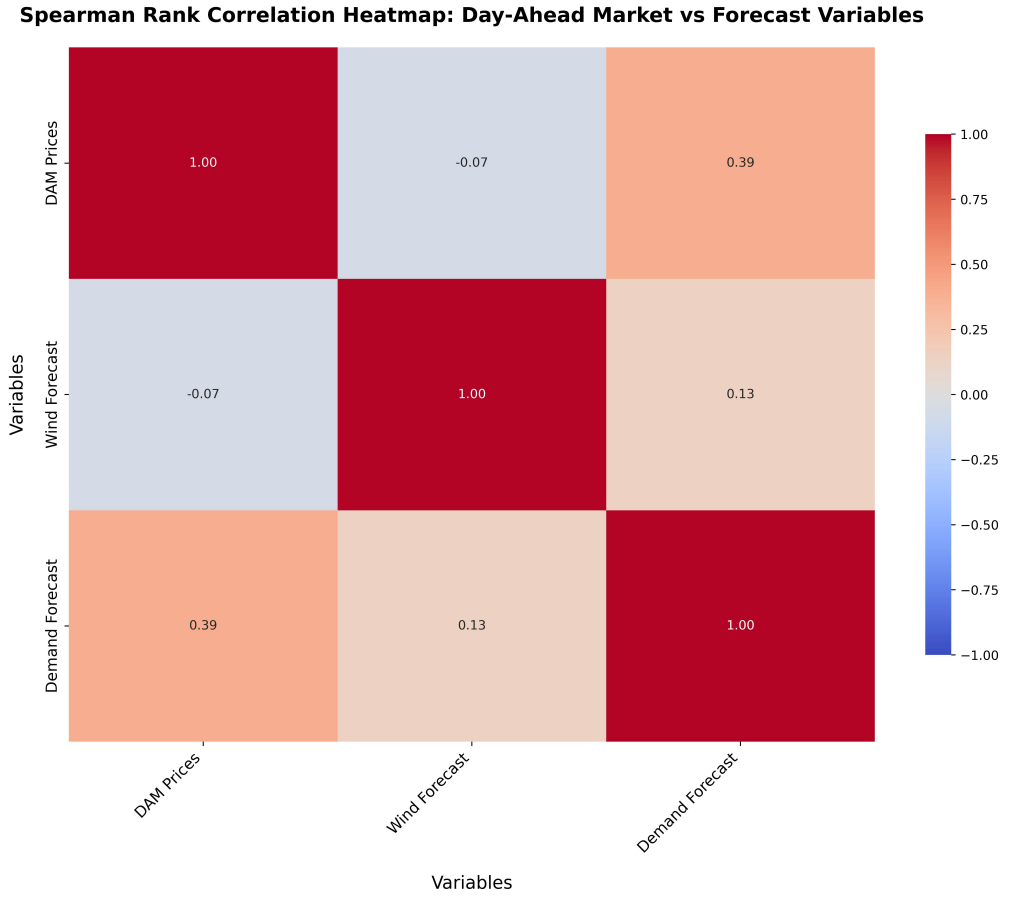


Figure 3.9: Spearman Rank Correlation Plots for the DAM Prices and predictors

with the *wind forecast* ($\rho = 0.13$), implying a slight tendency for demand and wind forecasts to increase together. Given that wind power represents only a portion of total demand, this relationship is not unexpected, though not particularly strong. Overall, these correlations indicate that in the DAM, *DAM prices* are more strongly driven by *demand* than by *wind forecasts*, consistent with our Pearson correlation analysis.

In the BM, the results show stronger correlations across several variables (see Figure 3.10). *BM prices* have a high positive correlation with *past BM prices* ($\rho = 0.73$), underscoring the importance of historical prices in predicting current BM prices. Similarly, both *gas prices* and *carbon prices* exhibit strong positive correlations with BM prices ($\rho = 0.67$ and $\rho = 0.65$, respectively), reflecting the influence of fossil fuel markets on BM dynamics. Notably, *BM Prices* have a strong Pearson correlation (0.67) with *Demand Forecasts*, with the relationship even more pronounced in the Spearman Rank Correlation, with ($\rho = 0.77$), suggesting that while the linear relationship is strong, the monotonic relationship

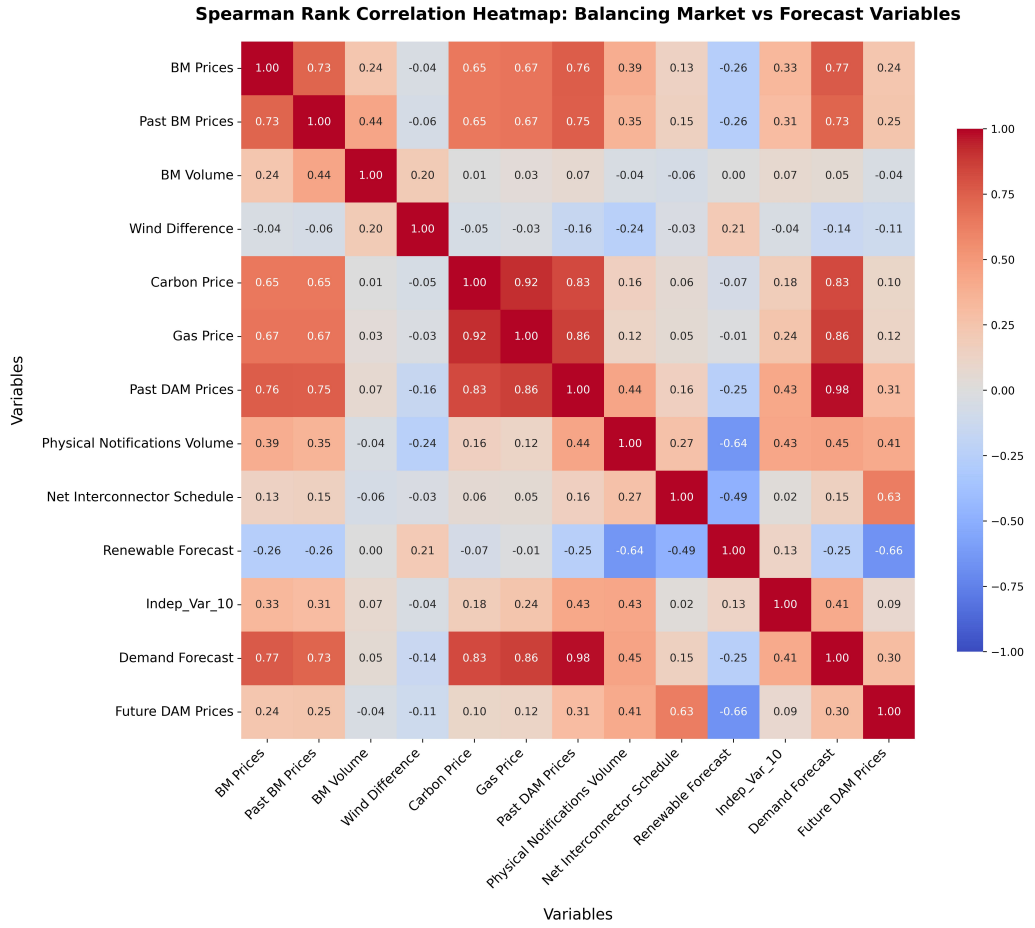


Figure 3.10: Spearman Rank Correlation Plots for the BM Prices and predictors

is even stronger, capturing non-linear effects between BM prices and demand. This suggests that BM prices and demand forecasts may follow a non-linear relationship that is better captured by the rank-based Spearman correlation. In contrast, *renewable energy forecasts* exhibit a negative correlation with *BM prices* ($\rho = -0.26$), suggesting that higher *renewable forecasts* are associated with lower *BM prices*. This is likely because cheaper renewable energy reduces the need for costly balancing interventions. Overall, BM prices are mainly driven by historical prices, fossil fuel markets, and demand forecasts, with renewable forecasts exerting downward pressure.

3.2.7 Decomposition

Decomposition is a crucial step in time series analysis, breaking down a time series y_t into trend T_t, seasonality S_t, and residual R_t components, expressed as $y_t = T_t + S_t + R_t$. We use STL decomposition with LOESS, as introduced by [RB90], which is particularly suited for handling complex seasonal patterns

and dynamic trends. Unlike classical methods, STL is non-parametric and effectively models non-linear trends and multiple seasonalities, making it ideal for dynamic electricity markets. This method was applied to key variables such as DAM prices, BM prices, and their predictors, providing a clearer understanding of each component's impact on market prices.

For each variable analysed, the smoothing parameters for the seasonal (S_t) and trend (T_t) components were optimized by adjusting seasonal windows n_s and trend windows n_t. This ensured that T_t accurately captured long-term movements while S_t reflected recurring patterns. The decomposed components were then analysed to assess their contributions, revealing insights into underlying structures such as long-term trends and recurring seasonal effects. This approach enhances understanding of market dynamics and improves forecasting and decision-making in electricity markets.

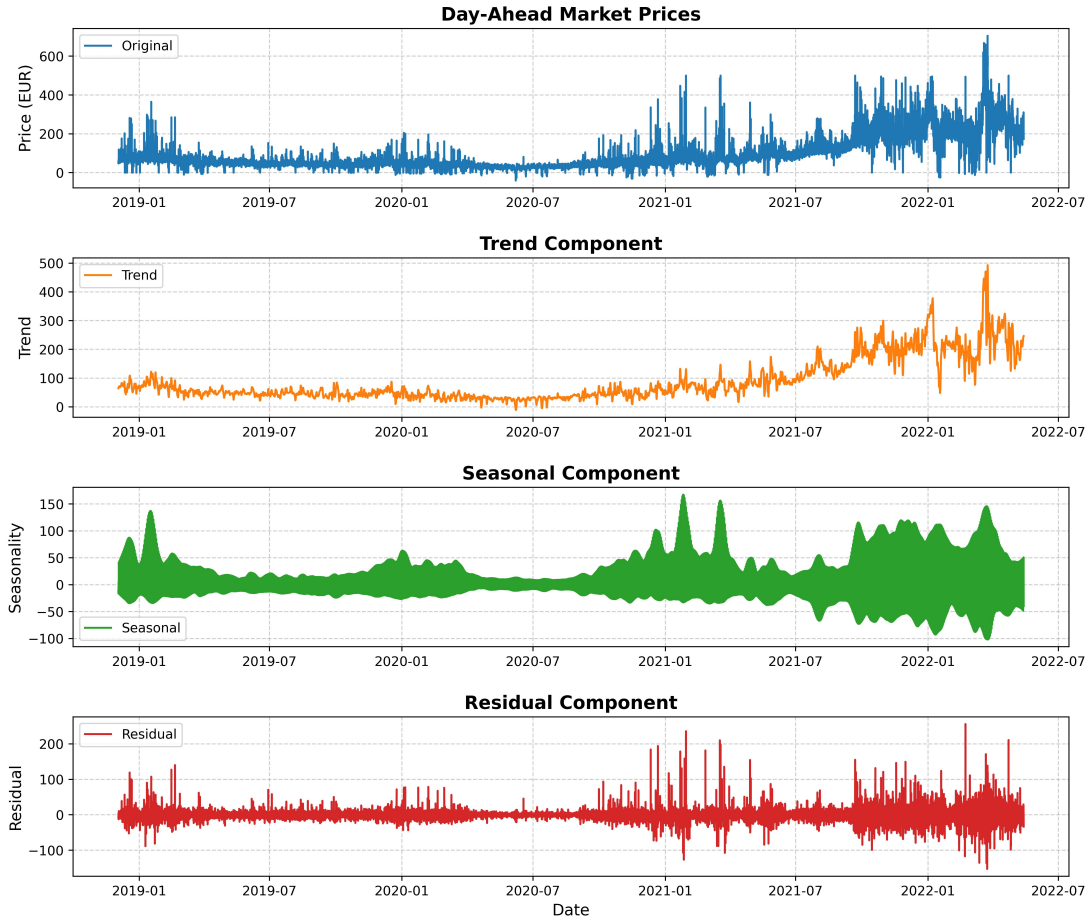


Figure 3.11: Seasonal and Trend Decomposition for DAM Prices

The STL decomposition of the DAM, as shown in Figure 3.11, reveals distinct patterns in price behaviour. The *trend* component shows a gradual upward tra-



Figure 3.12: Seasonal and Trend Decomposition for BM Prices

jectory, reflecting long-term growth in prices, likely driven by evolving market conditions and increased demand. The *seasonal* component captures regular, cyclical patterns, with notable fluctuations aligned with daily peaks in electricity consumption. Unlike the trend and residual components, the seasonal component appears much smoother, reflecting its focus on stable, recurring patterns that are less influenced by short-term volatility or abrupt changes in market dynamics. The *residual* component exhibits periods of heightened volatility, which could be attributed to external factors like sudden supply disruptions, regulatory changes, or unexpected shifts in demand. Together, these elements offer a clearer view of the drivers behind DAM price changes, helping improve forecasting models by identifying predictable seasonal behaviours and irregular shocks.

In contrast, the STL decomposition of the BM, as shown in Figure 3.12, indicates a relatively stable *trend*, with occasional fluctuations and aggressive price spikes towards the end. The *seasonal* component reveals clearer daily cycles,

although less pronounced than in the DAM, pointing to more stable, short-term price dynamics in the BM. The *residual* component presents occasional sharp spikes, signifying abrupt adjustments likely due to unexpected grid imbalances. However, the overall residual volatility is lower compared to the DAM, underscoring the BM's comparative stability. These decomposed components provide a comprehensive understanding of BM price behaviour, with significant implications for improving forecasting accuracy and market operations.

3.2.8 Conclusion

This chapter focuses on the description and analysis for the DAM and BM datasets used in this study. The datasets underwent thorough preprocessing, including handling missing data, outlier management, and data alignment, to ensure their suitability for time series modelling. Stationarity tests, such as the ADF and KPSS, confirmed that the DAM and BM series became stationary after differencing, validating their use in statistical and ML models. Autocorrelation and partial autocorrelation analyses revealed significant temporal dependencies, highlighting the daily and seasonal cycles that influence market dynamics.

Serial correlation tests, including the Ljung-Box and Breusch-Godfrey tests, indicated that the DAM model effectively captures time-dependent structures with no significant autocorrelation across various lags. In contrast, the BM model exhibited minor short-term autocorrelation that diminishes with longer lags. Heteroscedasticity tests, namely the Breusch-Pagan and White tests, identified significant heteroscedasticity in both DAM and BM models, suggesting that residual variances are not constant and are influenced by independent variables, reflecting the volatile nature of electricity markets.

Further analyses, including multicollinearity, correlation, decomposition, and residual diagnostics, assessed the robustness of the models. Multicollinearity tests showed minimal concerns for the DAM model, while the BM model exhibited moderate multicollinearity in some variables, reflecting redundancy yet preserving key market dynamics. Correlation analysis suggests that DAM prices are moderately associated with demand forecasts, while BM prices show correlations with past prices, fuel costs, and demand forecasts. STL decomposition provided insights into trends, seasonality, and residual volatility, aiding the understanding of price behaviours in both markets.

Overall, the preprocessing and diagnostic testing conducted in this chapter en-

sure that the DAM and BM datasets are robust and well-suited for the forecasting models developed in subsequent chapters. The insights gained from these analyses lay a solid foundation for improving predictive performance and optimising trading strategies, while also helping to understand and justify the experimental results. The code used for data processing and analysis is available at: GitHub Repository[3].

[3]https://github.com/ciaranoc123/Time-Series-Analysis

Chapter 4

Forecasting the Balancing Market Price

Accurately forecasting electricity prices in short-term markets is challenging due to high volatility, sharp price spikes, and complex seasonal patterns. The increased integration of renewable energy sources, along with the widespread deployment of Energy Storage Systems (ESS), has further complicated market dynamics, underscoring the need for advanced predictive methodologies [MABH16, Eur24]. In Chapter 2, we conducted a detailed review of electricity market structures and the statistical, Machine Learning (ML), and Deep Learning (DL) models used for Electricity Price Forecasting (EPF), along with their applications in related forecasting tasks such as load and wind prediction. Building on this foundation, this chapter focuses specifically on applying forecasting models to the highly volatile Balancing Market (BM), where unpredictable price fluctuations introduce unique challenges.

This chapter focuses on establishing a clear, standardised, BM forecasting framework. We adapt and benchmark various forecasting models that have proven effective in the Day-Ahead Market (DAM), assessing their suitability and performance in the BM. Building on the foundations laid in earlier chapters, the framework outlines steps for training, forecasting, and evaluating statistical, ML, and DL models tailored to the BM's unique dynamics. Our methodology leverages both DAM and BM datasets introduced in Chapter 3 incorporating both historical and forward looking predictors. We evaluate the performance of these models in forecasting the BM and examine how variations in training set sizes impact accuracy, using standard metrics such as Mean Absolute Er-

ror (MAE), Root Mean Squared Error (RMSE), and symmetric Mean Absolute Percentage Error (sMAPE).

The remainder of this chapter is organised as follows: Section 4.1 discusses the significance of the BM and the challenges it presents, especially in the context of increased renewable energy integration. Section 4.1 outlines the key role the BM plays within electricity markets. Section 4.2 presents the specifics of the forecasting problem and the associated methodology. Section 4.2.2 describes the models utilised in this study, including their configurations and adaptations for BM forecasting. The experimental results and their analysis are presented in Section 4.3, highlighting key findings and model comparisons. Finally, Section 4.4 concludes the chapter with observations on the unique challenges of BM forecasting and potential directions for future research.

4.1 The Importance of the Balancing Market

The BM, often referred to as the "market of last resort," is a final mechanism for trading electricity, designed specifically to address real-time supply-demand imbalances. It is inherently more volatile than both the DAM and Intra-Day Market (IDM), as participants rely on it to resolve unforeseen deviations from their contracted positions [OT19]. This unpredictability makes accurate EPF exceptionally challenging yet crucial for energy generators, buyers, and financial traders, who must navigate sudden price spikes and drops effectively [MP17]. For participants utilising ESS, which are highly sensitive to real-time price shifts, precise BM forecasts are essential to optimising operations and capturing arbitrage opportunities. However, as discussed in Chapters 2 and 3, the BM's evolving structure and limited dataset availability pose considerable challenges in testing, identifying, and selecting suitable forecasting models for this market.

Despite receiving less research focus, the BM plays a fundamental role in grid stability, particularly as renewable energy integration grows. Unlike forward markets like the DAM, which operate on planned schedules, the BM requires nearly instantaneous responses to generation and demand fluctuations, with price signals reflecting the immediate costs of these actions. This real-time balancing role has become increasingly crucial as renewables amplify the need for precise adjustments, demanding highly responsive and accurate forecasts to prevent imbalances that could destabilise the grid. However, forecasting the BM is complicated by regional variations in market rules, reliance on high-

frequency data, and the inherently unpredictable nature of balancing needs. Unforeseen events, such as equipment failures or abrupt demand changes, often lead to rapid price spikes or dips, challenging both traditional and advanced forecasting models to capture these dynamics accurately. Even minor inaccuracies in BM forecasts can result in financial losses or increased grid instability, given the speed and scale at which fluctuations occur in this market.

In contrast, the DAM has received substantial research focus, resulting in established methodologies that encompass a wide range of forecasting techniques, from traditional statistical models to advanced ML and DL approaches [Wer14, NWA17, LW19, LMDSW21]. DL models, particularly Recurrent Neural Network (RNN)s like Long Short-Term Memory (LSTM), have shown notable success in capturing temporal dependencies and nonlinear patterns in the DAM [UOT18, LDRDS18]. However, while these approaches have proven effective in the relatively stable DAM environment, they may lack the granularity and reactivity needed for the more volatile BM. This gap underscores a growing need for research dedicated to developing specialised BM forecasting methodologies that can account for its unique dynamics and deliver more reliable forecasts.

4.2 Methodology

To address the unique challenges of BM price forecasting, this framework is designed to capture the real-time dynamics of the market. In this section, we outline the analytical framework for forecasting BM prices, covering model configuration, training, and evaluation.

Section 4.2.1 describes the dataset splitting and hyperparameter optimisation procedures. Section 4.2.2 introduces the models used, detailing adaptations in their configurations to forecasting the BM. In Section 4.2.3, we present the key performance metrics and statistical tests employed to assess forecast accuracy. These metrics are carefully selected to account for the specific dynamics of electricity markets, where factors such as negative prices—frequent in both the DAM and particularly in the BM—render certain standard metrics, such as the Mean Absolute Percentage Error (MAPE), less applicable.

4.2.1 Experimental Configuration

In our analysis, the approach involves three main steps:

1. Hyperparameter tuning
2. Model training
3. Forecasting on unseen data

For each encountered model, we fit it using training datasets spanning 30, 60, 90, and 365 days. In the context of forecasting BM prices, several walk-forward validation techniques, including rolling [VS06], sliding window [AAK+22], expanding [Bro19], and walk-forward nested cross-validation [LB21], offer unique advantages. We chose rolling walk-forward validation for its focus on recent data, using a fixed-size training window that discards older observations as new data arrives [HA18, Mör19, BS20, MA21, YY24]. In our configuration, the model is retrained every 8 hours to incorporate the most recent data available.

Hyperparameters for the ML and DL models are determined based on validation set performance. We used 25% of the training set as a validation set, with a maximum validation window of 30 days for the 365-day dataset. To accommodate diverse hyperparameter configurations for validation and training sets, we repeat step 1 (hyperparameter tuning) every quarter for both ML and DL models, as illustrated in Figure 4.1. The process is as follows:

1. Given a model of interest and a starting time point t, obtain a validation set and determine hyperparameter settings.
2. Using hyperparameter settings from step 1, fit the model with training data.
3. Utilise the fitted model from step 2 to forecast the test set.
4. Return to step 1, set $t = t + 90$ days, and repeat the process.

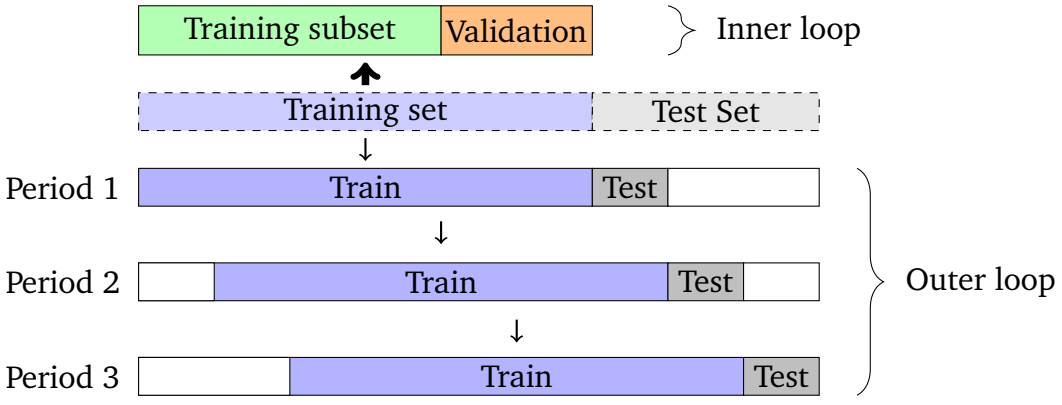


Figure 4.1: Rolling Walk-Forward Validation process for the first three periods in our experimental configuration.

When performing the hyperparameter optimisation step, we use various libraries/toolboxes. For ML models such as RF and SVR, we utilised the *Scikit-learn* library [PVG+11], and for XGB type models, we used the *XGBoost* library [CG16]. For the DL models (Single-Headed Deep Neural Network (SH-DNN) and Multi-Headed Deep Neural Network (MH-DNN)), we employed *Tensorflow* [Cho21] in conjunction with Talos [YS20] for the hyperparameters optimisation. For statistical model LEAR, we used the *epftoolbox* [LMDSW21]. A broad range of hyperparameter values was tested for the ML and DNN models to find optimal settings for each time period and training size.

4.2.2 Balancing Market Prediction Models

This section provides an overview of the models utilised for BM forecasting. Given the limited exploration of BM forecasting, we benchmarked a wide range of models, including statistical, ML, and DL approaches. The models were selected based on their performance in recent EPF literature across both DAM and BM domains. Building upon studies such as [LDRDS18, LPKC20, LB21] for DAM EPF and [Nar22, LPKC20, KEF15] for BM EPF, we identified consistent performers like LASSO/LEAR, XGB, and RF across both markets. Moreover, DNNs showed promising results specifically in the DAM setting [UOT18, LDRDS18]. Building on the model descriptions and their performance in EPF, as detailed in Chapter 2.2, we briefly summarise the key models considered:

- **Naive (Section 2.2.1)**: Utilises the preceding 8 hours of BM prices as a forecast for the subsequent 8 hours, capturing immediate pricing dynamics crucial in the real-time BM environment. This approach outperformed alternatives like considering the same hour of the previous day or the same weekday, which may overlook such dynamics. It shares similarities with methods in [BKM18b], which employed the DAM values from the same days a year prior as a seasonal naive DAM forecast. This represents our baseline model.
- **Auto-Regressive Integrated Moving Average (ARIMA) (Section 2.2.1)**: is a time series forecasting method that models the relationship between a series of observations and the time lagged versions of itself. It incorporates three key components: AR, Differencing (I), and MA. The ARIMA model is specified by three parameters: p, d, and q, representing the number of autoregressive terms, the degree of differencing, and the number of moving average terms, respectively. We employed ARIMA to capture

linear trends present in the data.

- **Lasso Estimated AutoRegressive (LEAR) (Section 2.2.1)**: A variant of the AR model that uses LASSO regularisation for feature selection, ensuring flexibility in high-dimensional settings. We adapted the implementation designed in [LMDSW21] to the BM structure of the input/output data. Similar to their work, we performed daily hyperparameter tuning for LEAR, estimating λ with the LARS method and in-sample AIC. The optimal λ from LARS was then used to re-calibrate LEAR through traditional coordinate descent.

- **K-Nearest Neighbors (KNN) (Section 2.2.2)**: KNN stands out as a non-parametric instance-based learning method renowned for its simplicity and effectiveness, particularly in scenarios where capturing intricate patterns within data is paramount. By leveraging the proximity of similar instances in the training set, KNN predicts an output by averaging the values of its "K" nearest neighbors. Its versatility in handling various distance metrics, such as Euclidean or Manhattan distance, enabling it to discern complex relationships within the data.

- **Support Vector Regression (SVR) (Section 2.2.2)**: It performs a non-linear mapping of the data to a higher-dimensional space where linear functions are used to perform regression. SVR performs well in high-dimensionality space [DBK+96]. In our case, the RBF kernel was chosen every time.

- **Random Forest (RF) (Section 2.2.2)**: It is an ensemble model that combines several regression trees to generate predictions. It is based on the principle of bagging - short for bootstrap aggregating, splitting the data into several subsets and choosing subsets randomly with replacement, thus combining individual models with low bias and high variance, which, when aggregated, reduce the variance and retain low bias.

- **Extreme Gradient Boosting (XGB) (Section 2.2.2)**: Similarly to the RF, XGB combines several regression trees to generate forecasts, but it is based on the principle of boosting - where learners are learned sequentially, combining models with high bias and low variance to reduce the bias while keeping a low variance. This makes it suitable for capturing complex relationships in BM price data.

- **Deep Neural Network (DNN) (Section 2.2.3)**: A simple extension of the

traditional Multi-Layer Perceptron, the first DL model for predicting DAM prices is a DNN used in [LDRDS18]. The chosen format for the single-headed DNN model is sequential, allowing for further hyperparameter optimisation with a range for the number of layers to be added for each block (where a block is comprised of between 1 to 3 connected layers) and activation function, see Figure 4.2b. The first layer flattens the data, followed by several dense layers and a single dropout layer after block 3. Their output is fed to a series of DNN blocks. The SH-DNN model is chosen for its ability to effectively capture intricate non-linear patterns present in the data.

- **Multi-Headed Recurrent Neural Network Deep Neural Network (MH-RNN-DNN) (Sections 2.2.3 & 2.2.3)**: A non-sequential network that incorporates historical and future-looking data using dense and recurrent layers, emphasising its ability to capture both temporal dependencies and high-dimensional relationships in the BM data. This hybrid approach aims to incorporate recurrent layers (LSTM) for capturing sequential patterns in time series data, along with fully-connected layers (DNN) focused on the future-looking data. These branches have their output concatenated into one vector, and this vector is fed into a regular output dense layer containing 16 neurons for the prediction of 8 hours (or 16-time stamps). For clarity on the outline of the network, see Figure 4.2a. The obtained optimal hyperparameters for the Multi-Headed RNN DNN are 3 layers for both the LSTM and DNN networks, making for an 8-layer deep network with the concatenated and final layer. A more detailed structure of the network configuration can be found in the shared repository.

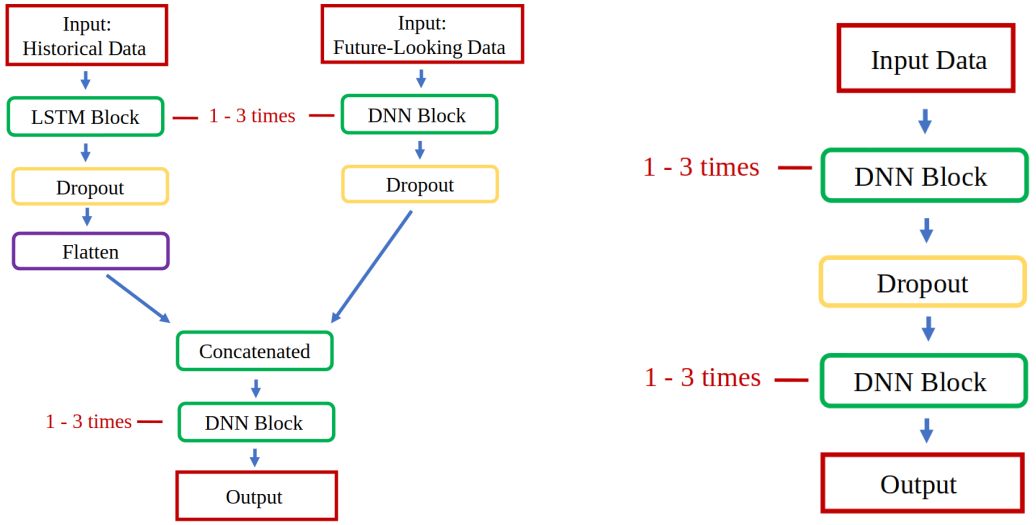
Details on the model configurations and hyperparameters are detailed in Chapter 2 and the code repository[1] for the evaluation of EPF models in the BM.

4.2.3 Point Forecasting Metrics

Commonly encountered metrics in the EPF literature used to evaluate forecast accuracy include MAE, RMSE [UTG18], as well as the sMAPE [Mak93]. Each metric provides a different perspective on model performance, making it important to consider them together for a comprehensive evaluation.

The MAE measures the average magnitude of the errors in a set of forecasts,

[1] https://github.com/ciaranoc123/Balance-Market-Forecast



(a) MH-RNN-DNN Model (b) SH-DNN Model

Figure 4.2: Comparison of the MH-RNN-DNN Model and SH-DNN Model.

without considering their direction. It is defined as follows:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t|, \tag{4.1}$$

where y_t and $\hat{y}_t$ are the actual and forecasted values, respectively, and n is the total number of observations. MAE is a straightforward metric that provides the mean error of the forecast, making it easy to interpret.

The RMSE is another common metric used to evaluate forecast accuracy. It penalises larger errors more than MAE, making it useful for identifying models that produce large deviations from actual values. RMSE is given by:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}. \tag{4.2}$$

While RMSE is sensitive to outliers, it highlights forecast error variance, offering a clearer picture of the model's performance under varying market conditions.

The sMAPE is preferred over MAPE due to its effectiveness in handling the frequent occurrence of prices near zero in the BM. It is defined as:

$$sMAPE = \frac{100}{n}\sum_{t=1}^{n}\frac{|y_t - \hat{y}_t|}{(|y_t| + |\hat{y}_t|)/2}. \tag{4.3}$$

sMAPE provides a percentage-based measure of forecast accuracy, normalising

the absolute errors by the sum of actual and forecasted values, which helps in comparing errors across different scales of data.

To determine if the differences in forecasting accuracy are statistically significant, we apply the Diebold-Mariano (DM) test [DM02]. The DM test evaluates the null hypothesis (H_0) that two models have equal predictive accuracy.

If H_0 is rejected, it indicates that one model significantly outperforms the other in terms of forecast accuracy. Given forecasted BM prices y_t and the associated errors $[\epsilon_1^{M_1}, \ldots, \epsilon_N^{M_1}]$ and $[\epsilon_1^{M_2}, \ldots, \epsilon_N^{M_2}]$, we compute the loss differentials and use the DM test to assess the following hypotheses:

$$\text{DM Test} = \begin{cases} H_0 : \mathbb{E}(d_k^{M_1,M_2}) \geq 0, \\ H_1 : \mathbb{E}(d_k^{M_1,M_2}) < 0. \end{cases} \tag{4.4}$$

If H_0 is rejected, it supports H_1, indicating that model M_2 has significantly better predictive accuracy than model M_1.

4.3 Experimental Results

In this section, we provide an in-depth analysis of the models' performance, focusing on accuracy, statistical significance, and computational efficiency. We first assess the accuracy of the predictions using key metrics such as MAE, RMSE, and sMAPE, while comparing the impact of different training horizons and model performance in both the BM and DAM. We then analyse the statistical significance of the differences between models, followed by an exploration of how price predictions vary throughout the day to identify which periods of the BM are more difficult to forecast. Lastly, we evaluate the computational efficiency of each model, considering their suitability for real-time forecasting.

4.3.1 Forecasting Accuracy

Table 4.1 provides a detailed comparison of the models' prediction accuracy across different training sizes, measured using MAE, RMSE, and sMAPE. This comparison allows us to assess the performance of each model in forecasting the BM under various data conditions. Additionally, we compare the BM results for the SH-DNN and LEAR models with their corresponding performance in the DAM, revealing notable differences in accuracy between the two markets.

LEAR emerges as the top-performing model for the BM, with few exceptions

	MAE					RMSE					sMAPE				
Training Size	30 days	60 days	90 days	Max Data	DAM	30 days	60 days	90 days	Max Data	DAM	30 days	60 days	90 days	Max Data	DAM
Model															
Naive	51.96	51.96	51.96	51.96		84.27	84.27	84.27	84.27						
LEAR	35.95	**34.12**	**33.57**	**32.82**	10.15	**57.89**	**56.38**	**56.34**	56.53	18.84	72.14	**67.19**	**66.20**	**64.93**	22.68
ARIMA	43.15	42.72	42.44	41.51		68.54	67.61	67.20	66.00		76.26	75.82	75.50	74.24	
XGB	**35.35**	34.40	34.29	33.71		60.46	58.03	56.89	57.03		69.71	67.62	67.12	66.89	
RF	36.15	35.24	34.92	34.66		58.62	57.27	56.66	**56.22**		**68.67**	68.55	68.45	67.18	
SVR	41.43	38.01	37.70	37.39		69.81	66.12	65.58	64.92		78.35	69.85	68.97	68.09	
KNN	45.85	44.94	44.71	44.55		69.79	68.88	69.20	68.59		79.83	77.97	76.98	74.57	
SH-DNN	40.14	38.67	38.31	37.41	**7.34**	64.13	68.91	62.25	61.95	**13.86**	77.85	72.10	72.05	70.81	**19.01**
MH-RNN-DNN	41.42	40.89	40.35	38.28		64.83	64.02	64.42	61.96		79.77	77.20	76.68	74.84	

Table 4.1: Balancing Market prediction accuracy for the different training sizes. In bold, the best-performing model. MAE is measured in €/MWh.

during shorter training periods. Its subtle adjustments lead to smoother predictions with lower variation. Figure 4.3 displays a comparison between the BM prices and LEAR's predictions. The model does not frequently attempt to predict significant spikes, which seems to contribute to its improved accuracy. Rather than capturing extreme price movements, LEAR consistently returns more stable forecasts, minimising large errors. Notably, increasing the training set size significantly improved LEAR's performance, with MAE decreasing by 3.31 as the training period expanded from 30 days to 310 days. ML models, includ-

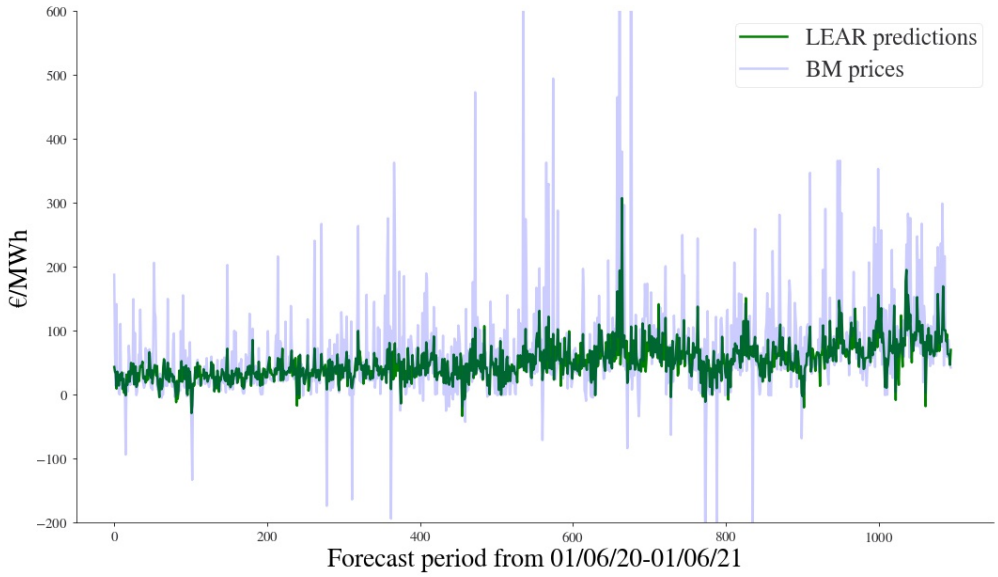


Figure 4.3: LEAR Forecast for BM

ing KNN, RF, XGB, and SVR displayed varied performance. XGB stood out by closely approaching the accuracy of LEAR and outperforming KNN, SVR, and RF. Both XGB and RF outperformed the DL models, while KNN and SVR struggled to capture meaningful data patterns, with KNN even failing to outperform ARIMA. These findings align with UK BM paper [LPKC20], confirming XGB's superiority over RF in a similar setting.

DL models exhibit inferior performance, consistently under-performing across all metrics, a result that mirrors previous studies [DBdV+19, Nar22]. While these models perform well in more stable environments like the DAM, they struggle with the BM's complexities, including overfitting to recent price spikes and difficulties in capturing complex BM dynamics. This divergence underscores the complexities inherent in electricity market forecasting. Figure 4.4

compares the performance of the DL models on part of the test set. In contrast to LEAR, see Figure 4.3, we can see that the DL model behaves overconfidently, predicting considerable spikes. If these predictions are incorrect, they are strongly penalised by the metrics compared to more cautious ones.

Between the DL models, the SH-DNN model outperformed the MH-RNN-DNN model, showing a 2.47 reduction in MAE as the training set size increased. While the SH-DNN model displayed better forecasting ability than XGB and LEAR during the least volatile period (first 3 months), its performance deteriorated as market volatility increased. The limited amount of training data may also explain the poor performance of DL models, which typically benefit from larger datasets due to their complex structural configurations and numerous parameters. On the other hand, the MH-RNN-DNN model, which performed exceptionally well in the DAM [LDRDS18], struggled in the BM context. This could be attributed to the current dataset's unsuitability for the split approach used, where data is fed into two separate branches. If past and future data heavily influence each other, both structures may fail to establish these relationships. In contrast, the SH-DNN model avoids assumptions about input data, enabling more comprehensive relationship-building.

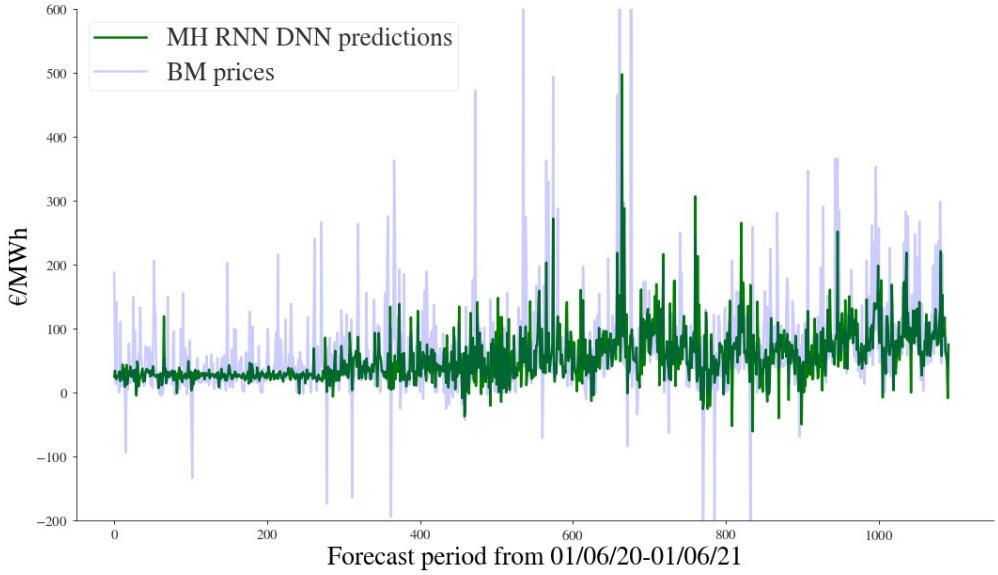


Figure 4.4: MH RNN DNN Forecasts for BM

We explored the impact of training data size on model performance, finding that all models benefited significantly from larger training sets. Expanding the

training window to 365 days led to an average reduction of 2.39 in MAE, 2.61 in RMSE, and 5.12 in sMAPE across the models. However, the DL models showed a smaller improvement compared to the ML models, despite the larger data availability. This suggests that while DL models may require even more data to fully optimise, access to a larger dataset consistently improves model accuracy overall, highlighting the critical role of data availability in BM price forecasting.

Statistical Test

Differences in accuracy across models can stem from the inherent stochasticity of the dataset. To compare model performance, the DM test is widely used in EPF literature [UOT18, LDRDS18, LMDSW21]. The DM test evaluates the null hypothesis (H_0), rejecting it if one model shows a statistically significant improvement in accuracy over another.

Table 4.2: DB test statistical significance. The rows represent model M_1 and the columns represent model M_2

	DB Test - MAE										
Model	LEAR	XGB	RF	SVR	SH DNN	MH RNN DNN	ARIMA	KNN	Naive
LEAR		✗	✗	✓	✓	✓	✓	✓	✓
XGB			✗	✓	✓	✓	✓	✓	✓
RF				✓	✓	✓	✓	✓	✓
SVR					✗	✗	✓	✓	✓
SH DNN						✗	✓	✓	✓
MH RNN DNN							✓	✓	✓
ARIMA								✓	✓
KNN									✓
Naive									

Table 4.2 presents the DM test results based on MAE values. The rows represent model M_1 and the columns represent model M_2, with the ✓ symbol indicating that M_2 significantly outperforms M_1 at a 95% confidence level. A ✗ symbol implies that the differences between the models are not statistically significant.

The results indicate that LEAR, XGB, and RF exhibit similar performance, with no statistically significant differences between them. These models consistently outperform the others, demonstrating their ability to navigate the volatility of the BM. In contrast, all models show statistically significant differences when compared to ARIMA, KNN, and Naive, which are among the weakest performers. ARIMA shows statistically significant differences from both KNN and Naive, highlighting their inferior performance. Meanwhile, SVR and the DL models (SH-DNN and MH-RNN-DNN) show no significant differences from one an-

other, suggesting their performance is relatively close but still inferior to the top-performing models (LEAR, XGB, and RF). However, all models significantly outperform the Naive baseline, confirming that even the lower-performing models provide meaningful predictive improvements over a simple heuristic approach.

Hourly Analysis

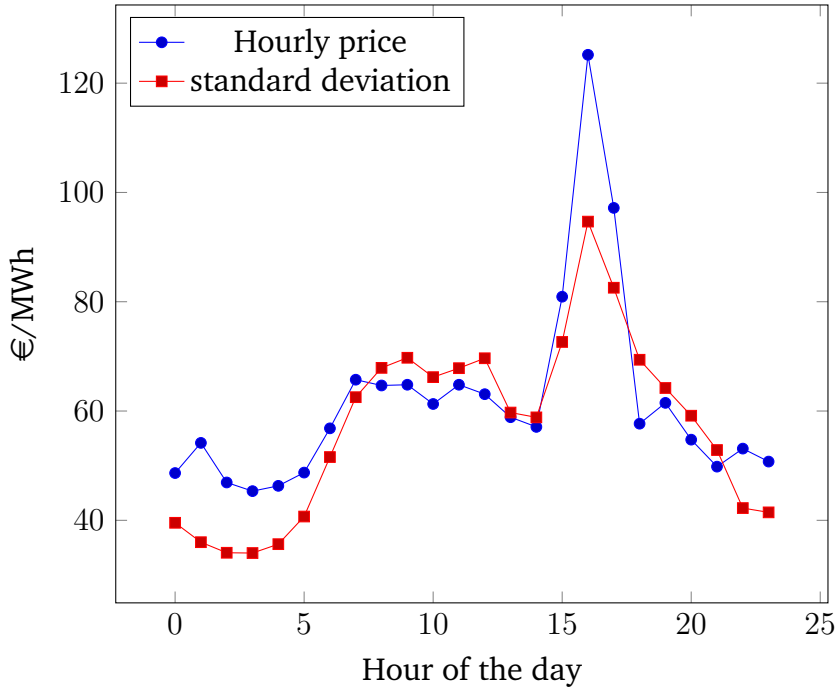
Analysing 24-hour trends provides clearer insights into performance differences, based on models trained with 365 days of data. Figure 4.5a shows the average price and its standard deviation, while Figure 4.5b shows the models' MAE both aggregated by the hour. There are considerable variations in prices throughout the day, with a strong peak occurring between 15:00 and 18:00, which is also the period of highest volatility. In contrast, during the night and early morning, the prices and their standard deviation are considerably lower. The forecast error of the predictors follows a similar pattern, with higher prices and volatility leading to less accurate predictions in absolute terms.

These trends strongly impact individual model performances. The DL models are more accurate on the night prices, where prices are more stable and uncertainty is lower. However, their attempts to predict the daytime peaks result in higher errors during more volatile periods. On the other hand, LEAR and XGB exhibit more stable error rates and are the most accurate during the volatile periods. For specific applications, such as determining the optimal time to charge a battery, techniques that excel in predicting lower and stable prices may be more desirable than those that perform best on average.

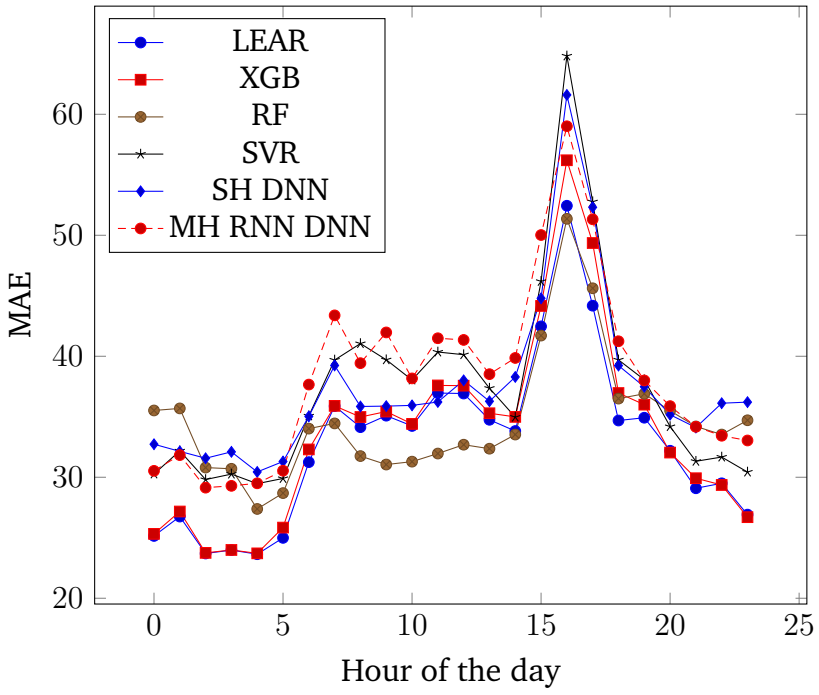
4.3.2 Comparison with DAM

To highlight the differences between the two markets, we applied the forecasting models presented in Section 4.2.2 to the DAM. Figure 4.6 illustrates the differences in price ranges and uncertainty between the two markets. The BM, in particular, experiences substantial volatility, with consecutive trading periods often witnessing extreme price fluctuations in both directions, for example:

- For the 4 BM settlement periods on 27/06/2020 in the 18:30-20:30 time horizon, prices veered from €2.01/MWh to €-390.07/MWh to €-38.69/MWh before ending at €3.12/MWh.
- On the 07/01/2021 15:00-16:30 prices jumped from €359.89/MWh to €1222.89/MWh, then to €1059.07/MWh and back to €320.88/MWh.



(a) BM Price mean and SD for each hour.



(b) Models' MAE for each hour.

Figure 4.5: Hourly breakdown of the BM prices and the models' forecast error.

For context, within the related 12-hour time block, the BM price ranged from a low of -€98.42/MWh to the aforementioned high of €1222.89/MWh.

While the DL models encounter challenges in BM predictions, despite the same

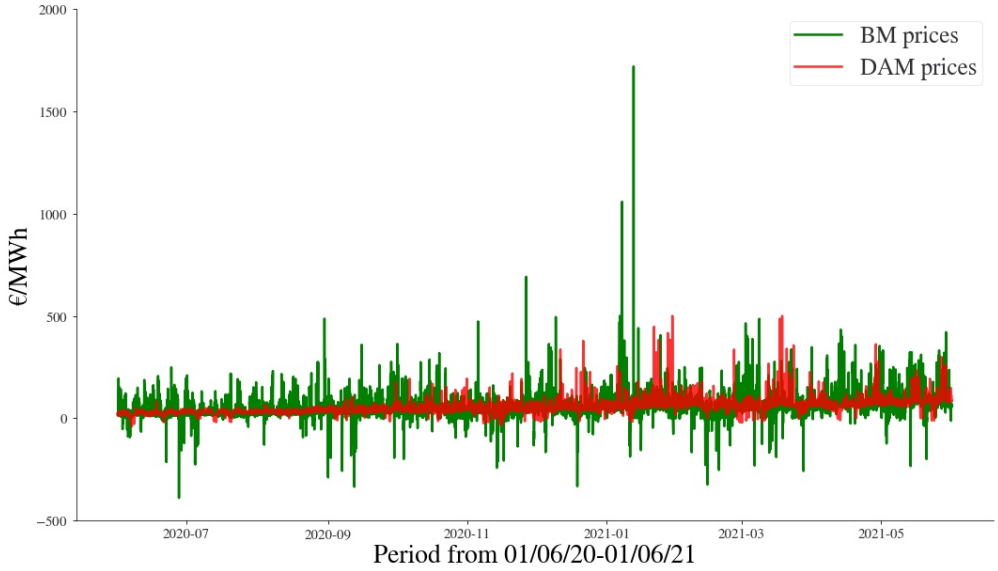


Figure 4.6: DAM and BM prices.

configuration as the DAM, they exhibit more accurate forecasts in the DAM domain, reflecting their ability to capture price spikes and market dynamics. DL models' superior performance in the DAM, as compared to the BM, underscores the impact of market structure and data characteristics on model performance. Specifically, the SH-DNN outperforms LEAR by a significant €2.81/MWh, confirming the results present in the literature [JLMW22]. In Figure 4.7, we can see the predictions of the SH-DNN in comparison to the DAM prices. Compared to Figure 4.4, we can see that the predicted peaks are more accurate. As previously mentioned, DNNs tend to overfit to recent spikes in the BM, leading to incorrect predictions of further price increases over the next 8 hours, whereas the more stable DAM market is better suited for DL approaches.

Additionally, the performance of these models may have been affected by relatively small training sample sizes, with models trained on less than a year of data generally showing poorer results, consistent with findings from [LDRDS18]. This comparison highlights the need for further research to develop more effective forecasting models tailored to the BM.

Residual Analysis

Residual analysis is a key aspect of time series modelling, aimed at evaluating model performance and identifying underlying issues. The residuals should

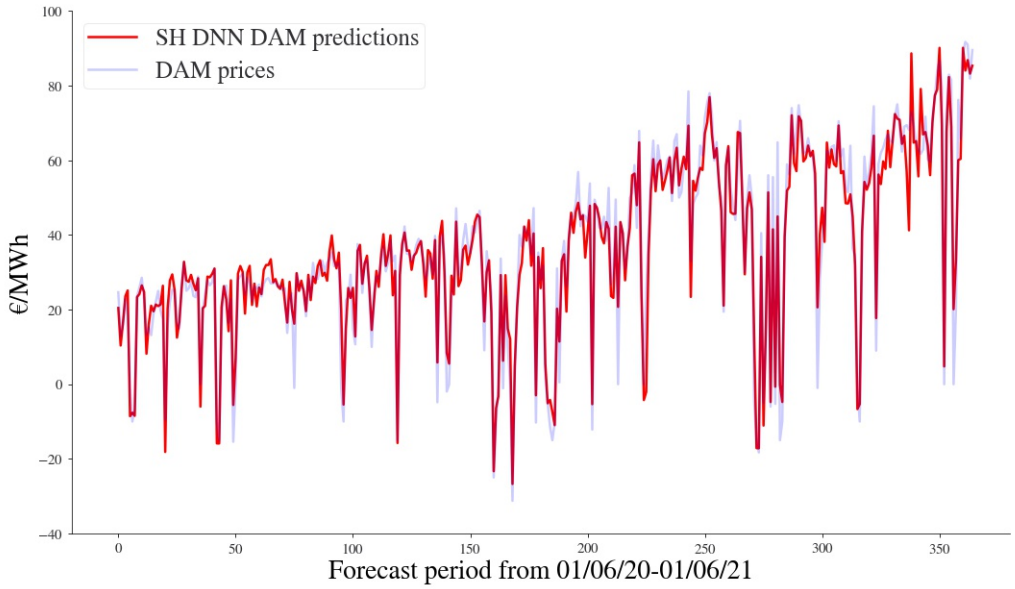


Figure 4.7: SH DNN Forecast for DAM

ideally be randomly dispersed around zero, indicating the model has accurately captured the data structure. In the context of EPF for the BM, residual plots are employed to visually assess the residuals' distribution, to assess non-linearity and heteroscedasticity. Patterns like a funnel shape may signal heteroscedasticity, while systematic curves can suggest non-linearity or model misspecification.

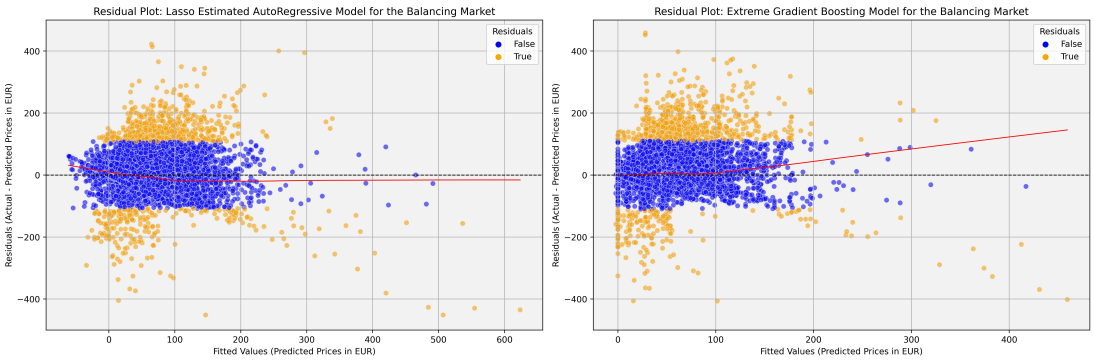


Figure 4.8: Residual Plots for Balancing Market Forecasts

In the residual plots for the LEAR and XGB models (Figure 4.8), distinct patterns emerge. The LEAR model shows a slight curved pattern in the residuals, indicating non-linearity, with systematic deviations from zero at both lower and higher fitted values. However, the spread of residuals remains fairly constant, suggesting that heteroscedasticity is not a significant concern. In contrast, the XGB model's residuals are more evenly distributed around zero, with less curvature,

indicating better handling of non-linearity. Although there is a slight upward trend in residuals at higher fitted values in the XGB model, it still effectively captures the overall data structure. Both models show some outliers, indicating periods of reduced forecasting accuracy, particularly during extreme price movements, but demonstrate reasonable capacity to model BM prices overall.

Computational Efficiency

In practical terms, in a live BM forecasting setting, participants should consider the time lag between available information (explanatory variables) and the computational time required for model output. This can vary based on factors such as the deployed model and available hardware. In our analysis, DL models were trained on a GPU (Nvidia GTX 1080 ti), while statistical and ML models were trained on the CPU (Intel(R) Core(TM) i7-6800K CPU @ 3.40GHz).

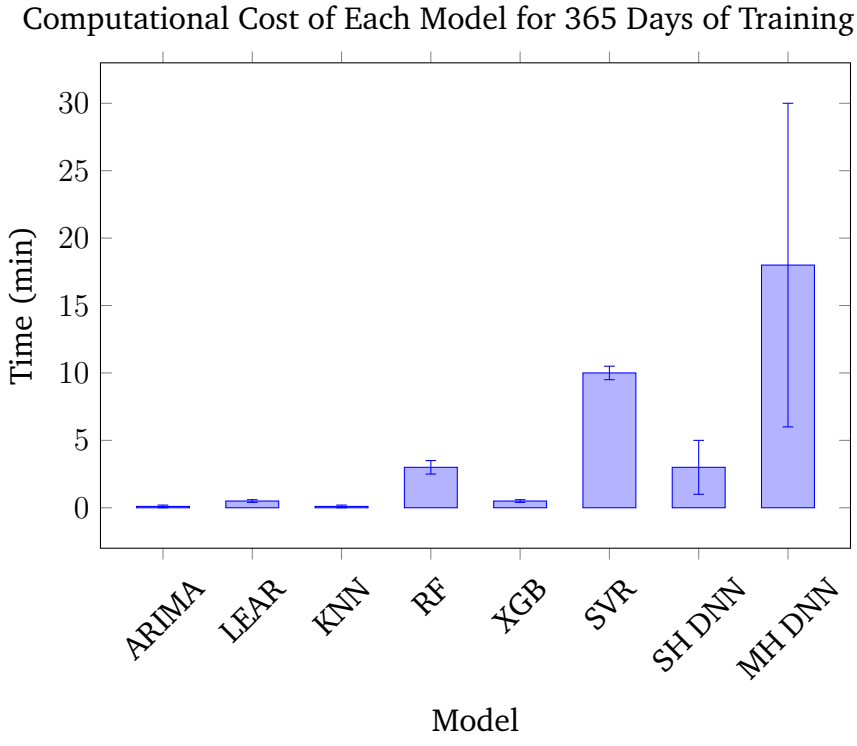


Figure 4.9: Computational Time Required for Training

Figure 4.9 illustrates the training time for all models, which is of considerable importance given the models need to be retrained quite often to incorporate the most recent training data. As expected, DL models take significantly longer to train and show the highest variability in computational costs. Factors like the number of parameters, dataset size, and the number of epochs contribute

to their performance. The MH model's complexity and parameter count further extend its training time. In contrast, LEAR and XGB strike an excellent balance between computational efficiency and forecast accuracy, making them ideal for real-time applications, while ARIMA and KNN prioritise speed over accuracy.

4.3.3 Limitations

Our models focus on providing forecasts that are accurate on average, which partially compromises their ability to identify extreme cases, such as exceptionally high or negative prices. This limitation arises because erroneous predictions of such extremes are heavily penalised by the training metrics. In applications where detecting these anomalies is critical, even at the cost of reduced accuracy, alternative approaches specifically designed for extreme event prediction in time series could be applied, such as those suggested by [DZP+19]. Additionally, adopting a different forecasting approach, such as using more granular 5-minute datasets or incorporating publicly available commercial order data into the model structure, could further improve overall accuracy.

Furthermore, the findings in this work are based on Irish market data from 2019 to 2022, and their generalisability may decrease as the Irish electricity generation and infrastructure evolve. For instance, the widespread adoption of ESS, which stores energy when prices are low and sells it when high to reduce volatility, could decrease price fluctuations and extreme events, potentially altering model performance rankings. However, such changes typically unfold gradually, and the forecasting framework presented here can be easily adapted to new datasets. Additionally, incorporating probabilistic forecasting and the inclusion of broader datasets could improve model performance.

4.4 Conclusion

In this chapter, we introduced a framework to benchmark predictive models for forecasting electricity prices in the BM. Through extensive numerical analysis, we evaluated statistical, ML, and DL approaches that have proven effective in similar tasks. While DL models excelled in the more stable DAM, they struggled to capture the uncertainty and volatility of the BM, where simpler approaches performed better. LEAR and tree ensemble techniques achieved the highest accuracy due to their conservative predictions, while DL models faced challenges with sudden price spikes, resulting in increased errors. Simpler models

like ARIMA and KNN lacked the representational power to model the dataset's complex patterns effectively.

The strong performance of both LEAR and XGB can be attributed to the built in l_1 and l_2 regularisation, which effectively mitigate multicollinearity and prevent overfitting in this high-dimensional BM dataset. This analysis also provided valuable insights into the dynamics of the Irish BM market. The original dataset, models, and framework are available under an open-source license.

This study suggests several potential future directions for BM price forecasting. The hourly analysis indicates that an algorithm portfolio, using different approaches for different parts of the day, could exploit the varying strengths of the models. Alternatively, combining predictions from different models in an ensemble could enhance robustness [BO24]. Exploring additional BM markets would also help generalise the findings and highlight key differences between markets. Some BM operators provide more granular 5-minute data, which could facilitate an alternative modelling approach. Additionally, greater data availability would support the use of transfer learning techniques.

Future research directions for the BM could explore generating probabilistic forecasts through modern methods like conformal prediction (discussed in Chapter 6) or approaching price prediction as a classification task using DNNs for spike detection. These forecasts could be leveraged to improve trading strategies (discussed in Chapter 5), in conjunction with the DAM, to improve financial outcomes for ESS participants.

Chapter 5

Dual Market Trading in the Day-Ahead and Balancing Markets with Quantile Forecasts

Building on the forecasting challenges and insights from the Balancing Market (BM) discussed in Chapter 4, this chapter explores how those forecasts can be leveraged in trading strategies, particularly through Battery Energy Storage Systems (BESS) for electricity arbitrage. Accurate Probabilistic Electricity Price Forecasting (PEPF) is fundamental to optimising trading decisions, increasing profitability, and mitigating risk. We focus on trading strategies that exploit price differentials across the Day-Ahead Market (DAM) and BM, integrating BM forecasts to optimise trading in a dual market approach. Our methodology uses the same DAM and BM datasets from Chapter 3, and forecasting approach outlined in 4, incorporating both price forecasts and practical BESS constraints to evaluate trading performance under real-world conditions.

This chapter emphasises quantile-based forecasts and BESS constraints to evaluate forecast-driven trading strategies. Our trading framework bridges theoretical models and real-world applications, utilising an 8-hour BM prediction horizon to align DAM forecasts with BM positions, improving decision-making and profitability. We evaluate the profitability of different trading constraints and assess the economic viability of small and large battery assets through a comprehensive 10-year scenario analysis. This work extends recent advancements in PEPF, highlighting the value of quantile forecasts in guiding battery selection, trading strategies, and economic outcomes in dynamic electricity markets.

The rest of this chapter is structured as follows: In Section 5.1, we explore the role of energy storage in electricity markets, focusing on its importance for trading and achieving renewable energy targets through efficient integration and price arbitrage. Section 5.2 provides an overview of our approach, models, specifics of the forecasting problem, and each trading strategy with its constraints. In the experimental results 5.3, we first assess the performance of each trading strategy and then simulate their impact on 10-year scenario analysis for four separate batteries. Finally, Section 5.4 concludes the chapter with observations and suggested future work.

5.1 Energy Storage and Trading Strategies to Address Curtailments

Modern power systems must transition to sustainable energy sources to meet rising demands and environmental goals. European electricity markets have experienced a remarkable surge in variable renewable generation, driven by policy incentives and renewable portfolio standards [FA23]. Over the past decade, Ireland has seen a rapid expansion of renewable generation in its electricity mix, rising from 21.7% in 2014 to 39.5% in 2022 [Eir22]. This integration brings volatility to the net power supply, as rapid and unpredictable changes in renewable output raise concerns about the reliability of power systems [MABH16]. In addition, curtailments from surplus renewable energy during low-demand periods hinder the expansion of renewable capacity. In Ireland, curtailment rates, illustrated in Figure 5.1, range from 3.0% to 5.9%, imposing significant economic burdens [YBC+22]. Many studies now focus on renewable integration, leveraging Machine Learning (ML) to forecast system penetration and manage these challenges [CMTP+24].

BESS are essential for mitigating supply-demand volatility from renewable output fluctuations, allowing efficient use of renewable energy, alleviating network constraints, improving grid efficiency, and facilitating electricity price arbitrage. They reduce reliance on fast-responding thermal plants and contribute to load shifting, power quality improvement, and overall grid stability. Moreover, the economic impact of curtailment, particularly energy wastage from inadequate demand, highlights the need for effective solutions to achieve a more sustainable and efficient energy future. The Irish Single Electricity Market includes the DAM for facilitating next-day trading through daily auctions and the BM for

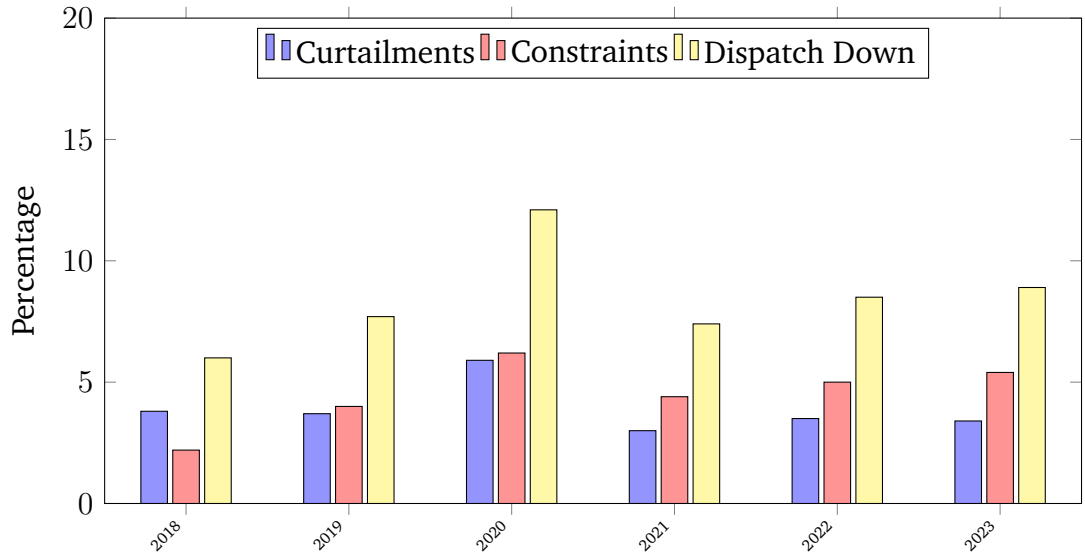


Figure 5.1: Annual percentages of curtailments, constraints, and dispatch down incidents, illustrating the impact of renewable energy challenges on wind output ([Eir23]).

balancing real-time supply and demand, both essential for grid stability. However, escalating market price volatility, fuelled by forecast errors in renewable electricity generation [DLM19, Mac20], presents challenges in maximising the potential of renewable energy integration [OT19]. Effective BESS arbitrage relies on accurate PEPF, increasing profitability while managing risk [BGK18].

5.2 Methodology

In this section, we outline the methodology used in our study to investigate the development of BESS trading strategies. These strategies encompass the optimisation, and in some cases, co-optimisation of revenue streams derived from participation in ancillary service markets, energy-only markets, or both. Just as ancillary service markets contribute to grid stability and reliability, BESS play an increasingly vital role in integrating renewable energy into the evolving energy landscape. To achieve this objective, our primary focus lies in the formulation of trading strategies aimed at maximising profits through energy arbitrage.

5.2.1 Forecasting and Trading Framework

In our analysis, the forecasting and trading approach is structured into five key steps, applicable to both markets. The first three steps follow the Analytical

Framework from Section 4.2 outlined in Chapter 4. Data collection, preparation, and pre-processing are carried out as described in Chapter 3. Following this the predictive models undergo hyperparameter optimisation for each three-month data subset, with model reliability ensured through a rolling walk-forward validation process where models are retrained every 8 hours using the latest data.

This chapter extends the previous framework by adding probabilistic forecasts and a trading layer for financial assessment, which includes quantile forecasting and strategy execution aimed at optimising profitability and risk management:

1. *Quantile Forecasting:* Following model optimisation, quantile forecasts are generated on unseen test data, with 24-hour forecasts for the DAM and 8-hour forecasts for the BM.
2. *Trading Strategy Execution:* Using the quantile predictions, we implemented a trading strategy that involves submitting bids and offers in both markets, purchasing during low-price periods, storing, and reselling during high-price periods.

This approach enables dynamic trading decisions based on predicted price distributions, optimising buy-sell operations for profitability and risk management.

Quantile Regression Models

All models in our study employ Quantile Regression (QR) from Section 2.3.3, a technique for estimating conditional quantiles, which allows the prediction of target values at specified quantiles (e.g., 0.1, 0.3, 0.5, 0.7, 0.9). These quantiles are used to define a "quantile pair," which establishes a forecast range, providing upper and lower bounds and offering potential values within a specified confidence level. For instance, a quantile pair of 0.1 and 0.9, denoted by $\alpha = 0.1$ and its complementary quantile $1 - \alpha = 0.9$, defines a forecast range covering the central 80% of possible outcomes, giving a nuanced view of the forecast distribution.

We utilise the same models as in Chapter 4, except for the removal of Auto-Regressive Integrated Moving Average, Naive, and Support Vector Regression models due to poor performance in forecasting accuracy. Lasso Estimated AutoRegressive was also removed due to consistently poor results, largely attributed to limited functionality in the available library. Additionally, we replaced Extreme Gradient Boosting with Light Gradient Boosting Machines (LGBM)

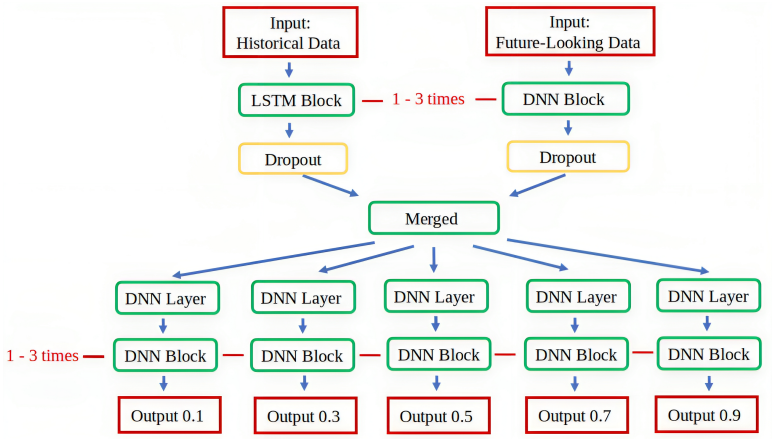


Figure 5.2: MH-RNN-DNN Model

to benefit from its faster training speed, lower memory usage, and better handling of high-dimensional datasets.

The models adapted for QR with the pinball loss function include:

- *K-Nearest Neighbors (KNN) (Section 2.2.2)*: A non-parametric algorithm that predicts based on the "K" nearest data points, capturing local patterns.
- *Random Forest (RF) (Section 2.2.2)*: An ensemble method that aggregates multiple decision trees to provide robust predictions.
- *LGBM (Section 2.2.2)*: A boosting algorithm that sequentially improves performance by focusing on data points with high residuals, making it effective for capturing complex temporal patterns.
- *Deep Neural Network (DNN) (Section 2.2.3)*: A versatile DL model with customisable hyperparameters, optimised for QR. The model predicts multiple quantiles using the pinball loss function, enabling robust forecasting under uncertainty. It processes both historical data and future projections, while a hybrid variant merges Long Short-Term Memory and DNN architectures. This hybrid design improves predictive capabilities by capturing sequential patterns from historical data and effectively incorporating future trends. The output layer produces forecasts for 24 hours (or 24 timestamps) in the DAM and 8 hours (or 16 timestamps) in the BM, capturing multiple quantile levels. See Figure 5.2 for a visual representation.

5.2.2 Trading Strategies

In this section, we present three trading strategies designed to optimise profits in electricity markets, considering constraints and price dynamics. These strategies are applied separately to the DAM and BM, while Section 5.2.2 introduces an adapted approach for holding DAM and BM positions concurrently.

These include:

1. *Trading Strategy 1 (TS1):* A rule-based heuristic approach utilising a straightforward single-trade strategy where the buy timestamp precedes the sell timestamp.
2. *Trading Strategy 2 (TS2):* An extension of TS1, increasing intraday trading frequency using an iterative multi-trade approach while maintaining the buy-before-sell constraint.
3. *Trading Strategy 3 (TS3):* An optimisation-based high-frequency strategy that enables flexible buy/sell timestamps while maintaining bottleneck-constrained purchase and sell volumes.

TS1: Single trade quantile strategy

TS1 is a rule-based heuristic trading strategy, incorporating quantile-based forecasts and BESS constraints similar to those in [UW21, Uni23b].

This strategy utilises quantile-based forecasts to optimise trading decisions involving a hypothetical battery with capacity B_c, no discharge limit, discharge efficiency $\mathcal{E}_d$, and charge efficiency $\mathcal{E}_c$. Over the time horizon of interest (i.e. 1 day for the DAM or 8 hours for the BM), a single buy-sell pair trade is permitted with the requirement that the buy trade occurs before the sell trade.

The key steps for TS1 are as follows:

1. *Identify Optimal Timestamps:* Determine timestamps for the minimum and maximum predicted prices for the specified quantile (α) and the complementary quantile (1-α), selecting the pair with the greatest price difference.
2. *Execute Trades:* Submit buy and sell orders at the identified timestamps.
3. *Calculate Profit:* Calculate profit as the difference between the sell and buy prices, adjusted for charge $\mathcal{E}_c$ and discharge $\mathcal{E}_d$ efficiencies.

The objective of the trading strategy is to maximise the expected profit by leveraging the quantile pair, $[\hat{q}_t^{\alpha}, \hat{q}_t^{1-\alpha}]$, forecasted prices, which are chosen to capture profitable trading opportunities based on predicted price levels. The objective function for this trading strategy is defined as:

$$\max \sum_{t=1}^{T} \left(\mathcal{E}_d \cdot \hat{q}_{t_2}^{\alpha} \cdot -\frac{\hat{q}_{t_1}^{1-\alpha} \cdot}{\mathcal{E}_c} \right) \tag{5.1}$$

where $\hat{q}_t^{\alpha}$ represents the model-predicted α-quantile price at time t, and $\hat{q}_t^{1-\alpha}$ denotes the complementary $(1-\alpha)$-quantile, which corresponds to the higher quantile price prediction, used here for comparison in deciding profitable trades. The total number of time periods, denoted as T, corresponds to 24 hourly periods for the DAM and 16 half-hour periods for the BM.

TS2: Multi-Trade strategy

TS2 extends TS1 by iteratively identifying multiple buy-sell pairs for the 24-hour period in DAM or 8-hour period in BM, maintaining the rule that buy timestamps precede sell timestamps while increasing intraday trading frequency.

The first two steps of TS2 are consistent with TS1. The subsequent three steps are specific to TS2:

3. *Create Additional Price Subsets:* To adhere to buy/sell constraints, the price data is divided into subsets, enabling the identification of new buy-sell pairs within each subset. The division typically occurs into one subset before the first trade (TS1) and another after it.

4. *Identify Buy-Sell Pairs:* Within each subset, as in TS1, identify the timestamps for the minimum and maximum predicted prices for the specified quantile (α) and the complementary quantile (1-α). For each pair within a subset, submit buy and sell orders for the pair with the greatest difference.

5. *Calculate Profit:* Calculate profit for each trade pair as the difference between the sell and buy prices, adjusted for efficiency, as in TS1.

While TS2 offers more trading opportunities than TS1, it is limited by the buy/sell order constraint.

TS3: High-Frequency Strategy

TS3 introduces an optimisation-based approach designed to tackle the inherent limitations of battery ramp rates. This challenge is particularly pronounced in larger batteries, where a full charge or discharge within a single trading cycle is typically unfeasible. By incorporating a bottleneck mechanism, TS3 enables greater flexibility in determining buy and sell timestamps, thereby removing the constraint requiring buy timestamps to precede sell timestamps.

The bottleneck in our research encompasses several real-world limitations modelled within TS3:

- **Ramp rate (R):** The maximum rate at which the battery can be charged or discharged, constrained by the battery's capacity and allowable charging speed. This limits the amount of energy that can be traded within a given time interval.
- **Discharge rate:** Constrained by the battery's current state of charge and the maximum allowable discharge rate, this determines how much energy can be released during trading.
- **Battery capacity (B_c):** Represents the total amount of energy available for trading. Additionally, a minimum charge level constraint is enforced to maintain battery health and operational reliability.
- **Energy efficiency ($\mathcal{E}_d$, $\mathcal{E}_c$):** Accounts for energy losses during charging and discharging processes. These efficiency factors are crucial for calculating actual profits from trading activities.

These constraints define the operational boundaries for TS3, influencing how the trading strategy is implemented. The key steps of TS3 are as follows:

1. *Find Min and Max Price Periods:* Similar to TS1 and TS2, this step involves identifying timestamps for the minimum predicted price for quantile (α) and the maximum predicted price for the complementary quantile (1-α). Unlike earlier strategies, the order of timestamps is flexible.
2. *Charging/Discharging Bottleneck:* This step focuses on the charging and discharging bottleneck. When the minimum price period precedes the maximum price period or vice versa, the algorithm dynamically adjusts the charging and discharging operations accordingly to maximise profitability. Further details on this bottleneck and the associated profit calculations are available in Algorithm 1.

3. *Iterate for Additional Trade Pairs:* The algorithm applies the same process to subsets of price data before, after, and in between each trade pair. This iterative approach allows for the exploration of additional trading opportunities and price subsets.

4. *Repeat for All Trade Pairs:* Continue iterating through all subsets until no further feasible buy/sell pairs are available.

Algorithm 1 Algorithm 1: TS3 Bottleneck

Initialise variables: c *(charge level)*, x_{buy} *(charging quantity)*, x_{sell} *(discharging quantity)*, p_{min} *(min price)*, p_{max} *(max price)*, *Profit*
Initialise constants: R *(ramp rate)*, B_c *(battery capacity)*, C_{min} *(min charge level)*
Charging/Discharging Bottlenecks
if Period of p_{min} < Period of p_{max} **then**
    Charging Strategy:
    $x_{buy} \leftarrow \min(B_c - c,\ R)$ ▷ Charging Bottleneck
    $c \leftarrow c + x_{buy}$
    $x_{sell} \leftarrow \min(c - C_{min},\ R)$ ▷ Discharging Bottleneck
    $c \leftarrow c - x_{sell}$
    Profit $\leftarrow (p_{max} \times x_{sell} \times \mathcal{E}_d) - (\frac{p_{min} \times x_{buy}}{\mathcal{E}_c})$
else
    Discharging Strategy:
    $x_{sell} \leftarrow \min(c - C_{min},\ R)$ ▷ Discharging Bottleneck
    $c \leftarrow c - x_{sell}$
    $x_{buy} \leftarrow \min(B_c - c,\ R)$ ▷ Charging Bottleneck
    $c \leftarrow c + x_{buy}$
    Profit $\leftarrow (p_{max} \times x_{sell} \times \mathcal{E}_d) - (\frac{p_{min} \times x_{buy}}{\mathcal{E}_c})$

The bottleneck mechanism dynamically adjusts the charging/discharging strategy based on whether the minimum price period occurs before or after the maximum price period. The algorithm ensures optimal trading decisions are made based on energy efficiency, ramp rate, and available battery capacity.

TS3-Dual: Dual Markets

Our dual markets approach, *TS3-Dual*, optimises trading decisions in both markets concurrently, ensuring synchronisation between them. This strategy capitalises on differences in granularity and lead times between the two markets:

- DAM positions are determined at time t for the time range of $t + 12$ to $t + 36$, with hourly granularity, covering a 24-hour period.

- BM positions, in contrast, are generated at time $t + 12$, 12 hours after DAM positions have been established. These BM positions are based on an 8-hour ahead forecast extending to $t + 20$, with half-hourly granularity.

The emphasis in TS3-Dual is on midday trading opportunities in the DAM, while the remaining time before and after DAM trades is allocated to the BM, capitalising on BM's consistent volatility. To achieve this, we create subsets of price data, denoted as period subset (PS), to optimise trades in both markets, divided as follows:

- Early-Bidding in the BM (PS1): Tailored for BM trades, PS1 focuses on time periods occurring before any DAM trades. It provides an opportunity for strategic planning and early bidding to optimise trading decisions in the BM market.
- DAM Trades - PS2: Dedicated to the DAM, this subset covers the time range starting from the initial trade pair's minimum period to the maximum period.
- Late-Bidding in the BM - PS3: Similar to PS1, PS3 serves BM trades but encompasses periods after DAM trades. It is strategically positioned to accommodate late bidding activities.

We follow the outline for TS3 in each subset, iterating through all subsets until no more purchase and sell pairs are available. Organising price data into these subsets enables TS3-Dual to efficiently manage trading opportunities in both the DAM and BM while ensuring synchronisation between them, as can be seen in Figure 5.3.

Code for all trading strategies can be found at GitHub[1].

5.3 Experimental Results

In this section, we present the results of our study, focusing on the practical applications of quantile forecasts in BESS trading within the DAM and BM. Our analysis evaluates three distinct trading strategies: TS1, TS2, and TS3, each incorporating quantile forecasts alongside trading constraints to optimise efficiency. Subsequent sections offer a detailed exploration of these strategies, exploring quantile pair selection, model choice, forecast accuracy, and highlight-

[1] `https://github.com/ciaranoc123/DAM-BM_Trading`

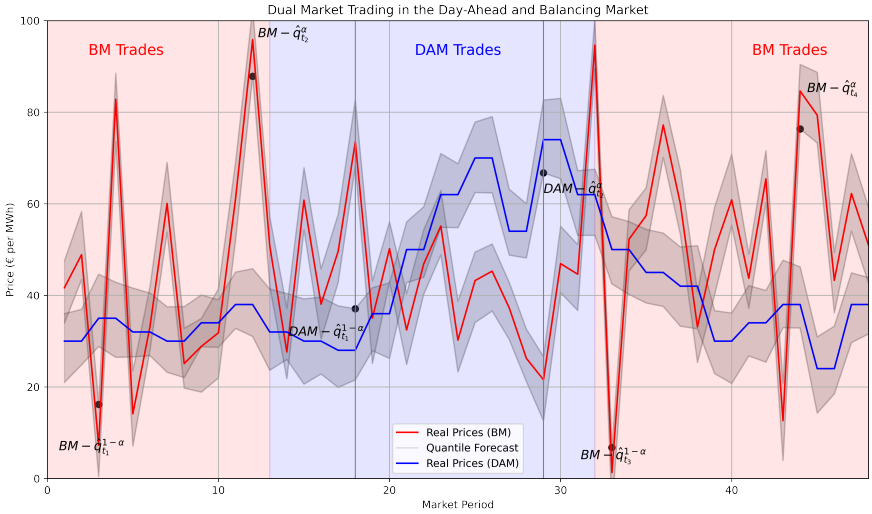


Figure 5.3: TS3-Dual Strategy

ing key distinctions between the DAM and BM that influence trading outcomes. Furthermore, we conduct a thorough BESS economic analysis, highlighting the economic implications of these strategies within both markets.

5.3.1 Trading Strategy Evaluation

This analysis involves a hypothetical battery with a capacity B_c of 1 MWh, an 80% discharge efficiency ($\mathcal{E}_d$), and a 98% charge efficiency ($\mathcal{E}_c$) [Uni23b]. Each strategy is tested over a 12-month period across both markets. To increase the likelihood of profitable trades, each strategy incorporates a minimum expected profit threshold. Specifically, a trade is executed only if the expected profit, calculated as $\mathcal{E}_d \cdot \hat{q}^{\alpha}_{t_2} - \frac{\hat{q}^{1-\alpha}_{t_1}}{\mathcal{E}_c}$, is non-negative. This threshold is particularly crucial in the BM, where increased volatility and unpredictable price dynamics amplify the risk of unprofitable trades. Furthermore, TS2, TS3, and TS3-Dual limit trading activity to a maximum of three trades per day, provided each trade meets the expected profit threshold. This limit is designed to prevent over-trading in low-margin conditions, thereby enhancing trade quality and supporting more consistent profitability across both markets.

TS3 emerges as the top-performing strategy in both markets, as illustrated in Figure 5.4. This improvement is due to its flexible timestamp order, which enables increased trading frequency compared to TS2, proving fundamental to maximising profits in a quantile-based context. By enabling non-sequential timestamp orders, TS3 capitalises on price fluctuations and manages battery state constraints more effectively. Despite the heightened risk in the BM of

trades falling short of the expected profit threshold, TS3 demonstrates a more aggressive profit increase due to the BM's half-hourly market structure, offering more trading opportunities compared to the DAM. TS3's ability to adjust buy/sell pairs based on quantile forecasts while managing battery constraints allows it to navigate BM's fluctuating conditions more effectively.

The inclusion of TS3-Dual, which synchronises trading between the DAM and BM, underscores the advantages of market diversification in energy trading by splitting trading windows across both markets to mitigate market-specific risks. This dual-market strategy leverages BM's intraday price fluctuations to identify profitable trades beyond the DAM alone. Table 5.1 provides a breakdown of financial outcomes achieved by TS3, evaluated across various quantile pairs and markets over the year. These results demonstrate each strategy's performance relative to benchmarks, including Perfect Foresight (PF) — a scenario where real market prices are used instead of forecasted prices, thus eliminating any forecasting error and representing the theoretical maximum profit achievable, and the 0.5-0.5 quantile pair serves as a baseline reference, representing the median forecast and allowing comparison with other quantile pairs to assess profitability across different confidence levels.

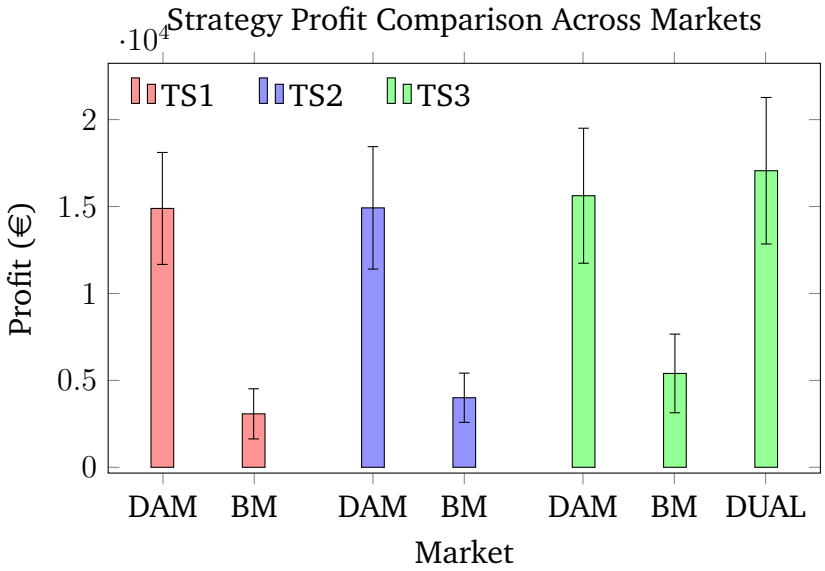


Figure 5.4: Average profit for TS1, TS2, & TS3 in each market. Error bars highlight the range for each quantile pair.

Trading in the day-ahead market

Across the selected trading strategies, all three strategies yield greater profits in the DAM than in the BM, with TS3 consistently outperforming TS1 and TS2, as

shown in Figure 5.4. Our analysis of TS3 in the DAM (Table 5.1) reveals that the baseline 0.5-0.5 quantile pair achieves the highest or near-highest profitability across all models. For KNN and LGBM, the 0.5-0.5 pair outperforms other quantile combinations, while for RF and DNN, it falls only slightly short of optimal pairs. This configuration produces the highest average profit of €17,011 across all models, capturing roughly 60% of the maximum potential profit (€28,212).

The relatively low error rates in the DAM limit the impact of quantile selection on profitability, suggesting that accounting for volatility is less critical than in the BM, where quantile selection has a more pronounced effect. In practical terms, this suggests that traders in the DAM could benefit more from focusing on efficiency-based strategies rather than complex quantile optimisations, as returns on such adjustments are modest in a more stable market environment.

Trading in the balancing market

In contrast to the DAM, the BM benefits more from quantile-based strategies due to its inherent volatility and forecasting challenges. Within TS3, as shown in Table 5.1, the 0.5-0.5 quantile averages €6,003, approximately 8% of the benchmark €71,669 achieved with PF. Higher quantile pairs such as 0.7-0.9 and 0.5-0.7, however, outperform this baseline 0.5-0.5 quantile pair. Specifically, the 0.7-0.9 quantile pair achieves the highest average profit of €7,052, approximately 9.8%, while the 0.5-0.7 pair follows closely with €6,966, representing 9.7%. The 0.1-0.9 quantile pair performs poorly in the BM due to its wider range, which captures more outliers and fails to align with the BM's high volatility and rapid price shifts. The 0.7-0.9 quantile pair is the top-performing choice for RF and LGBM, while KNN shows its best results with the 0.5-0.5 quantile, and DNN performs best with the 0.5-0.7 pair. This pattern suggests that higher quantile pairs may better capture extreme price values aligned with the BM's volatile nature, though results vary significantly among quantile pairs for TS3 and are highly model-dependent, as illustrated in Figure 5.5.

While BM's profitability consistently lags behind the DAM and PF benchmarks, with the best result achieving approximately 9.8% of the potential €71,669 from PF, this discrepancy highlights the forecasting difficulty and unpredictable nature of the BM. Nevertheless, it underscores the substantial, though challenging, profit potential for market participants who can accurately forecast in this volatile environment. The BM's frequent price spikes and dips present unique opportunities; with advanced models and strategic quantile choices, traders

Financial Profits from TS3: Bottleneck-Controlled Strategy										
Quantiles α		0.5-0.5	0.1-0.3	0.3-0.5	0.5-0.7	0.7-0.9	0.3-0.7	0.1-0.9	PF	Model Average
Market	**Model**									
DAM	KNN	**€14,764**	€12,570	€14,135	€13,557	€13,524	€13,036	€7,618	€28,212	€9,844
	RF	€18,722	€18,503	**€18,838**	€17,434	€16,916	€17,520	€15,015	€28,212	**€17,564**
	LGBM	**€19,420**	€16,446	€19,272	€18,032	€17,160	€18,178	€14,263	€28,212	€17,539
	DNN	€15,137	€14,593	**€15,740**	€14,914	€15,337	€15,196	€11,585	€28,212	€14,643
BM	KNN	**€4,358**	€1,047	€926	€788	€1,414	€68.22	€0	€71,669	€1,229
	RF	€7,860	€5,301	€7,955	€8,697	**€8,741**	€5,063	€0	€71,669	**€6,231**
	LGBM	€6,109	€4,408	€6,416	€9,391	**€9,713**	€5,630	€42	€71,669	€5,958
	DNN	€5,685	€4,898	€6,747	**€8,988**	€8,339	€5,664	€51	€71,669	€5,767
DUAL	KNN	**€16,070**	€12,883	€14,411	€13,792	€13,947	€13,055	€7,618	€82,683	€13,111
	RF	€21,079	€20,091	**€21,223**	€20,042	€19,537	€19,038	€15,015	€82,683	**€19,432**
	LGBM	**€21,250**	€17,766	€21,195	€20,848	€20,072	€19,866	€14,275	€82,683	€19,325
	DNN	€16,841	€16,061	€17,762	€17,609	**€17,837**	€16,894	€11,599	€82,683	€16,372
DAM	Average α	**€17,011**	€15,528	€16,996	€15,984	€15,734	€15,983	€12,120	€28,212	€15,622
BM	Average α	€6,003	€3,913	€5,511	€6,966	**€7,052**	€4,107	€23	€71,669	€4,797
DUAL	Average α	**€18,810**	€16,700	€18,648	€18,073	€17,848	€17,213	€12,126	€82,683	**€17,060**

Table 5.1: Financial Profits from TS3: Bottleneck-Controlled Strategy

can capitalise on these fluctuations. However, success demands precision and adaptability, as even small forecasting errors can greatly impact profits.

Trading Dual Market Positions

In the dual market approach, TS3-Dual capitalises on both the stability of the DAM and the volatility of the BM, aligning positions across both markets to maximise returns. As detailed in Table 5.1, the baseline 0.5-0.5 quantile pair consistently produces high profitability, achieving €18,810 and capturing 22.7% of the potential profit of €82,683. The 0.3-0.5 quantile pair follows closely, averaging €18,648, or 22.5% of the profit potential. This result highlights the adaptability of TS3-Dual in leveraging straightforward quantile pairs effectively, even as market conditions vary between the DAM and BM.

The LGBM model emerges as the top performer in TS3-Dual, attaining a profit of €21,250 with the 0.5-0.5 quantile pair, equivalent to 25.7% of the profit potential with PF. While the DAM achieves a higher percentage of profit potential alone (60%), its total profit remains lower compared to the dual market approach, highlighting the benefits of combining the DAM's stability with the BM's higher-reward potential, resulting in a more profitable outcome. Although the BM alone generates the lowest returns and percentage of the profit potential (9.8%), its inclusion in the dual market strategy raises total profits beyond what is achievable in either market individually. This outcome confirms the economic value of dual-market trading, with TS3-Dual consistently outperforming single-market strategies across all quantile pairs, as illustrated in Figure 5.4.

5.3.2 Model and Quantile Pair Performance

This section evaluates the profit performance of different quantile pairs and models across the DAM, BM, and the Dual market strategy for TS3 and TS3-Dual, providing insights into how model accuracy and quantile selection affect trading outcomes. Section 5.3.2 analyses the impact of non-conventional quantile pairs, while Section 5.3.2 explores the effects of model choice and forecast accuracy.

Performance of Non-Conventional Quantile Pairs

In analysing the performance of non-conventional quantile pairs for the DAM, BM, and Dual market strategies, it becomes apparent that the non-conventional

pairs (e.g., 0.3-0.5, 0.5-0.7, 0.7-0.9) provide more varied profit outcomes compared to conventional pairs (e.g., 0.5-0.5, 0.3-0.7, 0.1-0.9), see Figure 5.5. For DAM, non-conventional pairs show a relatively stable profit profile, with quantile pairs like 0.3-0.5 and 0.5-0.7 performing comparably to conventional pairs in terms of overall profitability and consistency. However, as seen from the narrower error bars in the DAM profits, these non-conventional pairs offer greater stability and less variability across different forecasting models, making them less sensitive to model-specific biases or fluctuations in predictions.

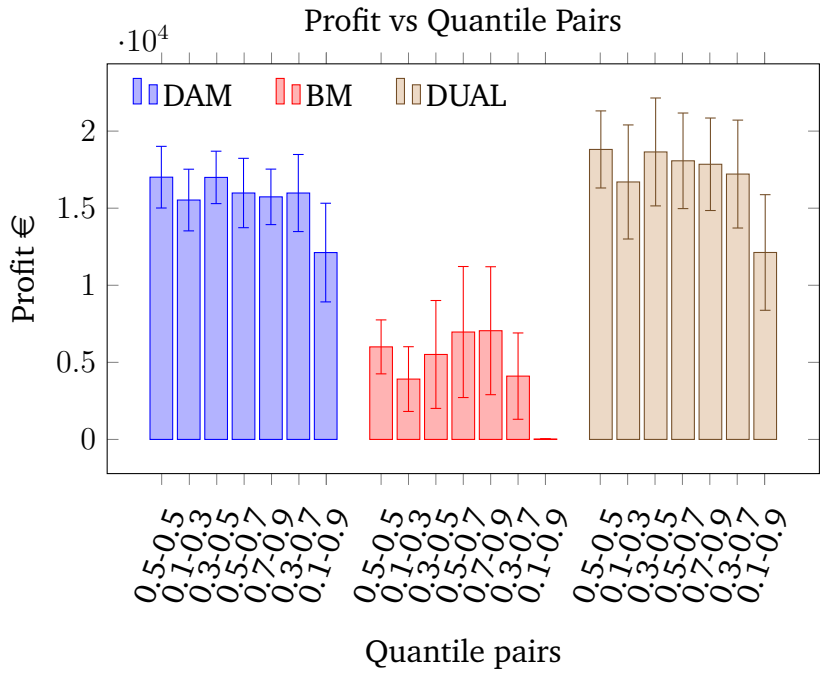


Figure 5.5: TS3 Profit for DAM, BM & DUAL quantile pairs. The error bars are relative to the different forecasting models.

In contrast, the BM shows lower profits and higher variability across both conventional and non-conventional pairs. Non-conventional pairs like 0.5-0.7 and 0.7-0.9 yield slightly higher profits than conventional pairs, though large error bars indicate model-dependent uncertainty. While the Dual strategy yields the highest profits for the baseline 0.5-0.5 quantile pair, non-conventional quantile pairs, particularly 0.3-0.5 and 0.5-0.7, also achieve consistent profits with moderate variability. Despite a general association between higher profits and increased variance in some pairs, these unconventional pairs effectively capture price fluctuations and manage risk, proving ideal for Dual market strategies.

Model Selection and Forecast Accuracy

This section highlights the importance of analysing model performance, forecast accuracy, and model selection to maximise profitability across different energy markets. Plot 5.6 illustrates the profit performance for different models (KNN, DNN, LGBM, and RF) for TS3 in the DAM, BM, and DUAL markets approach. In the DAM, the RF and LGBM models yield the highest profits with moderate variation, indicating stable performance across quantile pairs. The BM shows significantly lower profits for all models, with substantial variability, suggesting a high sensitivity to quantile pair selection and a less predictable environment. In the DUAL approach, RF and LGBM also perform well, with narrower error bars than BM but wider than DAM, showing a balance between profitability and sensitivity to quantile pairs. These results suggest that the RF and LGBM models are preferable for achieving stable outcomes in the DAM and BM markets, as well as in the combined DUAL approach.

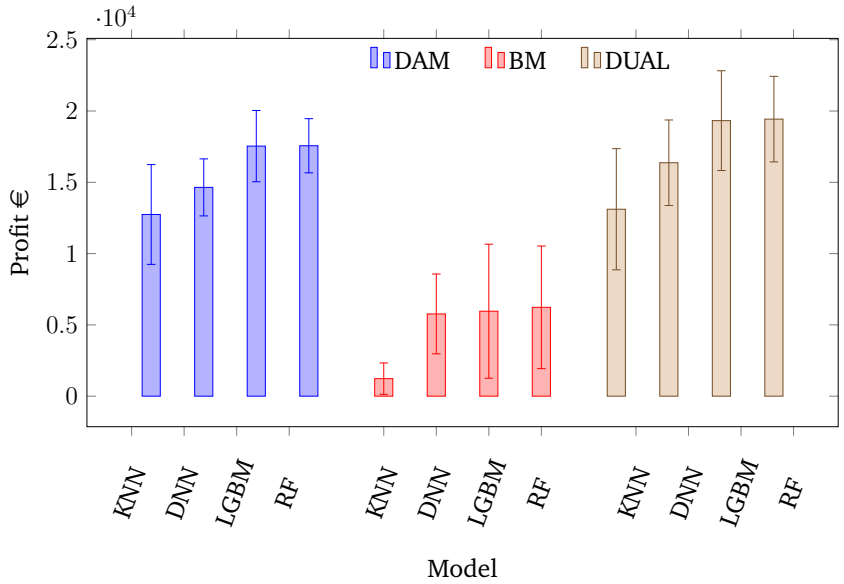


Figure 5.6: TS3 Profit for Models in the DAM, BM & Dual with Range in Quantile Pair Performance for each Model

The plot 5.7 highlights a trend between model accuracy (measured by pinball score) and profit across TS3 for the DAM, BM, and DUAL market approach. For TS3 in the DAM and TS3-DUAL, lower errors correlate strongly with higher profits, emphasising the impact of precise forecasting. However, the BM exhibits a more complex relationship; while lower accuracy does align with increased profits, the correlation is less pronounced for the first three models, suggesting that additional factors—such as price volatility and market-specific dynamics—might play a more substantial role in driving profitability in the BM.

Given the limited sample size of four models, these findings may require further validation. These findings underscore the fundamental role of forecast accuracy in driving profitability across the DAM, BM, and DUAL market approach, emphasising the need to prioritise high-accuracy models.

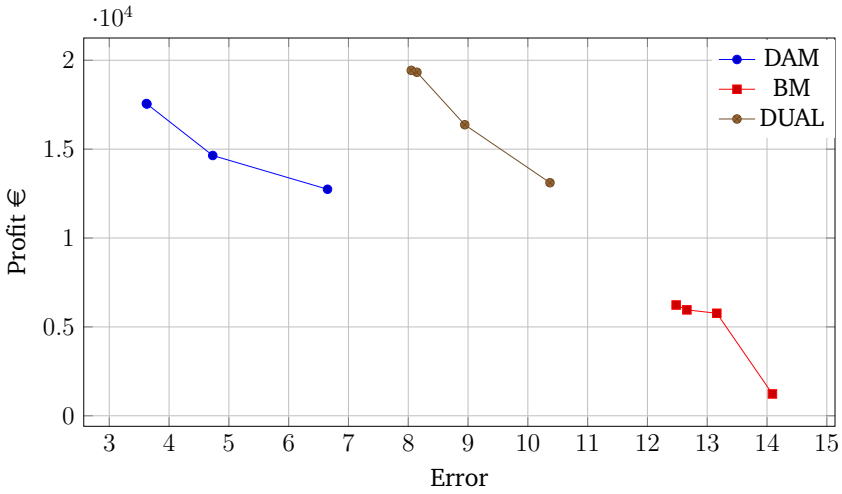


Figure 5.7: TS3 Aggregate Pinball Score of Model vs Profit

5.3.3 Economic Viability - Scenario Analysis

In this section, we present a comprehensive analysis of the economic viability of BESS configurations, exploring various factors that influence their returns and performance. Our study encompasses a scenario analysis focusing on key elements such as construction costs, maintenance expenses, degradation-related costs, efficiency losses, and market dynamics. The goal is to evaluate the profitability and economic outlook of different BESS units over a potential lifespan of 10-15 years.

In the following discussion, as detailed in [BB, Her18], we explore key factors influencing the returns derived from operating a BESS. These include:

- CAPEX: Battery purchase, including ancillary systems, installation, commissioning, permits, and civil works.
- Fixed O&M (€kW-year): Fixed maintenance costs and variable costs throughout the system's lifespan, based on operation hours or cycles. All batteries are modelled to have maintenance costs increasing at a rate of 2% per year, as per the Sandia report [RB18].
- Degradation-related costs: Expenses due to cycle-induced degradation,

potentially requiring cell addition or replacement, leading to reduced battery capacity and efficiency over time. Battery degradation is estimated at 1.55% per year, in line with lithium-ion battery data from [GMI+22].

- Efficiency costs: Costs related to round-trip efficiency losses (€/kWh). All battery options are modelled with a 95% charge & discharge efficiency.
- Depth of Discharge: Energy discharged as a percentage of rated capacity.
- Ramp Rate: Ratios of rated energy to rated power, e.g. 2-hour Tesla megapack, 1-hour lithium-ion battery in Sandia report [RB18].
- Calendar Life (years): Maximum lifespan from factors such as state of charge, e.g., Li-ion batteries end life below 60% rated energy. All batteries have an assumed lifespan of ∼10-15 years.
- Market registration costs for trading in I-SEM DAM and BM include an annual total of €18,294. These costs consist of a once-off entry fee, annual subscription fees, BM accession fee, BM participation fee, and variable trading fees, which apply per MWh traded. For a detailed fee breakdown, see [SEM23b] for I-SEM DAM & IDM and [SEM23a] for BM participation.

Items we will not factor in for issues on acquiring accurate data include:

- Warranty: (€/kWh-year): Annual fees for quality assurance.
- Insurance (€/kWh): Premiums against unforeseen risks.
- End-of-life costs: Expenses for disassembly, transportation to recycling facilities, and safe disposal of lithium-ion cells.
- Subsidies and government incentives: Potential financial support mechanisms that could offset initial investment costs. These vary widely in availability and impact.

Battery Options

We analyse various battery options, utilising configurations provided in [Com21], including smaller batteries with capacities of 3MWh (Scottish Power, Battery A) and 3.9 MWh (Tesla Megapack, Battery B) and larger batteries with capacities of 10MWh (Avolta storage, Battery C) and 39 MWh (10 Tesla Megapacks, Battery D). Revenue data for all batteries is derived from the LGBM model in TS3-Dual, utilising configurations specific to the selected battery setup.

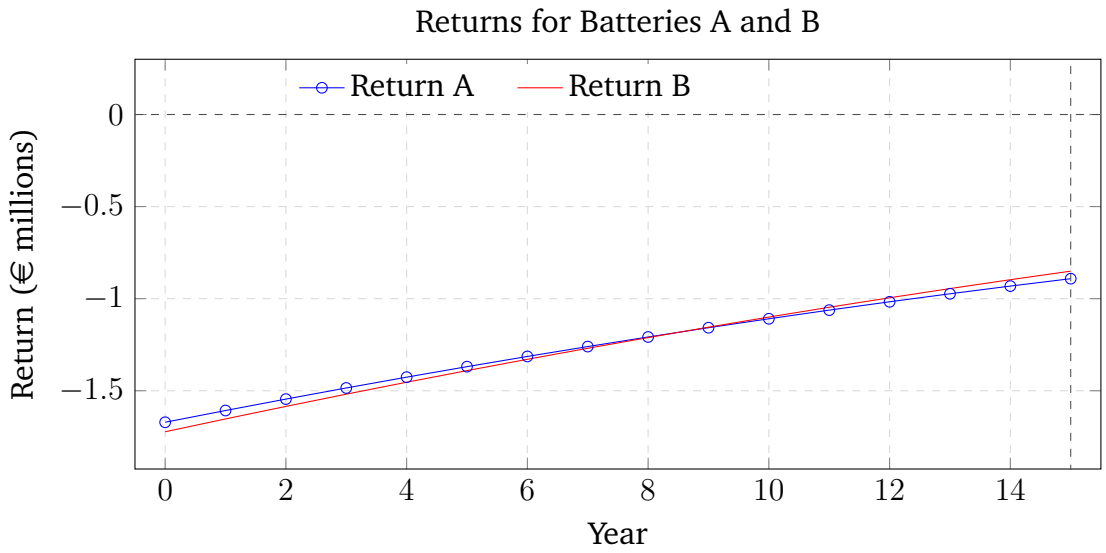


Figure 5.8: 15-year cost breakdown for the 3 MW Scottish Power and 3.9 MW Tesla capacity

Smaller Batteries: A & B, Larger Batteries: C & D

- Battery A, Scottish Power (3 MWh) - This battery features a 1-hour cycle, a maximum discharge capacity of 3 MWh, and operates with a charge/discharge efficiency of 95%. With a construction cost of €1,671,000.00 (€557 per kWh) and annual maintenance expenses of €11,000 (€4 per kWh), it represents a compact, efficient choice for shorter-duration applications.
- Battery B, Tesla Megapack (3.9 MWh) - Designed for a 2-hour cycle, Battery B shares similar charge/discharge efficiencies and ramp rates with Battery A. Its purchase costs amount to €1,722,628.98 (€447 per kWh), and annual maintenance costs are €7,730.96 (€2 per kWh). The extended capacity makes it well-suited for applications demanding longer discharge periods.
- Battery C, Avolta Storage (10 MWh) - This battery is configured with a 1-hour cycle, a maximum discharge capacity of 9 MWh, and a charge/discharge efficiency of 95%. With a construction cost of €4,850,000 (€485 per kWh) and annual maintenance costs of €100,000 (€10 per kWh), designed to balance capacity and efficiency, catering to applications with larger scale and short duration requirements.
- Battery D, Tesla Megapack*10 (38.5 MWh) - A larger-scale option, features a 2-hour cycle and shares similar charge/discharge efficiencies and ramp rates with Battery B. The purchase cost, accounting for a bulk dis-

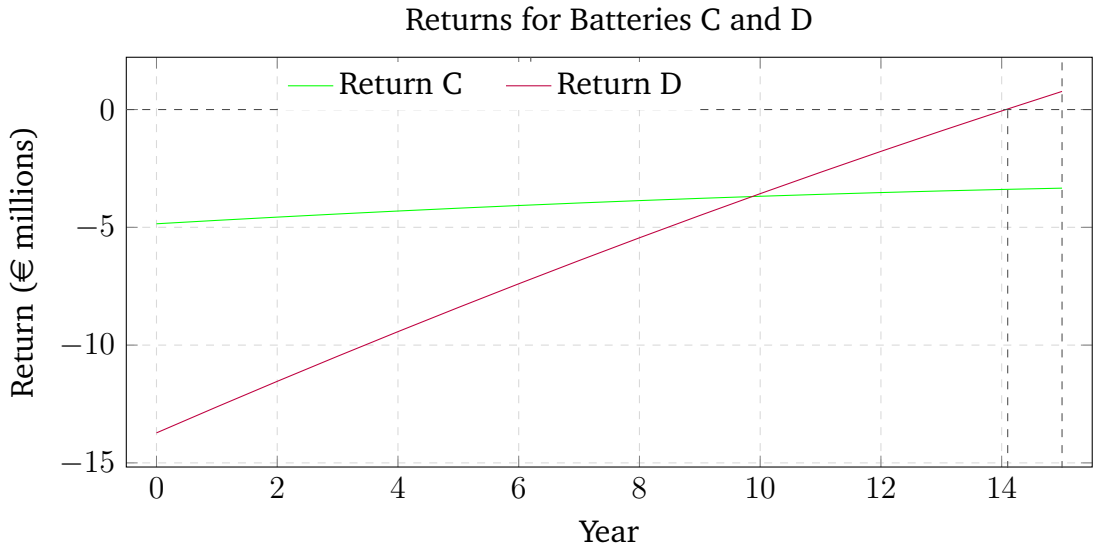


Figure 5.9: 15-year cost breakdown for Avolta Storage (C) and Tesla MegaPacks (D) totaling 38.5 MWh

count for ordering 10 units, is €13,726,186.17 (€356 per kWh), with annual maintenance costs of €61,593.89 (approximately €2 per kWh). This configuration, suitable for substantial energy storage needs, highlights the potential economic benefits of scaling up BESS installations.

Economic Viability of BESS Units

This section evaluates the economic viability of four BESS configurations, each with varying capacities and operational costs, across a 15-year horizon. Figures 5.8 and 5.9 provide detailed visualisations of the 15-year return on investment (ROI) for smaller (A and B) and larger (C and D) battery units. The analysis investigates each battery's path to profitability and highlights trends relevant to energy trading and investment decisions in the evolving energy storage market.

1. **Battery A (3 MWh, Scottish Power)** - With a construction cost of €557/kWh and annual maintenance costs of €11/kWh, Battery A remains unprofitable over its entire 15-year lifecycle. Despite a gradual reduction in losses, this unit ends with a total loss of €891,474 by year 15. Its performance reflects the limitations of smaller, short-duration storage options in terms of capital recovery, especially when operational in volatile markets without higher discharge capacities.

2. **Battery B (3.9 MWh, Tesla Megapack)** - Priced more competitively at €447/kWh with lower maintenance costs of €2/kWh, Battery B also fails

to achieve profitability, albeit with a slightly smaller end-of-life loss of €850,448. Its two-hour cycle time allows for longer discharge durations compared to Battery A, yet it remains economically constrained. This suggests that even with optimized efficiencies and slightly increased capacity, smaller batteries may struggle to be profitable in dual-market models, where higher-capacity, longer-duration options perform better.

3. **Battery C (10 MWh, Avolta Storage)** - As a mid-range storage option, Battery C incurs a construction cost of €485/kWh and maintenance expenses of €10/kWh. Despite offering a larger capacity suitable for frequent energy trading cycles, this battery also fails to break even, concluding with a loss of €3,337,555 by year 15. This result illustrates the challenges faced by medium-sized BESS units in balancing capital costs and revenue generation under typical market volatility, particularly when deployed in high-frequency, short-duration applications.

4. **Battery D (38.5 MWh, Tesla Megapack x10)** - The largest system, Battery D, achieves profitability, reaching a modest profit of €769,321 in its final operational year. With a significantly lower construction cost of €356/kWh (due to bulk purchasing) and maintenance costs of €2/kWh, this battery demonstrates that scale economies can significantly improve the financial outlook of BESS installations. Its two-hour cycle time and substantial capacity make it well-suited for dual-market operations, capitalising on both energy arbitrage and grid services for profitability.

These results emphasise the significant impact of battery size, cost structure, and operational efficiency on the economic viability of BESS in energy trading. Batteries A, B, and C remain unprofitable largely due to high upfront costs relative to their storage capacities and revenue potential, underscoring the importance of cost efficiencies achieved through bulk purchases and optimized maintenance strategies. The economic success of Battery D reflects industry trends that advocate for larger, consolidated BESS installations to achieve more favourable cost-per-kWh ratios. Additionally, larger storage capacities facilitate greater energy arbitrage opportunities and provide improved resilience against market fluctuations. Battery D's profitability, reached only in the final year, highlights the importance of aligning battery sizing with market demands, particularly in dual-market settings where larger units can better absorb price volatility and secure revenue through diverse applications.

Profitability in BESS also depends on strategic deployment across both the DAM

and BM markets. Larger capacities and longer cycle times, as seen with Battery D, enable more strategic positioning in response to price variations across both markets. This flexibility is less feasible with smaller units, which often lack the capacity to fully capitalise on market opportunities due to shorter discharge durations. For smaller BESS installations, profitability remains challenging without further advancements in cost reduction or market mechanisms that reward shorter-duration, high-frequency cycling. Conversely, larger-scale BESS are becoming viable investments, provided they are deployed with maximised charge/discharge flexibility and diversified revenue sources. In conclusion, the analysis reveals that achieving economies of scale, optimized cost structures, and operational flexibility is essential for BESS profitability. These findings highlight the need for continued cost reduction and technological advancements to make BESS viable across scales, with strategic sizing—particularly favouring larger installations in dual-market models—offering a pathway to strong economic outcomes in the evolving energy storage landscape.

The profitability challenges highlighted in this analysis underscore a critical need for targeted subsidies to support BESS deployment. Studies show that without subsidies addressing high capital costs and market integration barriers, economic viability often remains out of reach. For instance, [ZCDG21] emphasises that financial support mechanisms, such as self-consumption incentives and dynamic tariffs, are essential for enabling revenue generation and enhancing BESS appeal by leveraging renewable generation shifts and load management. Additionally, [FMC+23] argues that regulatory frameworks incorporating subsidies for peak-shaving or frequency regulation can significantly enhance grid stability and revenue consistency for storage operators in deregulated markets.

5.4 Conclusion

This study explored the role of quantile-based trading strategies in optimising BESS profitability across the DAM and BM. By examining three strategies—TS1, TS2, and TS3—and implementing a dual-market approach, TS3-Dual, we demonstrated the advantage of leveraging both markets to enhance returns. Among these, TS3-Dual proved the most effective, using flexible timestamp orders to balance the DAM's price stability with the BM's volatility, facilitating higher profitability through increased trading frequency. Our findings emphasise the significance of quantile pair selection, especially with high quan-

tiles such as 0.7-0.9 in the BM, which captured extreme price values effectively. In the DAM, the baseline 0.5-0.5 quantile pair generally performed best, highlighting the stable nature of this market. Additionally, model choice impacted profitability, with RF and LGBM delivering the most reliable results due to their adaptability and forecasting accuracy. The economic viability of BESS, particularly for larger configurations like Battery D, was underscored by the need for cost efficiencies and optimised battery sizing to sustain profitability in energy trading. This study demonstrates the potential of dual-market strategies for BESS trading and quantile forecasting in dynamic markets, underscoring the value of ongoing technological and methodological advancements to amplify BESS returns.

Future research could further enhance BESS trading by exploring alternative quantile forecasting techniques (see Chapter 6), implementing adaptive reinforcement learning approaches (e.g., [SK22]), and integrating intra-day markets for multi-market trading. These developments hold promise for strengthening BESS contributions to grid reliability and economic efficiency.

Chapter 6

Conformal Prediction for Probabilistic Electricity Price Forecasting

In the preceding chapters, we established a foundational understanding of Electricity Price Forecasting (EPF) and trading in electricity markets, focusing on the Day-Ahead Market (DAM) and Balancing Market (BM). Chapter 4 highlighted the limitations of point forecasts for capturing market uncertainties, pointing to the need for probabilistic methods. In Chapter 5, we address this suggestion, generating quantile forecasts for both markets and integrating these forecasts into trading strategies. While this provided an initial financial assessment of probabilistic methods, the forecasts were insufficient for achieving consistent profitability. The considerable shortfall between profits with quantile forecasts and those under Perfect Foresight (PF), particularly in the BM, where profits reach less than 10% of PF, highlights the challenges of forecasting BM volatility and underscores the need for improved probabilistic models to more reliably anticipate intraday price swings and capture rapid market fluctuations.

This chapter aims to address these gaps by advancing Probabilistic Electricity Price Forecasting (PEPF) for the DAM and BM using Conformal Prediction (CP) methods to improve the accuracy and reliability of forecasts. We extend CP techniques with time-series-adapted methods, specifically Ensemble Batch Prediction Intervals (EnbPI) and Sequential Predictive Conformal Inference (SPCI), to enhance forecast performance, with improved Prediction Interval (PI) coverage in both markets. Furthermore, to achieve a balance between the efficiency

of Quantile Regression (QR) and the validity of CP, we propose a new ensemble framework, Q-Ens, which combines QR, EnbPI, and SPCI for more accurate and reliable forecasts. This chapter further evaluates CP-enhanced forecasts using probabilistic metrics for efficiency and validity, along with financial performance analyses of TS1 and TS3 from Chapter 5 to assess practical impacts.

The rest of this chapter is structured as follows: In Section 6.1, we explore the importance of generating both accurate and reliable probabilistic forecasts in electricity markets. Section 6.2 details our methodological framework, including our forecasting models and probabilistic approaches. Section 6.3 discusses probabilistic metrics used to evaluate the efficiency and validity of our forecasts. The effectiveness of each forecasting approach, especially the proposed ensemble, is discussed through case study results in Section 6.4. Finally, Section 6.5 summarises our findings and suggests future work.

6.1 Beyond Efficiency Measures in Probabilistic Forecasting

EPF is crucial for market participants facing volatility, fluctuating demand, and the challenges of renewable integration. Power generators use forecasts to manage market risk and stabilise revenue, load-serving utilities rely on them to secure lower-cost energy and optimise procurement, and virtual traders leverage forecasts to estimate price spreads and exploit arbitrage opportunities. As renewable energy integration increases and smart grids evolve, accurate forecasting becomes more challenging, driving market volatility and complicating EPF in spot markets like the DAM, Intra-Day Market (IDM), and BM, where frequent price fluctuations due to renewable forecast errors pose significant challenges [OT19]. This underscores the need for accurate, reliable forecasts to stabilise production, optimise trading strategies, and promote policies that support financial health, grid stability, and sustainable energy goals.

There are two main EPF schemes: point forecasts and probabilistic forecasts. Extensive literature reviews cover these topics [NW18, ZS18], including our own review on uncertainty quantification in Section 2.3. Point forecasts provide a single price prediction at each time point, offering easy interpretation and a straightforward relationship between predictors and the target variable under homoscedasticity assumptions, but they lack uncertainty quantification. In contrast, probabilistic forecasts offer distributions that capture prediction

uncertainties. The shift towards PEPF in the DAM was highlighted by the 2014 Global Energy Forecasting Competition, which emphasised metrics like the pinball score, leading studies like [NW15] to propose Quantile Regression Averaging (QRA) for improved reliability in PEPF. Recent reviews in [KL22] have highlighted the relevance of probabilistic forecasting in the DAM for addressing uncertainties in smart grids, supply-demand dynamics, and price variations.

Despite the prevalence of QR and QRA, these methods have limitations in quantifying uncertainty, highlighting the need for alternative approaches. CP, introduced in [GVV98] and later extended to regression and classification domains by [VGS05] and [SV08], offers a promising alternative to QR by delivering forecasts that are both valid—meaning the predicted probabilities align closely with actual outcomes—and efficient, providing narrower, more precise PIs without losing accuracy. This combination of reliable alignment with outcomes and tight, actionable PIs makes CP especially valuable in settings requiring robust uncertainty quantification and adaptability without strict assumptions. This evolution addresses computational challenges while enhancing prediction accuracy and ensures that CP provides flexible PIs with predefined confidence levels across diverse models and scenarios. In PEPF, CP offers dynamic and symmetric PIs, providing consistent coverage with narrow PIs for managing market uncertainty [KZ21]. However, CP has limitations, requiring 25%-50% more historical data for calibration, and is not well suited for time series datasets with strong structural breaks, thus violating the exchangeability assumption. Recent adaptations for time series, such as EnbPI [XX21] and SPCI [XX23], offer promising, distribution-free PIs that do not rely on exchangeability assumptions.

6.2 Methodology

This section outlines our study's methodology, emphasising the key models and forecasting approaches. It is structured as follows: Section 6.2.1 introduces probabilistic forecasting models, followed by Section 6.2.2, which describes the uncertainty quantification approaches and our proposed ensemble.

6.2.1 Electricity Price Prediction Models

We use the same models as in Chapter 5, with two changes: Deep Neural Network models are excluded due to poor performance in TS3, and Lasso Estimated AutoRegressive (LEAR) is reintroduced. While LEAR was previously removed

due to its poor performance in quantile forecasts, its compatibility with model-agnostic CP libraries now makes it suitable for generating accurate probabilistic predictions. Thus, the models utilised in this chapter include:

- *LEAR (See Section 2.2.1)*: An Auto-Regressive model with Least Absolute Shrinkage and Selection Operator regularisation for feature selection and handling high-dimensional data.
- *K-Nearest Neighbors (KNN) (See Section 2.2.2)*: A non-parametric model predicting outputs based on the "K" nearest neighbors.
- *Random Forest (RF) (See Section 2.2.2)*: An ensemble model combining multiple regression trees to reduce variance and maintain low bias.
- *Light Gradient Boosting Machines (LGBM) (See Section 2.2.2)*: An ensemble model using boosting to reduce bias and variance.

6.2.2 Uncertainty Estimation

In the domain of EPF, probabilistic forecasting methods significantly advance beyond traditional point forecasts by capturing the uncertainty around future price estimates. This section examines various PEPF techniques for both the DAM and BM, covering methods such as QR, Split Conformal Prediction (SCP), EnbPI, SPCI, QRA, and our proposed QR-CP Ensemble (Q-Ens).

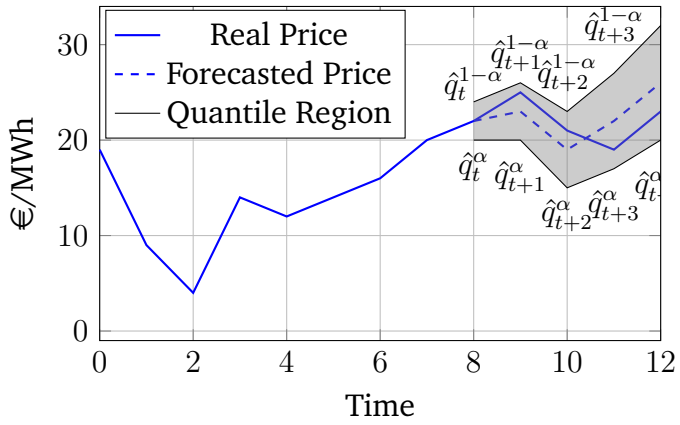


Figure 6.1: Quantile forecast of electricity prices

Our approach treats price prediction as a regression task, where the dataset $\mathcal{Z} = \{(\mathbf{x}_t, y_t)\}_{t=1}^{N}$ includes predictors $\mathbf{x}_t$ and corresponding response variables y_t. The conditional quantile function $Q_{y_t}(\alpha \mid \mathbf{x}_t)$ estimates the value below which a certain percentage α of the future price y_t is expected to fall, based

on predictors $\mathbf{x}_t$. We denote the lower quantile at level α as $\hat{q}_t^{\alpha}$ and the upper quantile at level $1-\alpha$ as $\hat{q}_t^{1-\alpha}$. In this probabilistic setting, as illustrated in Figure 6.1, the model f generates a PI defined by these quantiles:

$$\Gamma_{1-2\alpha}(\mathbf{x}_t) = [\hat{q}_t^{\alpha}, \hat{q}_t^{1-\alpha}] \tag{6.1}$$

This interval suggests that the future price y_t will fall within the range $[\hat{q}_t^{\alpha}, \hat{q}_t^{1-\alpha}]$ with a confidence level of $1-2\alpha$. For instance, if $\alpha = 0.1$, the PI represents an 80% confidence level ($1-2\alpha = 0.8$), implying an 80% probability that y_t will lie within this range. Further details on each method are in Chapter 2.

Quantile Regression

QR, as previously discussed in Section 2.3.3, extends ordinary least squares regression by estimating conditional quantiles of the response variable y_t, providing a richer perspective than mean estimation alone. QR reveals how predictors $\mathbf{x}_t$ influence the distribution of y_t at different quantiles.

1. *Model Formulation*: We start by expressing y_t as a function of predictors $\mathbf{x}_t$ and an error term ϵ_t:
$$y_t = \mathbf{x}_t^{\top}\beta + \epsilon_t \tag{6.2}$$
where y_t is the response variable for observation t, $\mathbf{x}_t$ is a K-dimensional vector of predictors, and β represents the regression coefficients.

2. *Objective*: QR aims to estimate the conditional quantile function, identifying values below which a proportion α of the data falls. The objective is to find coefficients $\hat{\beta}_{\alpha}$ that minimise a weighted sum of residuals, where the weights depend on the sign of the residuals and the quantile level α.

3. *Optimisation Problem*: For a quantile level α, the QR estimator $\hat{\beta}_{\alpha}$ solves:
$$\hat{\beta}_{\alpha} = \arg\min_{\beta} \sum_{t=1}^{N} \rho_{\alpha}(y_t - \mathbf{x}_t^{\top}\beta) \tag{6.3}$$
where $\rho_{\alpha}(u)$ is the "check function," which penalises over- and under-predictions differently based on α.

4. *Check Function Definition*: The check function $\rho_{\alpha}(u)$ is defined as:
$$\rho_{\alpha}(u) = \begin{cases} \alpha u, & \text{if } u \geq 0 \\ (\alpha - 1)u, & \text{if } u < 0 \end{cases} \tag{6.4}$$

This function gives a weight of α to positive residuals and $1-\alpha$ to negative residuals, enabling estimation of different quantiles by adjusting the penalties for over- and under-predictions.

5. *Properties of the QR Estimator*: The QR estimator $\hat{\beta}_\alpha$, defined in Equation 6.3, has well-established asymptotic properties. It is consistent, converging to the true value as sample size increases, and asymptotically normally distributed. Its covariance matrix is given by:

$$\text{Cov}(\hat{\beta}_\alpha) = \alpha(1-\alpha)\left[f_{y_t|\mathbf{x}_t}(\hat{q}_t^\alpha)^2 \cdot (X'X)^{-1}\right] \tag{6.5}$$

where $f_{y_t|\mathbf{x}_t}(\hat{q}_t^\alpha)$ is the conditional density of y_t at quantile α, which accounts for variability in the distribution around this quantile, and $(X'X)^{-1}$ adjusts for predictor scaling and correlations.

6. *Interval Prediction*: QR facilitates interval prediction by estimating upper and lower quantiles to form PIs. The PI for QR can be denoted as:

$$\Gamma_{1-2\alpha}(\mathbf{x}_t) = \left[\hat{q}_t^\alpha, \hat{q}_t^{1-\alpha}\right] \tag{6.6}$$

where $\hat{q}_t^\alpha$ and $\hat{q}_t^{1-\alpha}$ are the fitted quantile functions for the lower and upper bounds, respectively. For instance, $\alpha = 0.1$ gives an 80% confidence interval, illustrating QR's ability to capture distributional uncertainty.

Conformal Prediction

CP, as discussed in Section 2.3.4, is a flexible, distribution-free framework for quantifying the uncertainty of point predictions from models. CP generates prediction sets with a predefined confidence level, ensuring that the true value lies within these sets with a specified probability.

At the core of CP is the concept of nonconformity, which measures how unusual a new data point $(\mathbf{x}_{\text{new}}, y_{\text{new}})$ is compared to a set of previously observed examples. A nonconformity measure A is defined to assign a nonconformity score to each data point in the training set $\mathcal{Z}$. Traditional CP methods use p-values calculated from nonconformity scores to construct prediction sets. For a new example $\mathbf{x}_{\text{new}}$ with a candidate label y_{new}, the p-value is given by: $p(y_{\text{new}}) =$

$$\frac{|\{(\mathbf{x}_t, y_t) \in \mathcal{Z} : A(\mathbf{x}_t, y_t) \geq A(\mathbf{x}_{\text{new}}, y_{\text{new}})\}| + 1}{N+1} \tag{6.7}$$

where N is the size of $\mathcal{Z}$. The prediction set for $\mathbf{x}_{\text{new}}$ includes all candidate labels for which the p-value exceeds a certain significance level.

Split Conformal Prediction

SCP, introduced in Section 2.3.4 and also known as inductive CP, is a streamlined version of standard CP that improves computational efficiency by dividing the data into two distinct subsets: a training set and a calibration set. Here we outline the steps involved in applying SCP:

1. *Data Splitting*: The dataset $\mathcal{Z} = \{(\mathbf{x}_t, y_t)\}_{t=1}^{N}$ is split into two subsets:
   - *Training Set* $\mathcal{Z}_{\text{train}}$: Used to fit the predictive model.
   - *Calibration Set* $\mathcal{Z}_{\text{cal}}$: Used to compute nonconformity scores, measuring the degree to which predictions deviate from actual outcomes.
2. *Model Training*: A predictive model f is trained on the training set $\mathcal{Z}_{\text{train}}$, then used to generate predictions for the calibration set, enabling the computation of nonconformity scores.
3. *Nonconformity Score Calculation*: For each observation $(\mathbf{x}_t, y_t) \in \mathcal{Z}_{\text{cal}}$, the nonconformity score $A(\mathbf{x}_t, y_t)$ is calculated as:

$$A(\mathbf{x}_t, y_t) = |y_t - f(\mathbf{x}_t)| \tag{6.8}$$

   This score represents the absolute difference between the true value y_t and the predicted value $f(\mathbf{x}_t)$, with larger scores indicating greater nonconformity.
4. *Quantile Selection for Prediction Interval Construction*: After calculating nonconformity scores for the calibration set, a quantile λ_t (for time period t) is selected to define the desired confidence level $1 - 2\alpha$. This ensures that at least $1 - 2\alpha$ of calibration examples have nonconformity scores below λ_t, achieving PI coverage of $1 - 2\alpha$.
5. *Prediction Interval Construction*: For new input $\mathbf{x}_t$, the PI is constructed as:

$$\Gamma_{1-2\alpha}(\mathbf{x}_t) = [f(\mathbf{x}_t) - \lambda_t, f(\mathbf{x}_t) + \lambda_t] \tag{6.9}$$

   This interval provides a range of possible values for y_t, with coverage determined by quantile λ_t from the calibration set's nonconformity scores.

SCP balances computational efficiency and prediction robustness by calculating nonconformity scores on a smaller calibration set, avoiding the need to recompute predictions for each test example. This reduces computational load but relies on data exchangeability, limiting its applicability in time series settings.

Ensemble Batch Prediction Interval

EnbPI, outlined in Section 2.3.4, is a CP method designed to generate reliable PIs for non-exchangeable data, such as time series. EnbPI leverages ensemble learning and bootstrap sampling, achieving robust coverage without sacrificing data for calibration. This makes EnbPI computationally efficient and suitable for scenarios with complex data dependencies.

1. *Dataset and Bootstrap Sampling*: Given a dataset $\mathcal{Z}$, EnbPI generates B bootstrap samples from $\mathcal{Z}$. For each sample $\mathcal{Z}^{(b)}$, a predictive model $f^{(b)}$ is trained, creating an ensemble that reflects data variability.

2. *Point Prediction Generation*: Each model $f^{(b)}$ generates predictions for each observation in $\mathcal{Z}$. The ensemble prediction, $\hat{y}_t$, for input $\mathbf{x}_t$ is the average prediction across all models:

$$\hat{y}_t = \frac{1}{B} \sum_{b=1}^{B} f^{(b)}(\mathbf{x}_t) \tag{6.10}$$

3. *Nonconformity Score Calculation*: For each model $f^{(b)}$, residuals (nonconformity scores) are computed as the absolute difference between the observed value y_t and the predicted value $f^{(b)}(\mathbf{x}_t)$:

$$A^{(b)}(\mathbf{x}_t, y_t) = |y_t - f^{(b)}(\mathbf{x}_t)| \tag{6.11}$$

   These residuals capture prediction variability and help define the PI width.

4. *Quantile Selection for Prediction Interval*: For each model $f^{(b)}$, a quantile $\lambda_t^{(b)}$ at level $1 - 2\alpha$ is calculated from the residuals:

$$\lambda_t^{(b)} = \text{Quantile}_{1-2\alpha}(A_1^{(b)}, A_2^{(b)}, \ldots, A_N^{(b)}) \tag{6.12}$$

   The final threshold λ_t for PI construction is the average across all models:

$$\lambda_t = \frac{1}{B} \sum_{b=1}^{B} \lambda_t^{(b)} \tag{6.13}$$

5. *Prediction Interval Construction*: The PI for new input $\mathbf{x}_t$ is constructed as $\Gamma_{1-2\alpha}(\mathbf{x}_t)$, for a confidence level $1-2\alpha$, by averaging the predictions from all bootstrap models and incorporating the margin λ_t. The resulting PI is:

$$\Gamma_{1-2\alpha}(\mathbf{x}_t) = [\hat{y}_t - \lambda_t, \hat{y}_t + \lambda_t] \tag{6.14}$$

   This PI is expected to capture the true value y_t with confidence level $1-2\alpha$.

EnbPI avoids data-splitting by using leave-one-out residuals to leverage all available data, improving computational efficiency. This makes it well-suited for probabilistic forecasting in dynamic settings like energy markets, achieving consistent marginal coverage across the dataset and, in some cases, approximating conditional coverage, which adapts intervals to time-dependent patterns in the data.

Sequential Predictive Conformal Inference

SPCI, referenced in Section 2.3.4, extends EnbPI by adapting PIs to address time-dependent residuals and temporal dependencies. By using Quantile Regression Forests (QRF), SPCI estimates the conditional quantiles of residuals, making it effective for sequential, non-exchangeable time series data.

1. *Initial Model Training*: Train predictive model f on the dataset $\mathcal{Z}$. Calculate residuals ϵ_t over a rolling calibration set to capture prediction errors.
2. *Conditional Quantile Estimation with QRF*: Use QRF to estimate conditional quantiles of residuals from recent observations, capturing dependencies in the data. For each time t, QRF predicts $\hat{q}_t^{\alpha}$, allowing SPCI to adjust interval widths dynamically based on time-dependent uncertainty.
3. *Prediction Interval Construction*: Construct the PI for a new input $\mathbf{x}_t$ as:

$$\Gamma_{1-2\alpha}(\mathbf{x}_t) = [f(\mathbf{x}_t) - \lambda_t, f(\mathbf{x}_t) + \lambda_t] \tag{6.15}$$

   where λ_t is derived from the QRF-estimated quantiles of recent residuals, adjusting interval width to reflect current data patterns.
4. *Sequential Updating*: Update residuals by sliding the calibration window forward with each new observation, re-estimating quantiles to capture evolving patterns and adapt intervals accordingly.

SPCI's use of QRF for quantile estimation allows it to produce narrower PIs than

EnbPI, particularly in non-stationary settings, while maintaining valid coverage.

Quantile Regression Averaging

QRA, as described in Section 2.3.3, applies QR to a set of point forecasts from different models, allowing direct estimation of electricity spot price distribution without separating the probabilistic forecast into a point forecast and an error distribution. This method directly leverages electricity price distribution properties for improved forecasting accuracy. It considers M different point prediction models f^m and feeds their output to a probabilistic model:

$$\Gamma_{1-2\alpha}(\mathbf{x}_t, [f^1(\mathbf{x}_t), f^2(\mathbf{x}_t), \ldots, f^M(\mathbf{x}_t)]) = [\hat{q}_t^{\alpha}, \hat{q}_t^{1-\alpha}] \tag{6.16}$$

In our implementation, two sets of point forecasts are generated:

- QRA-R: Includes forecasts from models trained on different data contexts (full datasets, historical, and forward-looking variables) to capture diverse trends and improve reliability.
- QRA-CP: Consists of forecasts from various CP models like SCP, EnbPI, and SPCI, each contributing unique strengths to quantile predictions.

QR-CP Ensemble: A Robust Approach

In predictive modelling, uncertainty is categorised into epistemic and aleatoric types. Epistemic uncertainty arises from limited knowledge of the underlying model and can be reduced with more data or model improvements, while aleatoric uncertainty results from inherent variability in the data and cannot be mitigated by additional data. Our Q-Ens combines QR, EnbPI, and SPCI aims to address both uncertainties, improving the reliability and sharpness for PEPF. QR captures a wide range of possible outcomes by estimating multiple conditional quantiles, addressing both epistemic and aleatoric uncertainties. EnbPI reduces epistemic uncertainty by averaging predictions from bootstrap samples, lowering variance and reflecting aleatoric uncertainty in the constructed PIs. SPCI adapts dynamically to new data, updating PIs in real-time to reflect evolving data variability, capturing both types of uncertainty as they change over time.

The ensemble combines the strengths of these methods, working from a similar approach as EnbPI, averaging the models' outputs to both reduce epistemic uncertainty further, and improve overall forecast performance. The ensemble

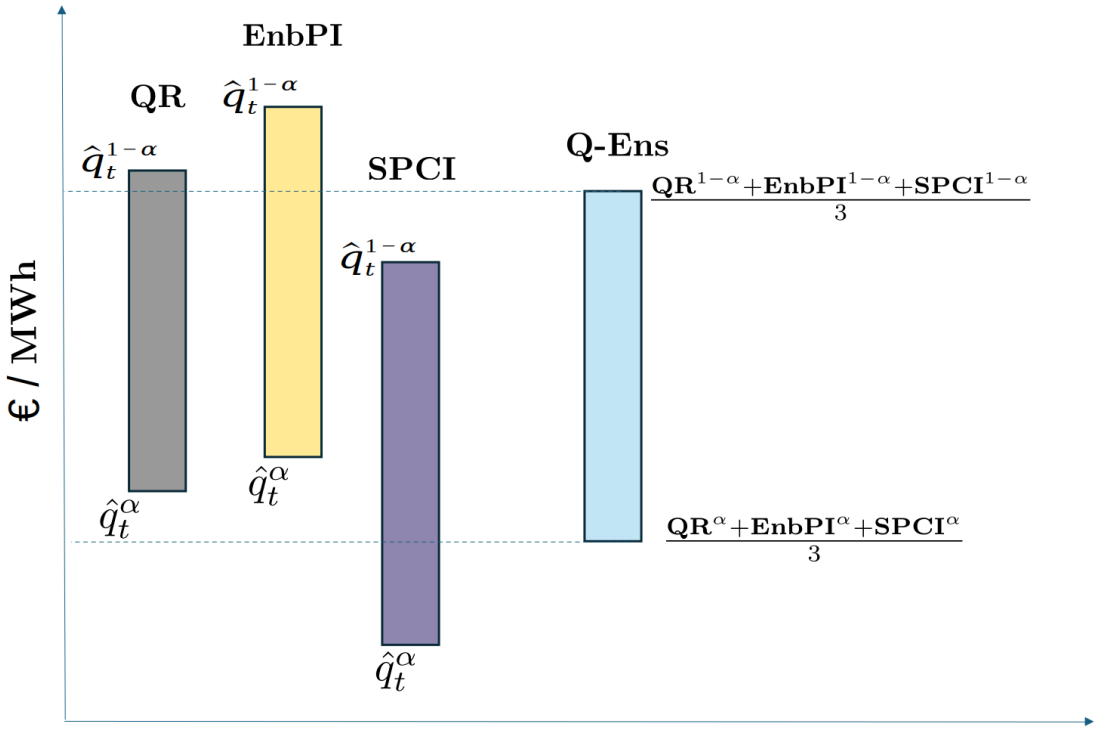


Figure 6.2: Construction of the Ensemble, Q-Ens.

prediction is obtained by averaging the quantile predictions from QR, EnbPI, and SPCI models: $\Gamma_{1-2\alpha} =$

$$[\frac{1}{3}(\text{QR}^{\alpha} + \text{EnbPI}^{\alpha} + \text{SPCI}^{\alpha}), \frac{1}{3}(\text{QR}^{1-\alpha} + \text{EnbPI}^{1-\alpha} + \text{SPCI}^{1-\alpha})] \tag{6.17}$$

with an outline of the PIs construction shown in Figure 6.2.

6.3 Probabilistic Forecasting Metrics

Our evaluation of probabilistic predictions hinges on two crucial dimensions: validity and efficiency. Efficiency, gauged through metrics like sharpness and interval width, enhances precision. Simultaneously, validity, including calibration and coverage, requires the generation of precise PIs to affirm the reliability of our forecasts.

6.3.1 Efficiency

Efficiency in probabilistic forecasting is crucial for precision. It is evaluated using two key metrics: the Pinball Score and Interval Width.

Pinball Score

In both the DAM and BM, sharpness plays an important role in accurate anticipation, hedging, and real-time adjustments. The Pinball Score [ZWS19], derived from the Pinball Loss function, measures the sharpness of the forecast based on the quantile prediction $\hat{q}_t^{\alpha}$ and the observed price y_t for time period t:

$$PS(\hat{q}_t^{\alpha}, y_t, \alpha) = \begin{cases} (1-\alpha)(\hat{q}_t^{\alpha} - y_t) & \text{for } y_t \leq \hat{q}_t^{\alpha} \\ \alpha(y_t - \hat{q}_t^{\alpha}) & \text{for } y_t \geq \hat{q}_t^{\alpha} \end{cases} \tag{6.18}$$

Here, α denotes the quantile level, such as 0.1 for the 10th percentile; hence, an 80% PI ($[\hat{q}_t^{\alpha}, \hat{q}_t^{1-\alpha}]$) corresponds to quantile levels 0.1 and 0.9.

To compare model performance over a given time horizon, we compute the Aggregate Pinball Score (APS) by aggregating the Pinball Scores across all forecasted quantiles and time periods. The APS provides a comprehensive measure of model efficiency by balancing sharpness across multiple quantiles and forecast steps, with lower values indicating sharper, more accurate predictions.

Interval Width

The Interval Width is another important metric for assessing the efficiency of probabilistic forecasts. While sharpness ensures reliability, efficiency is pivotal in refining precision, and it is intricately linked to the width of PIs. The Interval Width (IW_t) for a given PI can be defined as follows:

$$IW_t = \hat{q}_t^{1-\alpha} - \hat{q}_t^{\alpha} \tag{6.19}$$

where $\hat{q}_t^{1-\alpha}$ and $\hat{q}_t^{\alpha}$ are the upper and lower quantiles of the PI. Narrower intervals indicate higher efficiency, reflecting greater precision in the forecast.

6.3.2 Validity

Both reliability and accuracy of probabilistic forecasts are crucial, with a particular focus on evaluating their validity. Validity in probabilistic forecasting ensures that the PIs accurately reflect the true distribution of the observed data. Two key metrics for evaluating validity are coverage and the Winkler Score.

Coverage

We examine the models' coverage to highlight their ability to capture the true distribution, focusing on precision in capturing the price y_t within a predefined PI. The empirical coverage, indicating the agreement of predictions with the specified intervals, is denoted by the binary indicator I_t^α:

$$I_t^\alpha = \begin{cases} 1 & \text{if } y_t \in [\hat{q}_t^\alpha, \hat{q}_t^{1-\alpha}] \\ 0 & \text{if } y_t \notin [\hat{q}_t^\alpha, \hat{q}_t^{1-\alpha}] \end{cases} \tag{6.20}$$

The Coverage Probability (Λ) is then defined as:

$$\Lambda = \frac{1}{N} \sum_t^N I_t^\alpha \tag{6.21}$$

where N is the total number of forecasts. Ideally, the coverage probability should be close to the nominal level (e.g., 80% if α =0.1), indicating valid PIs.

Winkler Score

The Winkler Score (W_t) [TMB$^+$20] combines reliability and sharpness into a single metric, offering a comprehensive assessment of probabilistic forecasts. This score evaluates the interval width (IW_t), see Equation 6.19, of PIs based on the observed values (y_t), incorporating a penalty factor (τ) for deviations from the interval bounds:

$$W_t = \begin{cases} IW_t & \text{if } y_t \in [\hat{q}_t^\alpha, \hat{q}_t^{1-\alpha}] \\ IW_t + \frac{2}{\tau}(\hat{q}_t^\alpha - y_t) & \text{if } y_t < \hat{q}_t^\alpha \\ IW_t + \frac{2}{\tau}(y_t - \hat{q}_t^{1-\alpha}) & \text{if } y_t > \hat{q}_t^{1-\alpha} \end{cases} \tag{6.22}$$

The Winkler Score penalises larger deviations when observed prices fall outside the PIs, offering insight into the model's balance of accuracy and interval width.

6.4 Experimental Results

This section presents the results for each forecasting approach, with a focus on the performance of our QR-CP Ensemble (Q-Ens) on key metrics. We use the DAM and BM datasets from Chapter 3, and continue our analysis within the Irish Single Electricity Market (I-SEM), a market distinguished by high curtail-

ment rates due to surplus renewable generation during low-demand periods. Notably, in 2020, Northern Ireland and the Republic of Ireland had the highest curtailment rates in Europe, recorded at 13.5% and 11.1%, respectively [YBC+22].

Our evaluation examines the effectiveness of probabilistic forecasting methods across both the DAM (Section 6.4.1) and BM (Section 6.4.2) markets. The analysis is structured around two primary aspects: efficiency, which measures forecast sharpness through metrics like the APS and interval width, and validity, which assesses the reliability of forecasts based on coverage accuracy. These aspects are synthesized using the Winkler Score, which integrates both efficiency and validity into a single metric. This culminates in an assessment of each method's financial performance in Section 6.4.3.

For full reproducibility, the Python code used is available on GitHub[1].

6.4.1 Day-Ahead Market

The DAM dataset, introduced in Section 3.1.1, incorporates both historical and forward-looking data, providing insights into past market behaviour that support forecasting future conditions. Following the approach in Chapters 4 and 5, we use this dataset to predict DAM prices over the next 24 settlement periods by leveraging both types of data.

Efficiency in the Day-Ahead Market: Pinball Score & Interval Width

Table 6.1: APS for the DAM. Best approach for each regressor marked in bold, worst in red.

Model	QR	SCP	EnbPI	SPCI	QRA-R	QRA-CP	Q-Ens
KNN	6.65	7.79	**5.71**	6.09	6.20	6.01	5.96
LEAR	8.35	4.81	4.18	**4.08**	8.36	8.35	4.54
LGBM	3.62	3.67	3.76	3.71	3.67	3.66	**3.48**
RF	3.63	3.85	3.81	3.84	3.60	3.61	**3.59**
Avg.	5.56	5.03	**4.37**	4.43	5.43	5.36	4.39

Analysing the APS for DAM models in Table 6.1 highlights the strong performance for the Q-Ens approach, producing the lowest APS for RF and LGBM. This includes strong performance for CP approaches EnbPI and SPCI, in particular improving LEAR and KNN models, utilising CP's model agnostic feature.

[1] https://github.com/ciaranoc123/PEPF_Conformal

Time series adapted approaches EnbPI and SPCI also demonstrate notable improvements over SCP. QR and QRA models struggle, but this is largely as a result of their poor performance with LEAR, substantially reducing the average.

Looking further into efficiency, we evaluate the interval width of the models in Figure 6.3, centred around a median of 50 for clarity. The analysis of interval width across different models reveals key insights into model efficiency and precision; more efficient models produce smaller intervals. Figure 6.3 highlights

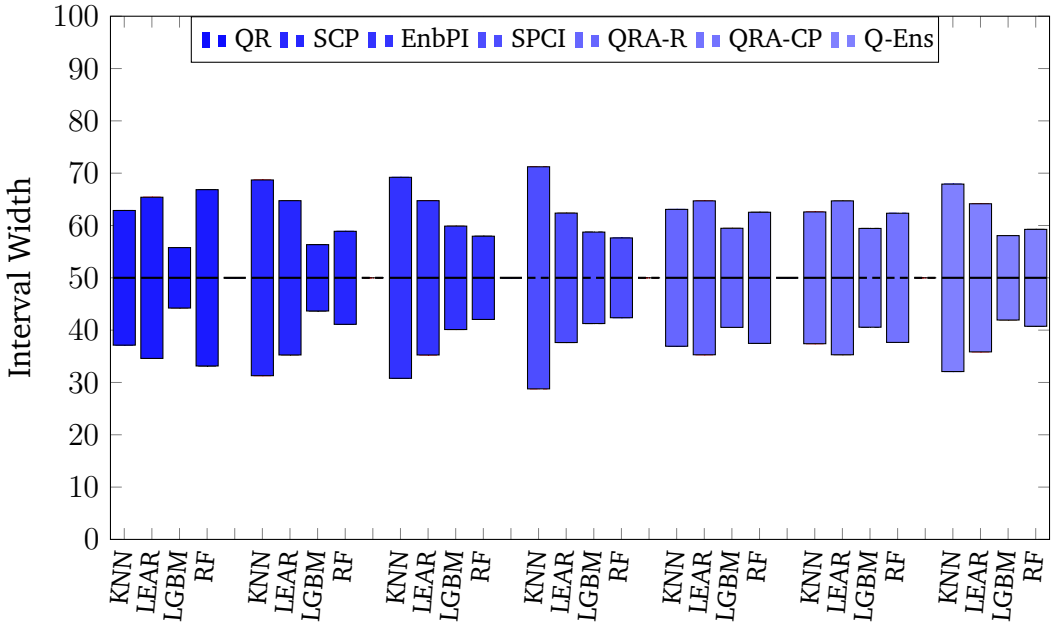


Figure 6.3: Interval Width for each model in the DAM

key differences between QR and CP approaches in interval width. CP methods (SCP, EnbPI, and SPCI) produce wider PIs for less accurate models like KNN and LEAR, while narrowing for more accurate ones like RF and LGBM. Notably, SPCI, designed to yield narrower intervals than EnbPI, succeeds marginally for LEAR, RF, and LGBM, but not for KNN. In contrast, QR and QRA approaches exhibit consistent interval widths independent of model accuracy, with QRA-CP and QRA-R producing similar intervals despite differing datasets. Q-Ens closely aligns with EnbPI and SPCI for KNN and LEAR. However, for LGBM, Q-Ens produces narrower PIs due to QR's inherently smaller widths, while for RF, it results in wider intervals due to QR's wider interval.

Validity in the Day-Ahead Market: Coverage

Figure 6.4 summarises coverage across the 0.1-0.9 quantile range in the DAM, highlighting each forecasting approaches' ability to achieve coverage. Aiming

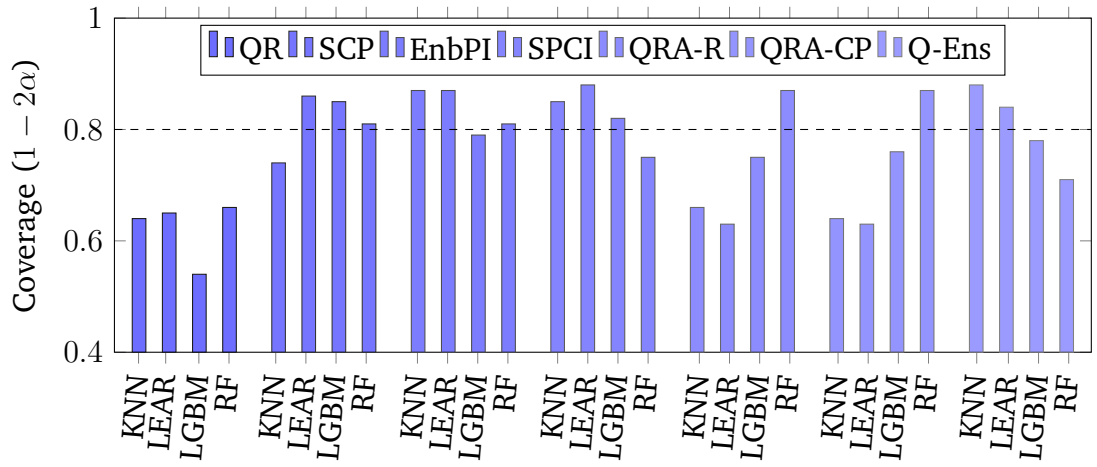


Figure 6.4: DAM Coverage for 0.1-0.9 quantile pair

for a coverage of 0.8 within the 0.1-0.9 quantile range, CP models consistently outperform QR counterparts. QR produces the worst results by a margin, failing to achieve coverage in a single model. Both QRA approaches begin to diverge from QR, improving coverage for most models, LGBM in particular, and achieving the desired 0.8 coverage for RF in both QRA-R and QRA-CP. Our Q-Ens approach yields slightly inferior results compared to EnbPI and SPCI. However, considering the lower APS for RF and LGBM models compared to EnbPI and SPCI, as shown in Table 6.1, this trade-off is justified.

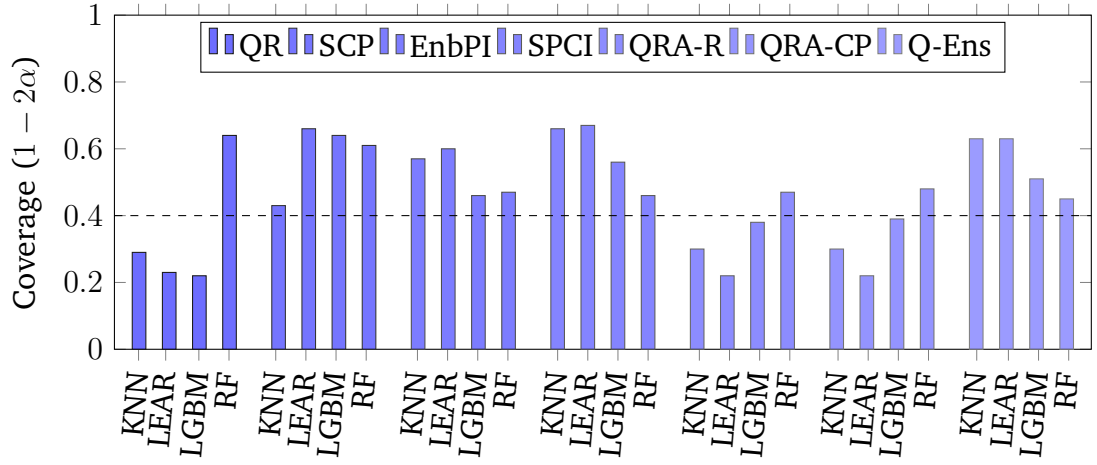


Figure 6.5: 0.3-0.7 quantile pair Coverage

Focusing on the 0.3-0.7 quantile pair with a target of 0.4 in Figure 6.4, we see a similar trend to Figure 6.4 with QR struggling and QRA improving QR for each model, with the exception of RF this time. All CP approaches achieve coverage for all models this time, highlighting the advantages of CP when it comes to improving a model's reliability. Our Q-Ens achieves coverage for all models,

with slightly lower scores than EnbPI and SPCI, but substantially better than QR, with the exception of RF.

Efficiency & Validity in the Day-Ahead Market: Winkler Score

Table 6.3 provides an overview of the Winkler Scores, highlighting each approaches' ability to tie together both efficiency and validity in our chosen forecasting models. Our approach Q-Ens excels producing the lowest average score, and the two lowest scores for all models with LGBM and RF.

Table 6.2: Winkler Scores for the DAM. Best approach for each regressor marked in bold, worst in red.

Model	QR	SCP	EnbPI	SPCI	QRA-R	QRA-CP	Q-Ens
KNN	57.02	64.16	51.47	56.16	53.70	**51.73**	54.33
LEAR	61.97	36.30	37.37	**35.06**	61.82	61.74	39.37
LGBM	29.87	30.39	32.76	31.31	31.50	31.47	**29.79**
RF	33.70	30.81	32.14	31.65	33.00	32.36	**29.99**
Avg.	42.94	40.42	38.43	38.55	45.01	44.33	**38.37**

Table 6.3: Winkler Scores for the DAM. Best approach for each regressor marked in bold, worst in red.

QR and QRA pay the price of their poor performance in coverage, producing the highest average Winkler scores. CP approaches produce lower Winkler scores, underscoring the benefits of their strong performance in coverage, with EnbPI strong performance just missing out on our Q-Ens approach.

6.4.2 Balancing Market

The BM dataset, introduced in Section 3.1.2, includes both historical and forward-looking data, providing insights into past market behaviour to support forecasting future conditions. Following the approach in Chapter 4, we forecast BM prices for the next 16 open settlement periods, beginning at $t + 2$. Although both markets have similar average prices (BM: €54.77/MWh, DAM: €53.77/MWh), the BM shows significantly higher price volatility, with a standard deviation of €64.60/MWh, compared to €37.38/MWh for the DAM.

Efficiency in the Balancing Market: Pinball Score & Interval Width

Analysing the APS for BM models in Table 6.4 follows a similar trend to Table 6.1 with Q-Ens showing strong performance producing the lowest average this

time, except we see EnbPI and SPCI, two of the top approaches for the DAM, produce the two highest average scores. SCP produces markedly better scores, with QR and QRA approaches also performing better than EnbPI and SPCI. The two lowest scores are achieved by LGBM with QRA-R and RF with QR.

Model	QR	SCP	EnbPI	SPCI	QRA-R	QRA-CP	Q-Ens
KNN	14.09	15.19	14.60	15.84	13.66	**13.49**	14.25
LEAR	15.22	**12.91**	13.96	13.79	15.33	15.22	13.21
LGBM	12.48	12.87	13.55	13.6	12.58	**12.41**	12.70
RF	**12.66**	13.01	13.34	13.48	12.79	12.89	12.88
Avg.	13.61	13.50	13.91	14.18	13.59	13.50	**13.26**

Table 6.4: APS for the BM. Best approach for each regressor marked in bold, worst in red.

Examining efficiency in the BM, we evaluate the interval width of the models in Figure 6.6. Compared to the average interval width for the DAM in Figure 6.3, all models have significantly wider interval widths, reflecting the greater inherent uncertainty in the BM. We see the same trend as the DAM with interval

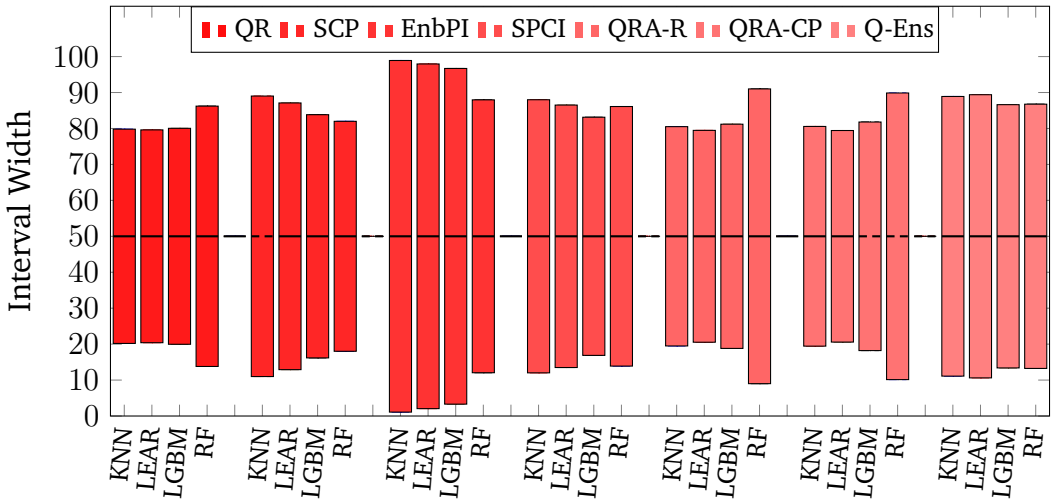


Figure 6.6: Interval Width for each model in the BM

width being independent of model accuracy for QR and QRA approaches, but again a strong correlation between interval width and model accuracy for all CP approaches, albeit less pronounced for SPCI than in the DAM. SPCI produces narrower interval widths than EnbPI as expected, but much more pronounced than in the DAM. Q-Ens produces wider interval widths than SPCI and QR, skewed by EnbPI large interval width.

Validity in the Balancing Market: Coverage

Figure 6.7 summarises coverage across the 0.1-0.9 quantile range in the BM. Aiming for a coverage of 0.8, all models come significantly closer to achieving the desired coverage compared to their performance in the DAM. CP approaches similarly excel, with most models achieving coverage. QRA again improves all QR models noticeably, with RF for both QRA approaches achieving coverage. Our Q-Ens fails to perform as well as any CP approach, but outperforms all QR and QRA models bar RF for QRA.

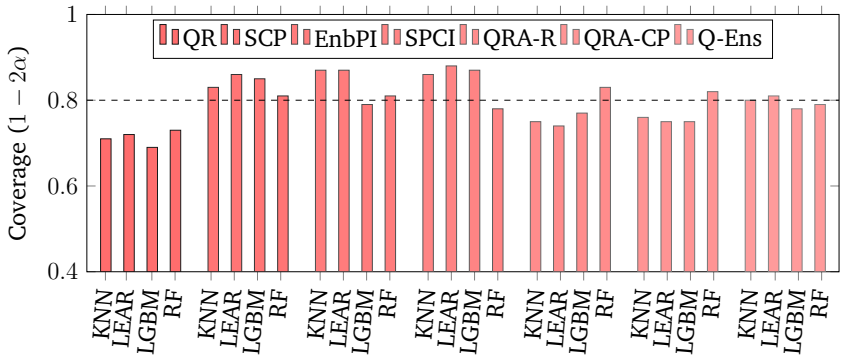


Figure 6.7: BM Coverage for 0.1-0.9 quantile pair

Focusing on the 0.3-0.7 quantile pair aiming for a coverage of 0.4 in Figure 6.8, again all CP approaches achieve coverage, including Q-Ens. QRA fails to

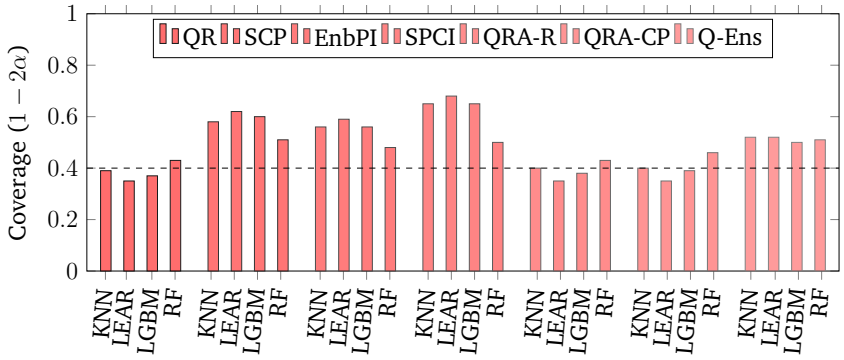


Figure 6.8: BM Coverage for 0.3-0.7 quantile pair

demonstrate a noticeable improvement off of QR, but all models come close to achieving coverage, with RF for QR and both QRA achieving the 0.4 target.

Efficiency & Validity in the Balancing Market: Winkler Score

Table 6.5 provides a overview for Winkler Scores in the BM, offering insights into each approach's ability to tie efficiency and validity together. The analysis

highlights the strong performance of our ensemble approach again, resulting in the second lowest Winkler score, slightly underperforming QRA-CP. QR and QRA demonstrate considerably improved performance than they did for Winkler scores in the DAM, with QRA-CP coming first, QR third, and QRA-R fourth. The last 3 spots are occupied by all three CP approaches with SPCI changing from just 0.07 off the top approach in the DAM, to the worst performer in the BM. QR achieves the lowest score of any model with LGBM, while Q-Ens has the second lowest with RF. These results are likely due to the closer coverage scores and narrower interval widths for the QR and QRA approaches.

Model	QR	SCP	EnbPI	SPCI	QRA-R	QRA-CP	Q-Ens
KNN	124.18	134.77	131.35	132.18	**119.21**	120.89	126.35
LEAR	127.79	**115.02**	122.63	126.16	129.03	127.75	124.09
LGBM	**107.34**	115.77	111.18	122.41	109.47	108.34	111.18
RF	114.37	113.88	111.84	114.32	116.15	115.07	**111.39**
Avg.	118.42	119.86	119.25	123.77	118.47	**118.01**	118.25

Table 6.5: Winkler Scores for the BM. Best approach for each regressor marked in bold, worst in red.

6.4.3 Financial Performance Analysis

For our financial evaluation, we utilise a Battery Energy Storage Systems (BESS). This section outlines our BESS trading strategy, crucial for ensuring grid stability and seamlessly integrating renewable energy into the dynamic energy landscape. This consists of our Trading Strategy 1 (TS1) and our high-frequency Trading Strategy 3 (TS3):

- TS1 is the same heuristic trading strategy found in Section 5.2.2. As in Section 5.2.2, this strategy utilises quantile-based forecasts to optimise trading decisions involving a hypothetical 1 MWh battery with no discharge limit, 80% discharge efficiency, and 98% charge efficiency. Over the specified time horizon (1 day for DAM or 8 hours for BM), a single buy-sell pair trade is permitted, with the requirement the buy trade occur before the sell trade.
- TS3 (Section 5.2.2) leverages optimisation techniques to maximize profits in electricity markets, limited to a maximum of 3 trades per day. It incorporates flexible buy/sell timestamps within bottleneck-constrained volumes, considering realistic constraints such as battery capacity, charg-

ing speed, discharging rate, minimum charge level, and energy efficiency.

- TS3-Dual (Section 5.2.2) is our extension of TS3 to optimise trading in both the DAM and BM by leveraging differences in market granularity and lead times. By synchronising positions across both markets, TS3-Dual maximises trading opportunities, focusing on midday trades in the DAM and taking advantage of the BM's higher volatility before and after DAM trades.

For more detailed information on the implementation of each trading strategy see Chapter 5.

Model	TS1 DAM	TS1 BM	TS3 DAM	TS3 BM	TS3-Dual
QR	€16,261	**€4,332**	€17,551	**€6,095**	€19,378
SCP	€15,819	€3,528	€16,621	€5,462	€18,026
EnbPI	€16,688	€3,995	€17,486	€5,649	€18,959
SPCI	€16,370	€3,750	€17,347	€5,658	€18,852
QRA-R	€15,771	€3,657	€16,790	€5,562	€18,318
QRA-CP	€15,782	€4,290	€16,684	€5,897	€18,505
Q-Ens	**€16,816**	€4,143	**€18,183**	€5,982	**€19,803**

Table 6.6: Financial Performance Comparison of RF & LGBM Models. Best approach marked in bold, worst in red.

Table 6.6 presents the financial performance of different forecasting approaches across trading strategies in both markets, focusing on top-performing models and excluding KNN and LEAR for clarity. SCP consistently yields the lowest returns, particularly in the BM and TS3-Dual, as well as in TS3 within the DAM, indicating it may struggle to capture the BM's volatility and the precision required for high-frequency trading strategies. QR consistently outperforms QRA, achieving better average results across all strategies, with QRA-R recording the lowest returns for TS1 in the DAM. However, QRA-CP demonstrates competitive results in the BM for TS1 and TS3, where it performs closer to QR. EnbPI and SPCI also exhibit solid performances: SPCI leads among CP methods in the DAM, while EnbPI shows an advantage in the BM and TS3-Dual, benefiting from its design for time series data. Notably, our QR-CP Ensemble (Q-Ens) achieves the highest average returns in the DAM and closely follows QR in the BM, achieving the top result overall in TS3-Dual. This consistent high performance highlights Q-Ens' ability to combine the strengths of individual models, making it particularly effective across the varying dynamics of both markets' trading environments.

Key Findings and Limitations

The application of CP techniques, particularly EnbPI and SPCI, improves the accuracy of PEPF in both markets, outperforming traditional QR models by providing more reliable PIs with consistent coverage. The ensemble approach, combining QR with EnbPI and SPCI, improves the balance between narrow intervals and coverage, resulting in more consistent and accurate forecasts. Financial analysis of these forecasting methods indicates that CP techniques, particularly our ensemble, Q-Ens, can increase profitability in trading algorithms; however, their dependence on large amounts of historical data and higher computational requirements may limit their practicality in real-time scenarios. While EnbPI and SPCI reduce computational overhead compared to full CP by avoiding data splitting (EnbPI) or using rolling windows (SPCI), their reliance on ensemble methods or residual bootstrapping still imposes higher costs than QR, particularly for high-frequency trading. These models can also face challenges when encountering extreme price events or sudden market shifts.

6.5 Conclusion

This study evaluated several probabilistic forecasting methods for electricity prices in the Irish DAM and BM, focusing on QR, QRA-R, QRA-CP, SCP, EnbPI, SPCI, and the proposed ensemble method Q-Ens. Our evaluation covered multiple dimensions—accuracy, reliability, interval width, coverage, and financial performance. We found that QR and QRA approaches performed well in efficiency metrics but fell short in coverage, where CP methods like SCP, EnbPI, and SPCI excelled. Notably, the ensemble approach (Q-Ens) demonstrated top-tier performance in both efficiency and validity, consistently delivering superior financial outcomes in single trade (TS1), high-frequency (TS3), and dual market (TS3-Dual) trading strategies, underscoring its practical value for market participants. The integration of CP techniques into EPF offers a promising way to improve forecast reliability and financial outcomes, especially as the energy market evolves and renewable energy integration increases. However, limitations were observed, particularly with SCP in the context of high-frequency trading and the BM's increased volatility, suggesting that CP methods may require further adaptation to handle such volatile conditions effectively. Additionally, dataset characteristics, such as high curtailment rates and market volatility specific to the I-SEM, influenced the forecasting performance, underscoring the importance of method selection based on market traits. The financial evalua-

tion, based on simulated trading, may not fully capture the complexities of real-world market conditions, including transaction costs, market liquidity, and sudden structural shifts. Therefore, further research should focus on testing these methods in diverse markets, including IDM, and exploring the practical limitations of CP models, particularly in environments with extreme price events or limited historical data. Future studies could also delve deeper into more granular 5-minute BM datasets and incorporate probabilistic forecasting frameworks that address transaction costs and market conditions. Additionally, improving the robustness of CP methods, particularly with respect to outliers and non-exchangeable time series, will be fundamental. Exploring the application of CP models in renewable energy forecasting, smart grid management, and real-time data streams presents a rich avenue for future development, which will help to create more adaptable and reliable forecasting systems in increasingly complex energy markets.

Chapter 7

Conclusions

This dissertation contributes to the field of electricity price forecasting and trading optimisation within the Irish Day-Ahead Market (DAM) and Balancing Market (BM). Addressing the challenges posed by renewable energy integration, this research develops a clear, standardised framework for BM forecasting, as well as new approaches surrounding dual-market trading strategies and probabilistic forecasting. The Irish electricity market presents unique challenges due to its high renewable energy penetration, particularly wind, and its isolated grid system, which limits interconnections with other markets. These factors lead to greater supply variability, higher curtailment levels, and increased price volatility compared to other regions. By improving forecasting accuracy and optimising trading within both markets, this work supports grid stability and market efficiency amid increasing renewable variability. We introduce structured methodologies that enhance both the approach to price prediction in the BM, as well as accuracy and trading strategy profitability. Furthermore, this research establishes a foundation for robust, adaptable, and economically viable forecasting and trading frameworks in electricity markets.

Review of Electricity Market Structure and Predictive Models

Chapter 2 provides an in-depth review of the structure and dynamics of electricity markets, along with a comprehensive analysis of the predictive models used within these contexts. By examining the distinct features and operational complexities of key electricity spot markets, the review highlights the need for tailored forecasting approaches that can adapt to varying market demands. Additionally, this chapter surveys a wide range of forecasting techniques, spanning from statistical and ML methods to probabilistic approaches that capture un-

certainty. The insights gained from this review directly inform the forecasting framework and trading strategies developed in this dissertation, highlighting the strengths and limitations of existing methods and underscoring the importance of probabilistic forecasts in effectively managing price volatility.

Irish Market Dataset Preparation and Analysis

In Chapter 3, we developed high-quality DAM and BM datasets, incorporating forward-looking and historical data from SEMO, with preprocessing focused on reproducibility and consistency. To ensure dataset suitability for forecasting, we conducted key time series tests, including stationarity, autocorrelation, partial autocorrelation, serial correlation, and heteroscedasticity checks. These tests revealed key statistical properties, guiding model selection and providing uniformly preprocessed datasets to support reproducibility and future research.

Forecasting Framework for the Balancing Market

In Chapter 4, we presented a standardised framework for forecasting BM prices, a market characterised by high volatility with limited research. By evaluating a range of predictive models—including statistical, ML, and DL approaches—this framework identifies models best suited to the BM's unique dynamics. The LEAR model was particularly effective for BM forecasts, outperforming DL models, which were superior in the DAM, underscoring the need for market-specific forecasting approaches tailored to a market's characteristics.

Dual-Market Trading for Improved Economic Viability

In Chapter 5, we introduced high-frequency trading strategies for the DAM and BM, optimised through the use of quantile-based forecasts. Building trading strategies around BESS demonstrated the economic advantages of increased trading frequency and dual-market participation. The Scenario analysis aims to offer practical tools for real-world market participants to improve trading precision and efficiency under varying market conditions. However, after accounting for BESS trading costs in both markets, losses for three systems and slight profits for the others suggest that alternative investments may yield better returns.

Integration of Conformal Prediction for Probabilistic Forecasts

Chapter 6 presents our contributions to PEPF by integrating CP techniques, including EnbPI and SPCI, into our forecasting framework. Unlike traditional quantile-based methods, CP techniques result in improved validity, providing consistent coverage, especially within the BM, where reliable PIs are crucial for effective risk management. By combining CP with QR, we introduced an ensemble method that balances efficiency and coverage, resulting in a more adaptive approach that strengthens financial outcomes for market participants.

In conclusion, this dissertation has made substantial advancements in EPF and trading optimisation within both the DAM and BM. Through a comprehensive literature review, the creation and analysis of DAM and BM datasets for future testing, the development of a standardised BM forecasting framework, the implementation of dual-market trading strategies, and the integration of CP methods, this research addresses the pressing challenges posed by renewable energy variability. The methodologies presented not only improve forecasting accuracy and trading profitability but also improve the resilience and adaptability of market operations. By establishing a clear and structured framework, this work provides researchers, traders, and policymakers with key tools to support market stability, optimise resource allocation, and manage risk. These contributions lay a foundation for future innovations in forecasting and trading, promoting a more efficient, sustainable, and reliable energy ecosystem.

7.1 Future Work

While this dissertation has made significant progress in forecasting and trading within the DAM and BM, several promising research directions remain. This section outlines potential areas for expansion, including increasing data granularity, expanding datasets with new variables, broadening applications to additional markets, refining trading methodologies, and developing advanced forecasting techniques. These directions provide opportunities to improve the adaptability, precision, and scalability of these approaches to meet the evolving demands of electricity markets amid the increasing renewable integration.

Balancing Market Forecasting with High-Frequency Data

Improving forecasting accuracy in the BM requires capturing the high-frequency, short-term dynamics that characterise this volatile market. Integrating 5-minute granularity datasets into BM forecasting can provide models with the temporal resolution needed to detect rapid fluctuations, price spikes, and market anomalies. This level of granularity can enable a more responsive forecasting framework, improving the precision of short-term predictions and supporting the grid's stability by anticipating imbalances in real time. With improved temporal granularity, forecasting models may better capture BM behaviour, offering market participants clearer insights into short-term risks and opportunities, which can contribute to improved reliability and stability in grid operations.

In addition to temporal granularity, expanding the dataset with additional predictors could further refine BM forecasting accuracy by incorporating factors influencing price movements. While DAM prices are currently used as a predictor for BM prices, adding Intra-Day Market (IDM) data could improve the model's responsiveness to intra-day fluctuations that impact the BM. Additionally, real-time data on generation forecasts, fuel prices, and renewable output could provide essential context, allowing models to account for supply-demand imbalances caused by renewable variability, fuel cost shifts, and unexpected outages. Optimising predictor combinations and granularity levels can create a more adaptive, data-rich approach to account for short-term BM volatility.

Scalability Across Diverse Market Structures

Broadening the scope of forecasting and trading frameworks to include additional markets and regions would provide valuable insights into the adaptability, robustness, and scalability of the methodologies developed in this research. Extending these frameworks to the IDM and Ancillary Services Market would be a logical progression, as each of these markets presents unique forecasting challenges. The IDM operates with near-real-time adjustments, requiring models that can handle high-frequency data and rapidly shifting supply-demand dynamics, while the Ancillary Services Market focuses on maintaining grid reliability through services like frequency regulation, reserve capacity, and voltage control. Analysing these markets could lead to a comprehensive, multi-market forecasting and trading framework, offering improved profits, better risk management, as well as new insights into managing operational stability across

various market types, each with distinct regulatory and operational demands.

In parallel, assessing the scalability of these forecasting and trading methodologies across different geographic regions would help evaluate their generalisability and robustness under varied market conditions. Geographic regions often differ in their levels of renewable energy integration, regulatory frameworks, and grid infrastructure, all of which impact market volatility and trading behaviour. By adapting the methodologies to account for region-specific factors—such as renewable output, local pricing structures, and market coupling mechanisms—can enhance their global relevance and effectiveness, enabling more informed, context-sensitive decisions in diverse electricity markets.

Dynamic Trading Strategies with Real-Time Data Integration

Future work in trading optimisation could leverage models that integrate closer to real-time data, helping better adapt to market fluctuations. Building upon the trading strategies outlined in Chapter 5, leveraging high-frequency, real-time data—such as market prices, intra-hour renewable generation forecasts, and weather updates—could improve automated trading frameworks that continuously adjust to evolving conditions. In fast-moving environments like the BM, this dynamic adjustment could increase forecasting precision and strategy responsiveness, improving the accuracy of real-time decisions.

Reinforcement Learning (RL) offers a promising direction for adaptive trading strategies, enabling models to learn optimal actions over time by refining policies through feedback from market interactions. In highly volatile settings like the BM, RL agents can help develop policies that respond to shifts in price, supply, and demand, while dynamic programming could complement RL by optimising multi-stage trading decisions, improving performance across multiple markets. Such techniques are especially valuable in multi-market trading, where strategies need to align real-time adjustments in the BM with forecasts made in the DAM. Future research could explore these adaptive methods within multi-market trading contexts, focusing on high-frequency data conditions to optimise profitability, resilience, and strategic alignment between markets.

Moreover, understanding the impact of trading volume on price dynamics is critical for refining trading strategies, especially in the BM where volume fluctuations can amplify price sensitivity. As the volume traded in the BM grows, price elasticity will increase, affecting both risk and profitability, necessitating

adaptive strategies that account for volume effects in trading. Investigating the relationship between trading volume and price response, including factors such as liquidity and supply-demand elasticity, would be invaluable for market participants. Such insights could improve market participation models by equipping trading frameworks with volume-driven price adjustments, ultimately improving risk management and strategic positioning across electricity markets.

Hybrid and Interpretable Forecasting Models

Future advancements in forecasting could prioritise developing models that are not only highly adaptable and reliable but also interpretable for broader industry adoption. Hybrid forecasting models, which integrate statistical methods with ML and DL techniques, are especially promising for volatile and complex environments like the BM. These models can combine the interpretability of statistical techniques with the non-linear modelling capabilities of ML and DL to capture both short-term price trends and broader, more complex patterns. In markets with frequent shifts in renewable generation and load demand, multi-stage hybrid models tailored to each market's characteristics could improve forecast accuracy and reliability by adapting quickly to changing conditions.

Advances in probabilistic forecasting also present new ways to address uncertainty in price predictions. CP has shown particular promise in generating reliable PIs by adjusting to various levels of uncertainty and providing coverage guarantees even in highly stochastic environments. Future research could expand CP with ensemble approaches and multi-step-ahead capabilities to improve accuracy over extended horizons. A comprehensive evaluation comparing CP with Bayesian Neural Network (BNN), bootstrapping, and ensemble probabilistic methods could enhance uncertainty estimation, providing a versatile toolkit for resilient forecasts that can quantify risk effectively, thus supporting improved decision-making for traders, system operators, and grid managers.

As forecasting models become increasingly complex, improving their explainability is essential for building trust, fostering industry-wide adoption, and ensuring compliance with regulatory standards. Techniques such as SHapley Additive exPlanations [TSU+24] and Local Interpretable Model-Agnostic Explanations [BDS+24] provide valuable insights into model behaviour by identifying how input features impact predictions, helping to uncover biases or hidden correlations that may affect pricing. In regulated markets like the energy sector,

where transparency and accountability are critical, improved interpretability facilitates better informed, transparent decisions, supporting the adoption of sophisticated forecasting tools [SCEZ23], fostering a stable and reliable foundation for market participants.

In summary, by pursuing these future directions, further research can build on and extend the forecasting and trading frameworks introduced in this dissertation. Improvements in granularity, scalability, adaptability, and explainability will enable electricity markets to better account for the complexities introduced by renewable energy integration. These advancements promise to support greater market stability, efficiency, and profitability in the transition to a sustainable energy future. As the market evolves, these contributions will equip operators, traders, and policymakers with resilient, data-driven strategies essential for managing the challenges of high-renewable, volatile energy systems.

Acronyms

ACF Autocorrelation Function. 71, 73–76

ADF Augmented Dickey-Fuller. 71–73, 91

APS Aggregate Pinball Score. v, 148, 150, 152–154

AR Auto-Regressive. 19, 20, 22, 23, 35, 48, 75, 76, 97, 98, 140

ARIMA Auto-Regressive Integrated Moving Average. 5, 18–23, 26, 34, 35, 58, 97, 102, 103, 105, 111, 112, 116

ARIMAX Auto-Regressive Integrated Moving Average with Exogenous input. 21

ARMA Auto-Regressive Moving Average. 48

ARMAX Auto-Regressive Moving Average with Exogenous input. 49

ARX Auto-Regressive with Exogenous input. 23, 50

BESS Battery Energy Storage Systems. 2, 3, 6, 8, 113–115, 118, 122, 123, 130, 133–136, 156, 161

BiLSTM Bi-directional Long Short-Term Memory. 29, 35, 36

BM Balancing Market. iii–v, x, xi, 2–9, 11, 14–18, 21, 27–29, 31, 36, 48, 49, 53, 55, 57–60, 65–74, 76, 78, 80, 81, 83–101, 103–114, 116–119, 122–125, 127–131, 135–138, 140, 148–150, 153–165

BNN Bayesian Neural Network. 47, 48, 165

CNN Convolutional Neural Networks. 30, 32–36

CP Conformal Prediction. xi, 4, 6, 8, 10, 38, 42–46, 49, 52, 58, 137–140, 142–144, 146, 149–159, 162, 165

CQR Conformalised Quantile Regression. 42–45

DAM Day-Ahead Market. iii, iv, x, xi, 2, 3, 5–9, 11–18, 21–24, 27–29, 31–36, 42, 44, 45, 48–50, 53, 55–64, 66, 68–76, 78, 80–87, 89, 91–95, 97, 99, 101, 103, 104, 106, 108, 111–114, 116–119, 121–125, 127–131, 134–140, 148–151, 153–158, 160–164

DL Deep Learning. x, 5, 7, 10, 12, 13, 15, 17, 22, 23, 28–31, 34–36, 58, 93, 95–97, 99, 103–108, 110, 111, 117, 161, 165

DM Diebold-Mariano. 101, 105

DNN Deep Neural Network. 10, 28, 30–33, 35, 43, 48, 49, 58, 97–99, 108, 112, 117, 125, 129, 139

DP Dynamic Programming. 56

EFB Exclusive Feature Bundling. 29

EnbPI Ensemble Batch Prediction Intervals. xi, 4, 10, 38, 45–47, 52, 58, 137–140, 144–147, 150–154, 157, 158, 162

EPF Electricity Price Forecasting. 3, 4, 7–10, 17, 18, 21–36, 38, 44, 46–49, 52, 53, 57, 59, 60, 71, 76, 81, 93, 94, 97, 99, 105, 109, 137, 138, 140, 158, 162

ES Exponential Smoothing. 23

ESS Energy Storage Systems. 54, 56–58, 93, 94, 111, 112

FARX Full Auto-Regressive with Exogenous input. 23, 35

GARCH Generalised Auto-Regressive Conditional Heteroskedasticity. 21–24, 35, 42, 49, 58

GOSS Gradient-based One-Side Sampling. 29

GRU Gated Recurrent Units. 30, 32–35

I-SEM Irish Single Electricity Market. 13, 14, 60–62, 65, 66, 114, 131, 149, 158

IDM Intra-Day Market. x, 2, 3, 9, 11, 13, 14, 16–18, 33, 36, 45, 56, 94, 131, 138, 159, 163

KNN K-Nearest Neighbors. 24, 25, 98, 102, 103, 105, 111, 112, 117, 125, 129, 140, 150, 151, 157

KPSS Kwiatkowski-Phillips-Schmidt-Shin. 71–73, 91

LASSO Least Absolute Shrinkage and Selection Operator. 18, 21, 22, 24, 35, 43, 76, 79, 97, 98, 140

LEAR Lasso Estimated AutoRegressive. x, 5, 10, 18, 19, 21, 22, 27, 35, 97, 98, 101–106, 108, 109, 111, 112, 116, 139, 140, 150, 151, 157, 161

LGBM Light Gradient Boosting Machines. v, 24, 28, 29, 116, 117, 125, 127, 129, 131, 136, 140, 150–154, 156, 157

LM Lagrange Multiplier. 79, 80

LOESS Locally Estimated Scatterplot Smoothing. 88

LP Linear Programming. 10, 53–55, 57, 58

LSTM Long Short-Term Memory. 21, 30–36, 42, 49, 58, 95, 99, 117

MA Moving Average. 19, 20, 97
MAE Mean Absolute Error. v, 93, 99–106
MAPE Mean Absolute Percentage Error. 95, 100
MH-DNN Multi-Headed Deep Neural Network. 97
MH-RNN-DNN Multi-Headed Recurrent Neural Network Deep Neural Network. iii, 99, 100, 102, 104, 105, 117
MILP Mixed Integer Linear Programming. 10, 53, 55–58
ML Machine Learning. x, 5–8, 10, 12, 13, 15, 17, 21, 23, 24, 26, 29–31, 33–36, 39, 50, 53, 55, 58, 81, 91, 93, 95–97, 103, 105, 110, 111, 114, 160, 161, 165

NN Neural Network. 26, 30, 31, 33–35, 39, 44, 48, 49

PACF Partial Autocorrelation Functions. 71, 73, 75, 76
PCA Principal Component Analysis. 25, 33, 42
PDF Probability Distribution Function. 49
PEPF Probabilistic Electricity Price Forecasting. 4, 6, 8, 34, 38, 42–50, 52, 113, 115, 137, 139, 140, 146, 158, 162
PF Perfect Foresight. 124, 125, 127, 137
PI Prediction Interval. 4, 38, 42–50, 52, 137, 141–145, 148, 149
PIs Prediction Intervals. 4, 6–8, 139, 142, 144–149, 151, 158, 162, 165

QR Quantile Regression. xi, 4, 6, 8, 10, 38, 40–45, 47–49, 58, 81, 116, 117, 138–142, 146, 147, 149, 151–158, 162
QRA Quantile Regression Averaging. xi, 4, 6, 8, 10, 38, 42–44, 49, 50, 58, 139, 140, 146, 151–158
QRF Quantile Regression Forests. 145

RBF Radial Basis Function. 26, 29, 35, 98
RF Random Forest. v, x, 10, 24–28, 36, 43, 58, 81, 97, 98, 102, 103, 105, 106, 117, 125, 129, 136, 140, 150–157
RL Reinforcement Learning. 56, 164
RMSE Root Mean Squared Error. 94, 99–101, 105
RNN Recurrent Neural Network. 23, 31, 33, 34, 48, 95, 99

SCP Split Conformal Prediction. 10, 38, 42, 44, 45, 52, 140, 143, 144, 146, 151, 154, 157, 158
SEMO Single Electricity Market Operator. 5, 7, 161
SH-DNN Single-Headed Deep Neural Network. iii, 97, 99–102, 104, 105, 108
sMAPE symmetric Mean Absolute Percentage Error. 94, 99–101, 105

SOM Self-Organising Maps. 26, 35

SPCI Sequential Predictive Conformal Inference. xi, 4, 10, 38, 45–47, 52, 58, 137–140, 145–147, 150–154, 156–158, 162

STL Seasonal and Trend. 34, 71, 88–91

SVM Support Vector Machines. 25, 28, 36

SVR Support Vector Regression. 23–26, 30, 35, 97, 98, 102, 103, 105, 116

TF Transfer Function. 23

TFT Temporal Fusion Transformer. 30, 34

TS1 Trading Strategy 1. iii, 118–120, 122, 124, 135, 138, 156–158

TS2 Trading Strategy 2. iii, 118–120, 122–124, 135

TS3 Trading Strategy 3. iii–v, 118, 120–131, 135, 138, 139, 156–158

TSO Transmission System Operator. 14, 15, 62, 63, 65, 68, 69

VIF Variance Inflation Factor. 71, 81–83

XGB Extreme Gradient Boosting. x, 10, 24, 25, 27–29, 31, 34, 36, 58, 81, 97, 98, 102–106, 109–112, 116

Bibliography

[AAK+22] Ali Najem Alkawaz, Abdallah Abdellatif, Jeevan Kanesan, Anis Salwa Mohd Khairuddin, and Hassan Muwafaq Gheni. Day-ahead electricity price forecasting based on hybrid regression model. *IEEE Access*, 10:108021–108033, 2022.

[AAN+23] Md Shadman Abid, Hasan Jamil Apon, Imtiaz Mahmud Nafi, Ashik Ahmed, and Razzaqul Ahshan. Multi-objective architecture for strategic integration of distributed energy resources and battery storage system in microgrids. *Journal of Energy Storage*, 72:108276, 2023.

[AB21] Ekaterina Abramova and Derek Bunn. Optimal daily trading of battery operations using arbitrage spreads. *Energies*, 14(16):4931, 2021.

[ACT24] Anastasios Angelopoulos, Emmanuel Candes, and Ryan J Tibshirani. Conformal pid control for time series prediction. *Advances in neural information processing systems*, 36, 2024.

[AE19] Sara Atef and Amr B Eltawil. A comparative study using deep learning and support vector regression for electricity price forecasting in smart grids. In *2019 IEEE 6th international conference on industrial engineering and applications (ICIEA)*, pages 603–607. IEEE, 2019.

[AFRB17] José R Andrade, Jorge Filipe, Marisa Reis, and Ricardo J Bessa. Probabilistic price forecasting for day-ahead and intraday markets: Beyond the statistical model. *Sustainability*, 9(11):1990, 2017.

[AGMRJS11] Andrés M Alonso, Carolina García-Martos, Julio Rodríguez, and María Jesús Sánchez. Seasonal dynamic factor analysis and bootstrap inference: application to electricity market forecasting. *Technometrics*, 53(2):137–151, 2011.

[AHZ+24] Saud Alghumayjan, Jiajun Han, Ningkun Zheng, Ming Yi,

and Bolun Xu. Energy storage arbitrage in two-settlement markets: A transformer-based approach. *arXiv preprint arXiv:2404.17683*, 2024.

[AJ19] Tehreem Ashfaq and Nadeem Javaid. Short-term electricity load and price forecasting using enhanced KNN. In *2019 International Conference on Frontiers of Information Technology (FIT)*, pages 266–2665. IEEE, 2019.

[AKM+19] Mughees Ali, Zahoor Ali Khan, Sana Mujeeb, Shahid Abbas, and Nadeem Javaid. Short-term electricity price and load forecasting using enhanced support vector machine and k-nearest neighbor. In *2019 Sixth HCT Information Technology Trends (ITT)*, pages 79–83. IEEE, 2019.

[Alb18] Abdullah Albaker. *Optimal Operation of Integrated Microgrids*. PhD thesis, University of Denver, 2018.

[APS22] Mario Arrieta-Prieto and Kristen R Schell. Spatio-temporal probabilistic forecasting of wind power for multiple farms: A copula-based hybrid model. *International Journal of Forecasting*, 38(1):300–320, 2022.

[BA24] Dane Birkeland and Tarek AlSkaif. Research areas and methods of interest in european intraday electricity market research—a systematic literature review. *Sustainable Energy, Grids and Networks*, page 101368, 2024.

[BB] Richard Baxter and Ray Byrne. Doe eshb chapter 25: Energy storage system pricing.

[BBC+20] Mokhtar Bozorg, Antonio Bracale, Pierluigi Caramia, Guido Carpinelli, Mauro Carpita, and Pasquale De Falco. Bayesian bootstrap quantile regression for probabilistic photovoltaic power forecasting. *Protection and Control of Modern Power Systems*, 5(3):1–12, 2020.

[BBGL24] Alessandro Brusaferri, Andrea Ballarino, Luigi Grossi, and Fabrizio Laurini. On-line conformalized neural networks ensembles for probabilistic forecasting of day-ahead electricity prices. *arXiv preprint arXiv:2404.02722*, 2024.

[BCB14] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

[BCI+22] Sergio Beltrán, Alain Castro, Ion Irizar, Gorka Naveran, and

Imanol Yeregui. Framework for collaborative intelligence in forecasting day-ahead electricity price. *Applied Energy*, 306:118049, 2022.

[BCT18] Kyle Bassett, Rupp Carriveau, and David S-K Ting. Energy arbitrage and market opportunities for energy storage facilities in ontario. *Journal of Energy Storage*, 20:478–484, 2018.

[BD18] V Naren Bharatwaj and Aodhagan Downey. Real-time imbalance pricing in i-sem-ireland's balancing market. In *2018 20th National Power Systems Conference (NPSC)*, pages 1–6. IEEE, 2018.

[BDS+24] Lukas Baur, Konstantin Ditschuneit, Maximilian Schambach, Can Kaymakci, Thomas Wollmann, and Alexander Sauer. Explainability and interpretability in electric load forecasting using machine learning techniques–a review. *Energy and AI*, page 100358, 2024.

[BGDGR23] Anna Gloria Billé, Angelica Gianfreda, Filippo Del Grosso, and Francesco Ravazzolo. Forecasting electricity prices with expert, linear, and nonlinear models. *International Journal of Forecasting*, 39(2):570–586, 2023.

[BGDJB19] Tom Brijs, Frederik Geth, Cedric De Jonghe, and Ronnie Belmans. Quantifying electricity storage arbitrage opportunities in short-term electricity markets in the cwe region. *Journal of Energy Storage*, 25:100899, 2019.

[BGK18] Derek W Bunn, Angelica Gianfreda, and Stefan Kermer. A trading-based evaluation of density forecasts in a real-time electricity market. *Energies*, 11(10):2658, 2018.

[BHTL22] Tasarruf Bashir, Chen Haoyong, Muhammad Faizan Tahir, and Zhu Liqiang. Short term electricity load forecasting using hybrid prophet-lstm model optimized by bpnn. *Energy reports*, 8:1678–1686, 2022.

[BKM18a] Rita Beigaite, Tomas Krilavičius, and Ka Lok Man. Electricity price forecasting for nord pool data. In *2018 International Conference on Platform Technology and Service (PlatCon)*, pages 1–6, 2018.

[BKM18b] Rita Beigaite, Tomas Krilavičius, and Ka Lok Man. Electricity price forecasting for nord pool data. In *2018 International Conference on Platform Technology and Service (PlatCon)*, pages 1–6, 2018.

[BKW05] David A Belsley, Edwin Kuh, and Roy E Welsch. *Regression diagnostics: Identifying influential data and sources of collinearity*. John Wiley & Sons, 2005.

[BLZ16] Asmae Berrada, Khalid Loudiyi, and Izeddine Zorkani. Valuation of energy storage in energy and regulation markets. *Energy*, 115:1109–1118, 2016.

[BMPV19] Alessandro Brusaferri, Matteo Matteucci, Pietro Portolani, and Andrea Vitali. Bayesian deep learning based method for probabilistic forecast of day-ahead electricity prices. *Applied Energy*, 250:1158–1175, 2019.

[BMR+20] Alessandro Brusaferri, Matteo Matteucci, Danial Ramin, Stefano Spinelli, and Andrea Vitali. Probabilistic day-ahead energy price forecast by a mixture density recurrent neural network. In *2020 7th International Conference on Control, Decision and Information Technologies (CoDIT)*, volume 1, pages 523–528. IEEE, 2020.

[BMVT21] Kushagra Bhatia, Rajat Mittal, Jyothi Varanasi, and MM Tripathi. An ensemble approach for electricity price forecasting in markets with renewable energy resources. *Utilities Policy*, 70:101185, 2021.

[BO24] Adela Bâra and Simona-Vasilica Oprea. Predicting day-ahead electricity market prices through the integration of macroeconomic factors and machine learning techniques. *International Journal of Computational Intelligence Systems*, 17(1):10, 2024.

[BP79] Trevor S Breusch and Adrian R Pagan. A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the econometric society*, pages 1287–1294, 1979.

[BPPE14] Kyle Bradbury, Lincoln Pratson, and Dalia Patiño-Echeverri. Economic viability of energy storage systems based on price arbitrage potential in real-time us electricity markets. *Applied Energy*, 114:512–519, 2014.

[Bre78] Trevor S Breusch. Testing for autocorrelation in dynamic linear models. *Australian economic papers*, 17(31), 1978.

[BRM16] Antonio Bello, Javier Reneses, and Antonio Muñoz. Medium-term probabilistic forecasting of extremely low prices in electricity markets: Application to the spanish case. *Energies*, 9(3):193, 2016.

[BRMD16] Antonio Bello, Javier Reneses, Antonio Muñoz, and Andrés Delgadillo. Probabilistic forecasting of hourly electricity prices in the medium-term using spatial interpolation techniques. *International Journal of Forecasting*, 32(3):966–980, 2016.

[Bro19] Jason Brownlee. Comparing classical and machine learning algorithms for time series forecasting. *Machine Learning Mastery, Australia*, 2019.

[BS20] Lukas Börjesson and Martin Singull. Forecasting financial time series through causal and dilated convolutional neural networks. *Entropy*, 22(10):1094, 2020.

[BY24] Çağatay Berke Bozlak and Claudia Fernanda Yaşar. An optimized deep learning approach for forecasting day-ahead electricity prices. *Electric Power Systems Research*, 229:110129, 2024.

[CBL+23] Thibault Cordier, Vincent Blot, Louis Lacombe, Thomas Morzadec, Arnaud Capitaine, and Nicolas Brunel. Flexible and systematic uncertainty estimation with conformal prediction via the mapie library. In *Conformal and Probabilistic Prediction with Applications*, pages 549–581. PMLR, 2023.

[CCL+17] Igor M Coelho, Vitor N Coelho, Eduardo J da S Luz, Luiz S Ochi, Frederico G Guimaraes, and Eyder Rios. A gpu deep learning metaheuristic based model for time series forecasting. *Applied Energy*, 201:412–418, 2017.

[CDCG22] Tingting Cai, Mingyu Dong, Ke Chen, and Taorong Gong. Methods of participating power spot market bidding and settlement for renewable energy systems. *Energy Reports*, 8:7764–7772, 2022.

[CDZD19] Hangyang Cheng, Xiangwu Ding, Wuneng Zhou, and Renqiang Ding. A hybrid electricity price forecasting model with bayesian optimization for german energy exchange. *International Journal of Electrical Power & Energy Systems*, 110:653–666, 2019.

[CENC03] Javier Contreras, Rosario Espinola, Francisco J Nogales, and Antonio J Conejo. ARIMA models to predict next-day electricity prices. *IEEE transactions on power systems*, 18(3):1014–1020, 2003.

[CG16] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international*

conference on knowledge discovery and data mining, pages 785–794, 2016.

[CHKO04] Jesús Crespo Cuaresma, Jaroslava Hlouskova, Stephan Kossmeier, and Michael Obersteiner. Forecasting electricity spot-prices using linear univariate time-series models. *Applied Energy*, 77(1):87–106, 2004.

[Cho21] Francois Chollet. *Deep learning with Python*. Simon and Schuster, 2021.

[CK07] Michael P Clements and Jae H Kim. Bootstrap prediction intervals for autoregressive time series. *Computational statistics & data analysis*, 51(7):3580–3594, 2007.

[CLF+11] David Connolly, Henrik Lund, Paddy Finn, Brian Vad Mathiesen, and Martin Leahy. Practical operation strategies for pumped hydroelectric energy storage (phes) utilising electricity price arbitrage. *Energy Policy*, 39(7):4189–4196, 2011.

[CLMCC+22] Sergio Cantillo-Luna, Ricardo Moreno-Chuquen, Harold R Chamorro, Jose Miguel Riquelme-Dominguez, and Francisco Gonzalez-Longatt. Locational marginal price forecasting using svr-based multi-output regression in electricity markets. *Energies*, 15(1):293, 2022.

[CMS22] Luis Corona, Asuncion Mochon, and Yago Saez. Electricity market integration and impact of renewable energy sources in the central western europe region: Evolution since the implementation of the flow-based market coupling mechanism. *Energy Reports*, 8:1768–1788, 2022.

[CMTP+24] Javier Cardo-Miota, Rohit Trivedi, Sandipan Patra, Shafi Khadem, and Mohamed Bahloul. Data-driven approach for day-ahead system non-synchronous penetration forecasting: A comprehensive framework, model development and analysis. *Applied Energy*, 362:123006, 2024.

[Com21] SEM Committee. Sem plexos model validation 2021-2029 and backcast report, Dec 2021.

[COS13] Peter Cramton, Axel Ockenfels, and Steven Stoft. Capacity market fundamentals. *Economics of Energy & Environmental Policy*, 2(2):27–46, 2013.

[CPEM05] Antonio J Conejo, Miguel A Plazas, Rosa Espinola, and Ana B Molina. Day-ahead electricity price forecasting using the

wavelet transform and ARIMA models. *IEEE transactions on power systems*, 20(2):1035–1042, 2005.

[CPH+18] Radhakrishnan Angamuthu Chinnathambi, Siby Jose Plathottam, Tareq Hossen, Arun Sukumaran Nair, and Prakash Ranganathan. Deep neural networks for day-ahead electricity price markets. In *2018 IEEE electrical power and energy conference (EPEC)*, pages 1–6. IEEE, 2018.

[CS20] Pornchai Chaweewat and Jai Govind Singh. An electricity price interval forecasting by using residual neural network. *International Transactions on Electrical Energy Systems*, 30(9):e12506, 2020.

[CSMZ21] Armin Cosic, Michael Stadler, Muhammad Mansoor, and Michael Zellinger. Mixed-integer linear programming based optimization strategies for renewable energy communities. *Energy*, 237:121559, 2021.

[CW10] Jinxing Che and Jianzhou Wang. Short-term electricity prices forecasting based on support vector regression and autoregressive integrated moving average modeling. *Energy Conversion and Management*, 51(10):1911–1917, 2010.

[CWMJ19] Yiyuan Chen, Yufeng Wang, Jianhua Ma, and Qun Jin. Brim: An accurate electricity spot price prediction scheme-based bidirectional recurrent neural network and integrated market. *Energies*, 12(12):2241, 2019.

[CXW+20] Ren Cai, Sen Xie, Bozhong Wang, Ruijiang Yang, Daosen Xu, and Yang He. Wind speed forecasting based on extreme gradient boosting. *IEEE Access*, 8:175063–175069, 2020.

[CZC18] Zihan Chang, Yang Zhang, and Wenbo Chen. Effective adam-optimized LSTM neural network for electricity price forecasting. In *2018 IEEE 9th international conference on software engineering and service science (ICSESS)*, pages 245–248. IEEE, 2018.

[CZC19] Zihan Chang, Yang Zhang, and Wenbo Chen. Electricity price prediction based on hybrid model of adam optimized LSTM neural network and wavelet transform. *Energy*, 187:115804, 2019.

[DBdV+19] Jonathan Dumas, Ioannis Boukas, Miguel Manuel de Villena, Sébastien Mathieu, and Bertrand Cornélusse. Probabilistic forecasting of imbalance prices in the belgian context. In *2019 16th International Conference on the European Energy Market (EEM)*,

pages 1–7. IEEE, 2019.

[DBK+96] Harris Drucker, Christopher J Burges, Linda Kaufman, Alex Smola, and Vladimir Vapnik. Support vector regression machines. *Advances in neural information processing systems*, 9, 1996.

[DBW23] Nicolas Dewolf, Bernard De Baets, and Willem Waegeman. Valid prediction intervals for regression problems. *Artificial Intelligence Review*, 56(1):577–613, 2023.

[DF81] David A Dickey and Wayne A Fuller. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica: journal of the Econometric Society*, pages 1057–1072, 1981.

[DGRR22] Fabrizio Durante, Angelica Gianfreda, Francesco Ravazzolo, and Luca Rossini. A multivariate dependence analysis for electricity prices, demand and renewable energy sources. *Information Sciences*, 590:74–89, 2022.

[DGV09] Abdou Kâ Diongue, Dominique Guegan, and Bertrand Vignal. Forecasting electricity spot market prices with a k-factor GIGARCH process. *Applied energy*, 86(4):505–510, 2009.

[DKD09] Armen Der Kiureghian and Ove Ditlevsen. Aleatory or epistemic? does it matter? *Structural safety*, 31(2):105–112, 2009.

[DLM19] Linh Phuong Catherine Do, Štefan Lyócsa, and Peter Molnár. Impact of wind and solar production on electricity prices: Quantile regression approach. *Journal of the Operational Research Society*, 70(10):1752–1768, 2019.

[DlTRM+24] Jorge De la Torre, Leticia R Rodriguez, Francisco EL Monteagudo, Leonel R Arredondo, and José B Enriquez. Electricity price forecast in wholesale markets using conformal prediction: Case study in mexico. *Energy Science & Engineering*, 12(3):524–540, 2024.

[DM02] Francis X Diebold and Robert S Mariano. Comparing predictive accuracy. *Journal of Business & economic statistics*, 20(1):134–144, 2002.

[Dud16] Grzegorz Dudek. Multilayer perceptron for gefcom2014 probabilistic electricity price forecasting. *International Journal of Forecasting*, 32(3):1057–1060, 2016.

[Dud18] Grzegorz Dudek. Probabilistic forecasting of electricity prices using kernel regression. In *2018 15th International Conference*

on the European Energy Market (EEM), pages 1–5. IEEE, 2018.

[DZFG24] Grégoire Dutot, Margaux Zaffran, Olivier Féron, and Yannig Goude. Adaptive probabilistic forecasting of french electricity spot prices. *arXiv preprint arXiv:2405.15359*, 2024.

[DZP+19] Daizong Ding, Mi Zhang, Xudong Pan, Min Yang, and Xiangnan He. Modeling extreme events in time series prediction. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1114–1122, 2019.

[EC10] Jim Eyer and Garth Corey. Energy storage for the electricity grid: Benefits and market potential assessment guide. *Sandia National Laboratories*, 20(10):5, 2010.

[Eir22] EirGrid. Renewable energy, 2022. System and Renewable Data Reports: Data relating to system and renewable energy use in our electricity system.

[Eir23] EirGrid. Eirgrid group, 2023. Data download summary report.

[ERH21] Anselm Eicke, Oliver Ruhnau, and Lion Hirth. Electricity balancing as a market equilibrium. 2021.

[EUP24] EUPHEMIA. Euphemia public description: PCR market coupling algorithm, 2024.

[Eur24] Eurostat. Statistics explained: Renewable energy statistics, 2024. Share of renewable energy more than doubled between 2004 and 2020.

[FA23] Ann-Katrin Fleck and Vasilios Anatolitis. Achieving the objectives of renewable energy policy–insights from renewable energy auction design in europe. *Energy Policy*, 173:113357, 2023.

[FBCRT22] Rina Foygel Barber, Emmanuel J Candes, Aaditya Ramdas, and Ryan J Tibshirani. Conformal prediction beyond exchangeability. *arXiv e-prints*, pages arXiv–2202, 2022.

[Fér24] Olivier Féron. Probabilistic forecasting of electricity spot prices, 2024. Technical report, Institut Louis Bachelier.

[FMC07] Shu Fan, Chengziong Mao, and Luonan Chen. Next-day electricity-price forecasting using a hybrid network. *IET generation, transmission & distribution*, 1(1):176–182, 2007.

[FMC+23] Roberto Dias Filho, Amanda CM Monteiro, Tatiane Costa, Andrea Vasconcelos, Ana Clara Rode, and Manoel Marinho. Strategic guidelines for battery energy storage system deployment: Regulatory framework, incentives, and market planning. *Ener-*

gies, 16(21):7272, 2023.

[FSD16] Felipe Feijoo, Walter Silva, and Tapas K Das. A computationally efficient electricity price forecasting model for real time energy markets. *Energy Conversion and Management*, 113:27–35, 2016.

[GAG+19] M Ghayekhloo, R Azimi, M Ghofrani, MB Menhaj, and E Shekari. A combination approach based on a novel data clustering method and bayesian recurrent neural network for day-ahead price forecasting of electricity markets. *Electric Power Systems Research*, 168:184–199, 2019.

[GBC16] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

[GDCPSS24] Sujan Ghimire, Ravinesh C Deo, David Casillas-Pérez, and Sancho Salcedo-Sanz. Two-step deep learning framework with error compensation technique for short-term, half-hourly electricity price forecasting. *Applied Energy*, 353:122059, 2024.

[GFF+22] Arpad Gellert, Ugo Fiore, Adrian Florea, Radu Chis, and Francesco Palmieri. Forecasting electricity consumption and production in smart homes through statistical methods. *Sustainable Cities and Society*, 76:103426, 2022.

[GGBGDM08] Javier García González, Julián Barquín Gil, and Pablo Dueñas Martínez. A hybrid approach for modeling electricity price series in the medium term. 2008.

[Gir16] Godekere P Girish. Spot electricity price forecasting in indian electricity market using autoregressive-GARCH models. *Energy Strategy Reviews*, 11:52–57, 2016.

[GMI+22] David Gräf, Julian Marschewski, Lukas Ibing, David Huckebrink, Marc Fiebrandt, Götz Hanau, and Valentin Bertsch. What drives capacity degradation in utility-scale battery energy storage systems? the impact of operating strategy and temperature in different grid applications. *Journal of Energy Storage*, 47:103533, 2022.

[GMM21] ILR Gomes, Rui Melicio, and VMF Mendes. A novel microgrid support management system based on stochastic mixed-integer linear programming. *Energy*, 223:120030, 2021.

[GMO16] Camino González, José M Mira, and José A Ojeda. Applying multi-output random forest models to electricity price forecast. 2016.

[God78] Leslie G Godfrey. Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. *Econometrica: Journal of the Econometric Society*, pages 1293–1301, 1978.

[GSB+23] Subhankar Ghosh, Yuanjie Shi, Taha Belkhouja, Yan Yan, Jana Doppa, and Brian Jones. Probabilistically robust conformal prediction. *Proceedings of the 39th Conference on Uncertainty in Artificial Intelligence (UAI)*, 216:681–690, 31 Jul–04 Aug 2023.

[GUO23] Salih Gunduz, Umut Ugurlu, and Ilkay Oksuz. Transfer learning for electricity price forecasting. *Sustainable Energy, Grids and Networks*, 34:100996, 2023.

[Gür18] Turgut M Gür. Review of electrical energy storage technologies, materials and systems: challenges and prospects for large-scale grid storage. *Energy & Environmental Science*, 11(10):2696–2767, 2018.

[GVV98] A Gammerman, V Vovk, and V Vapnik. Learning by transduction. *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence (UAI)*, 1998.

[GWA23] William Gomez, Fu-Kwun Wang, and Zemenu Endalamaw Amogne. Electricity load and price forecasting using a hybrid method based bidirectional long short-term memory with attention mechanism model. *International Journal of Energy Research*, 2023(1):3815063, 2023.

[GWBS22] Paolo Gabrielli, Moritz Wüthrich, Steffen Blume, and Giovanni Sansavini. Data-driven modeling for long-term electricity price forecasting. *Energy*, 244:123107, 2022.

[HA18] Rob J Hyndman and George Athanasopoulos. *Forecasting: principles and practice*. OTexts, 2018.

[Hac06] Ian Hacking. *The emergence of probability: A philosophical study of early ideas about probability, induction and statistical inference*. Cambridge University Press, 2006.

[HBK+16] Lars Ivar Hagfors, Derek Bunn, Eline Kristoffersen, Tiril Toftdahl Staver, and Sjur Westgaard. Modeling the uk electricity price distributions using quantile regression. *Energy*, 102:231–243, 2016.

[HDY+12] Geoffrey Hinton, Li Deng, Dong Yu, George E Dahl, Abdelrahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Van-

houcke, Patrick Nguyen, Tara N Sainath, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal processing magazine*, 29(6):82–97, 2012.

[Her18] Bruce Herzer. What Does Battery Storage Cost? - Invinity. `https://invinity.com/what-does-battery-storage-cost/`, 2018.

[HHF23] Mohammadreza Heidarpanah, Farhad Hooshyaripor, and Meysam Fazeli. Daily electricity price forecasting using artificial intelligence models in the iranian electricity market. *Energy*, 263:126011, 2023.

[HLT+22] Jianming Hu, Qingxi Luo, Jingwei Tang, Jiani Heng, and Yuwen Deng. Conformalized temporal convolutional quantile regression networks for wind power interval forecasting. *Energy*, 248:123497, 2022.

[Hoc97] S Hochreiter. Long short-term memory. *Neural Computation MIT-Press*, 1997.

[Hor96] Stephen C Hora. Aleatory and epistemic uncertainty in probability elicitation with an example from hazardous waste management. *Reliability Engineering & System Safety*, 54(2-3):217–223, 1996.

[HOT06] Geoffrey E. Hinton, Simon Osindero, and Yee-Whye Teh. A Fast Learning Algorithm for Deep Belief Nets. *Neural Computation*, 18(7):1527–1554, 07 2006.

[HSW+24] Siwan Huang, Jianheng Shi, Baoyue Wang, Na An, Li Li, Xuebing Hou, Chunsen Wang, Xiandong Zhang, Kai Wang, Huilin Li, et al. A hybrid framework for day-ahead electricity spot-price forecasting: A case study in china. *Applied Energy*, 373:123863, 2024.

[HW21] Eyke Hüllermeier and Willem Waegeman. Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning*, 110:457–506, 2021.

[IB+18] Valentin Ilea, Cristian Bovo, et al. European day-ahead electricity market coupling: Discussion, modeling, and case study. *Electric Power Systems Research*, 155:80–92, 2018.

[IMK+22] Rashed Iqbal, Hazlie Mokhlis, Anis Salwa Mohd Khairuddin, Syafiqah Ismail, and Munir Azam Muhammad. Optimized gated

recurrent unit for mid-term electricity price forecasting. *Computer Systems Science & Engineering*, 43(2), 2022.

[IOH22] Izzaamirah Ishak, Nor Salwati Othman, and Nor Hamisham Harun. Forecasting electricity consumption of malaysia's residential sector: Evidence from an exponential smoothing model. *F1000Research*, 11:54, 2022.

[Jan24] Joanna Janczura. Expectile regression averaging method for probabilistic forecasting of electricity prices. *arXiv preprint arXiv:2402.07559*, 2024.

[JAS11] Tina Jakaša, Ivan Andročec, and Petar Sprčić. Electricity price forecasting—ARIMA model approach. In *2011 8th international conference on the European energy market (EEM)*, pages 222–225. IEEE, 2011.

[JAVWVH24] Jef Jonkers, Diego Nieves Avendano, Glenn Van Wallendael, and Sofie Van Hoecke. A novel day-ahead regional and probabilistic wind power forecasting framework using deep cnns and conformalized regression forests. *Applied Energy*, 361:122900, 2024.

[JBA22] Vilde Jensen, Filippo Maria Bianchi, and Stian Normann Anfinsen. Ensemble conformalized quantile regression for probabilistic time series forecasting. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[JH18] LianLian Jiang and Guoqiang Hu. Day-ahead price forecasting for electricity market using long-short term memory recurrent neural network. In *2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, pages 949–954. IEEE, 2018.

[JKG19] Nils Jakob Johannesen, Mohan Kolhe, and Morten Goodwin. Deregulated electric energy price forecasting in nordpool market using regression techniques. In *2019 IEEE Sustainable Power and Energy Conference (iSPEC)*, pages 1932–1938. IEEE, 2019.

[JLMW22] Arkadiusz Jędrzejewski, Jesus Lago, Grzegorz Marcjasz, and Rafał Weron. Electricity price forecasting: The dawn of machine learning. *IEEE Power and Energy Magazine*, 20(3):24–31, 2022.

[JP15] Daniel R Jiang and Warren B Powell. Optimal hour-ahead bidding in the real-time electricity market with battery storage using approximate dynamic programming. *INFORMS Journal on*

Computing, 27(3):525–543, 2015.

[JPDW24] He Jiang, Sheng Pan, Yao Dong, and Jianzhou Wang. Probabilistic electricity price forecasting based on penalized temporal fusion transformer. *Journal of Forecasting*, 2024.

[JPMN14] Tryggvi Jónsson, Pierre Pinson, Henrik Madsen, and Henrik Aalborg Nielsen. Predictive densities for day-ahead electricity prices using time-adaptive quantile regression. *Energies*, 7(9):5523–5547, 2014.

[JW10] Joanna Janczura and Rafal Weron. An empirical comparison of alternate regime-switching models for electricity spot prices. *Energy economics*, 32(5):1059–1073, 2010.

[KAU21] Mahmut Kara, Kazim Baris Atici, and Aydin Ulucan. Price and volatility forecasting in electricity with support vector regression and random forest. *Applied Operations Research and Financial Modelling in Energy: Practical Applications and Implications*, pages 101–124, 2021.

[KBJ78] Roger Koenker and Gilbert Bassett Jr. Regression quantiles. *Econometrica: journal of the Econometric Society*, 1978.

[KEF15] Gro Klæboe, Anders Lund Eriksrud, and Stein-Erik Fleten. Benchmarking time series based forecasting models for electricity balancing market prices. *Energy Systems*, 6:43–61, 2015.

[KFA+20] Zahoor Ali Khan, Sahiba Fareed, Mubbashra Anwar, Afrah Naeem, Hira Gul, Arooj Arif, and Nadeem Javaid. Short term electricity price forecasting through convolutional neural network. In *Web, Artificial Intelligence and Network Applications: Proceedings of the Workshops of the 34th International Conference on Advanced Information Networking and Applications (WAINA-2020)*, pages 1181–1188. Springer, 2020.

[KG17] Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? *Advances in neural information processing systems*, 30, 2017.

[KGPA24] Alkiviadis Kitsatoglou, Giannis Georgopoulos, Panagiotis Papadopoulos, and Herodotus Antonopoulos. An ensemble approach for enhanced day-ahead price forecasting in electricity markets. *Expert Systems with Applications*, 256:124971, 2024.

[KH18] Ping-Huan Kuo and Chiou-Jye Huang. An electricity price forecasting model by hybrid structured deep neural networks. *Sus-*

tainability, 10(4):1280, 2018.

[KH19] Christopher Koch and Lion Hirth. Short-term electricity trading for system balancing: An empirical analysis of the role of intraday trading in balancing germany's electricity system. *Renewable and Sustainable Energy Reviews*, 113:109275, 2019.

[KK19] Maciej Kostrzewski and Jadwiga Kostrzewska. Probabilistic electricity price forecasting with bayesian stochastic volatility models. *Energy Economics*, 80:610–620, 2019.

[KL22] Hosna Khajeh and Hannu Laaksonen. Applications of probabilistic forecasting in smart grids: A review. *Applied Sciences*, 12(4):1823, 2022.

[KLF+21] Manuel Tong Koecklin, Genaro Longoria, Desta Z Fitiwi, Joseph F DeCarolis, and John Curtis. Public acceptance of renewable electricity generation and transmission network developments: Insights from ireland. *Energy Policy*, 151:112185, 2021.

[KNC13] Abbas Khosravi, Saeid Nahavandi, and Doug Creighton. Quantifying uncertainties of neural network-based electricity price forecasts. *Applied energy*, 112:120–129, 2013.

[KNT24] Deniz Kenan Kılıç, Peter Nielsen, and Amila Thibbotuwawa. Intraday electricity price forecasting via LSTM and trading strategy for the power market: A case study of the west denmark dk1 grid region. *Energies*, 17(12):2909, 2024.

[Kot22] Panagiotis-Christos Kotsias. Forecasting electricity prices for intraday markets with machine learning: An exploratory comparison of the state of the art. 2022.

[KPSS92] Denis Kwiatkowski, Peter CB Phillips, Peter Schmidt, and Yongcheol Shin. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics*, 54(1-3):159–178, 1992.

[KR05] Christopher R Knittel and Michael R Roberts. An empirical examination of restructured electricity prices. *Energy Economics*, 27(5):791–817, 2005.

[KRKB23] Emil Kraft, Marianna Russo, Dogan Keles, and Valentin Bertsch. Stochastic optimization of trading strategies in sequential electricity markets. *European Journal of Operational Research*,

308(1):400–421, 2023.

[KSH12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.

[KSN23] Nadja Klein, Michael Stanley Smith, and David J Nott. Deep distributional time series models and the probabilistic forecasting of intraday electricity prices. *Journal of Applied Econometrics*, 38(4):493–511, 2023.

[KUZ+17] Dheepak Krishnamurthy, Canan Uckun, Zhi Zhou, Prakash R Thimmapuram, and Audun Botterud. Energy storage arbitrage under day-ahead and real-time price uncertainty. *IEEE Transactions on Power Systems*, 33(1):84–93, 2017.

[KW23] Gaurav Kapoor and Nuttanan Wichitaksorn. Electricity price forecasting in new zealand: A comparative analysis of statistical and machine learning models with feature selection. *Applied Energy*, 347:121446, 2023.

[KX19] Orhan Altuğ Karabiber and George Xydis. Electricity price forecasting in the danish day-ahead market using the TBATS, ANN and ARIMA methods. *Energies*, 12(5):928, 2019.

[KZ21] Christopher Kath and Florian Ziel. Conformal prediction interval estimation and applications to day-ahead and intraday power markets. *International Journal of Forecasting*, 37(2):777–799, 2021.

[LB78] Greta M Ljung and George EP Box. On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303, 1978.

[LB21] Wei Li and Denis Mike Becker. Day-ahead electricity price prediction applying hybrid models of lstm-based deep learning methods and feature selection algorithms under consideration of market coupling. *Energy*, 237:121543, 2021.

[LDRDS18] 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.

[LDRVDS18] Jesus Lago, Fjo De Ridder, Peter Vrancx, and Bart De Schutter. Forecasting day-ahead electricity prices in europe: The importance of considering market integration. *Applied energy*, 211:890–903, 2018.

[LGT10] Whei-Min Lin, Hong-Jey Gow, and Ming-Tang Tsai. An enhanced radial basis function network for short-term electricity price forecasting. *Applied Energy*, 87(10):3226–3234, 2010.

[LMDSW21] Jesus Lago, Grzegorz Marcjasz, Bart De Schutter, and Rafał Weron. Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. *Applied Energy*, 293:116983, 2021.

[LPKC20] Alexandre Lucas, Konstantinos Pegios, Evangelos Kotsakis, and Dan Clarke. Price forecasting for the balancing energy market using machine-learning regression. *Energies*, 13(20):5420, 2020.

[LS22] Stefan Lamp and Mario Samano. Large-scale battery storage, short-term market outcomes, and arbitrage. *Energy Economics*, 107:105786, 2022.

[LS24] Francesco Lisi and Ismail Shah. Joint component estimation for electricity price forecasting using functional models. *Energies*, 17(14):3461, 2024.

[LSGSEC+22] Miguel López Santos, Xela García-Santiago, Fernando Echevarría Camarero, Gonzalo Blázquez Gil, and Pablo Carrasco Ortega. Application of temporal fusion transformer for day-ahead pv power forecasting. *Energies*, 15(14):5232, 2022.

[LSH22] Malte Lehna, Fabian Scheller, and Helmut Herwartz. Forecasting day-ahead electricity prices: A comparison of time series and neural network models taking external regressors into account. *Energy Economics*, 106:105742, 2022.

[LSTL23] Huixin Liu, Xiaodong Shen, Xisheng Tang, and Junyong Liu. Day-ahead electricity price probabilistic forecasting based on shap feature selection and lstnet quantile regression. *Energies*, 16(13):5152, 2023.

[LUW24] Arkadiusz Lipiecki, Bartosz Uniejewski, and Rafał Weron. Postprocessing of point predictions for probabilistic forecasting of day-ahead electricity prices: The benefits of using isotonic distributional regression. *Energy Economics*, page 107934, 2024.

[LW19] Shuman Luo and Yang Weng. A two-stage supervised learning approach for electricity price forecasting by leveraging different data sources. *Applied energy*, 242:1497–1512, 2019.

[LWB+19] Boqiang Lin, Wei Wu, Mengqi Bai, Chunping Xie, and Jonathan

Radcliffe. Liquid air energy storage: Price arbitrage operations and sizing optimization in the gb real-time electricity market. *Energy Economics*, 78:647–655, 2019.

[LXX24] Junghwan Lee, Chen Xu, and Yao Xie. Transformer conformal prediction for time series. *arXiv preprint arXiv:2406.05332*, 2024.

[MA21] Waleed Mahmood and Ercan Avşar. One step ahead prediction of ozone concentration for determination of outdoor air quality level. *MANAS Journal of Engineering*, 9(1):45–54, 2021.

[MABH16] Carlo Brancucci Martinez-Anido, Greg Brinkman, and Bri-Mathias Hodge. The impact of wind power on electricity prices. *Renewable Energy*, 94:474–487, 2016.

[Mac20] Katarzyna Maciejowska. Assessing the impact of renewable energy sources on the electricity price level and variability–a quantile regression approach. *Energy Economics*, 85:104532, 2020.

[Mak93] Spyros Makridakis. Accuracy measures: theoretical and practical concerns. *International journal of forecasting*, 9(4):527–529, 1993.

[Mal23] Seyed Matin Malakouti. Use machine learning algorithms to predict turbine power generation to replace renewable energy with fossil fuels. *Energy Exploration & Exploitation*, 41(2):836–857, 2023.

[MAM24] Mohammad Reza Monjazeb, Hossein Amiri, and Akram Movahedi. Wholesale electricity price forecasting by quantile regression and kalman filter method. *Energy*, 290:129925, 2024.

[Mee20] Leonardo Meeus. *The evolution of electricity markets in Europe*. Edward Elgar Publishing, 2020.

[MFS15] Dylan McConnell, Tim Forcey, and Mike Sandiford. Estimating the value of electricity storage in an energy-only wholesale market. *Applied Energy*, 159:422–432, 2015.

[MHH+14] Jie Mei, Dawei He, Ronald Harley, Thomas Habetler, and Guannan Qu. A random forest method for real-time price forecasting in new york electricity market. In *2014 IEEE PES general meeting| conference & exposition*, pages 1–5. IEEE, 2014.

[MJ19] Sana Mujeeb and Nadeem Javaid. Esaenarx and de-relm: Novel schemes for big data predictive analytics of electricity load and price. *Sustainable Cities and Society*, 51:101642, 2019.

[MJA+19] Sana Mujeeb, Nadeem Javaid, Mariam Akbar, Rabiya Khalid, Orooj Nazeer, and Mahnoor Khan. Big data analytics for price and load forecasting in smart grids. In *Advances on Broadband and Wireless Computing, Communication and Applications: Proceedings of the 13th International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA-2018)*, pages 77–87. Springer, 2019.

[MK21] Gholamreza Memarzadeh and Farshid Keynia. Short-term electricity load and price forecasting by a new optimal LSTM-NN based prediction algorithm. *Electric Power Systems Research*, 192:106995, 2021.

[MN24] Katarzyna Maciejowska and Weronika Nitka. Multiple split approach–multidimensional probabilistic forecasting of electricity markets. *arXiv preprint arXiv:2407.07795*, 2024.

[MNW16] Katarzyna Maciejowska, Jakub Nowotarski, and Rafał Weron. Probabilistic forecasting of electricity spot prices using factor quantile regression averaging. *International Journal of Forecasting*, 32(3):957–965, 2016.

[MNWZ23] Grzegorz Marcjasz, Michał Narajewski, Rafał Weron, and Florian Ziel. Distributional neural networks for electricity price forecasting. *Energy Economics*, 125:106843, 2023.

[MODJ23] Thomas Mercier, Mathieu Olivier, and Emmanuel De Jaeger. The value of electricity storage arbitrage on day-ahead markets across europe. *Energy Economics*, 123:106721, 2023.

[Mör19] Mathis Mörke. Marcos lópez de prado: Advances in financial machine learning: Wiley, 2018, 400 pp, usd 50.00, approx. eur 45, isbn: 978-1-119-48208-6, 2019.

[MP17] Nicoló Mazzi and Pierre Pinson. 10 - wind power in electricity markets and the value of forecasting. In George Kariniotakis, editor, *Renewable Energy Forecasting*, Woodhead Publishing Series in Energy, pages 259–278. Woodhead Publishing, 2017.

[MR22] Tensay Hadush Meles and Lisa Ryan. Adoption of renewable home heating systems: An agent-based model of heat pumps in ireland. *Renewable and Sustainable Energy Reviews*, 169:112853, 2022.

[MRRFJC16] Claudio Monteiro, Ignacio J Ramirez-Rosado, L Alfredo Fernandez-Jimenez, and Pedro Conde. Short-term price fore-

casting models based on artificial neural networks for intraday sessions in the iberian electricity market. *Energies*, 9(9):721, 2016.

[MRRFJR18] Claudio Monteiro, Ignacio J Ramirez-Rosado, L Alfredo Fernandez-Jimenez, and Miguel Ribeiro. New probabilistic price forecasting models: Application to the iberian electricity market. *International Journal of Electrical Power & Energy Systems*, 103:483–496, 2018.

[MS19] Gernot Müller and Armin Seibert. Bayesian estimation of stable carma spot models for electricity prices. *Energy Economics*, 78:267–277, 2019.

[MSU23] Katarzyna Maciejowska, Tomasz Serafin, and Bartosz Uniejewski. Probabilistic forecasting with factor quantile regression: Application to electricity trading. *arXiv preprint arXiv:2303.08565*, 2023.

[MSU24] Katarzyna Maciejowska, Tomasz Serafin, and Bartosz Uniejewski. Probabilistic forecasting with a hybrid factor-qra approach: Application to electricity trading. *Electric Power Systems Research*, 234:110541, 2024.

[MUW20a] Grzegorz Marcjasz, Bartosz Uniejewski, and Rafał Weron. Beating the naïve—combining LASSO with naïve intraday electricity price forecasts. *Energies*, 13(7):1667, 2020.

[MUW20b] Grzegorz Marcjasz, Bartosz Uniejewski, and Rafał Weron. Probabilistic electricity price forecasting with narx networks: Combine point or probabilistic forecasts? *International Journal of Forecasting*, 36(2):466–479, 2020.

[MUW22] Katarzyna Maciejowska, Bartosz Uniejewski, and Rafał Weron. Forecasting electricity prices. *arXiv preprint arXiv:2204.11735*, 2022.

[MWZ+22] Anbo Meng, Peng Wang, Guangsong Zhai, Cong Zeng, Shun Chen, Xiaoyi Yang, and Hao Yin. Electricity price forecasting with high penetration of renewable energy using attention-based LSTM network trained by crisscross optimization. *Energy*, 254:124212, 2022.

[Nar22] Michał Narajewski. Probabilistic forecasting of german electricity imbalance prices. *Energies*, 15(14):4976, 2022.

[NCCE02] Francisco Javier Nogales, Javier Contreras, Antonio J Conejo,

and Rosario Espínola. Forecasting next-day electricity prices by time series models. *IEEE Transactions on power systems*, 17(2):342–348, 2002.

[Ner71] Marc Nerlove. Time series analysis, forecasting, and control., 1971.

[NM24] Daniel Nickelsen and Gernot Müller. Bayesian hierarchical probabilistic forecasting of intraday electricity prices. *arXiv preprint arXiv:2403.05441*, 2024.

[NSSK23] Amril Nazir, Abdul Khalique Shaikh, Abdul Salam Shah, and Ashraf Khalil. Forecasting energy consumption demand of customers in smart grid using temporal fusion transformer (tft). *Results in Engineering*, page 100888, 2023.

[NSV16] David Newbery, Goran Strbac, and Ivan Viehoff. The benefits of integrating european electricity markets. *Energy Policy*, 94:253–263, 2016.

[NW15] Jakub Nowotarski and Rafał Weron. Computing electricity spot price prediction intervals using quantile regression and forecast averaging. *Computational Statistics*, 30:791–803, 2015.

[NW18] 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.

[NWA17] Bijay Neupane, Wei Lee Woon, and Zeyar Aung. Ensemble prediction model with expert selection for electricity price forecasting. *Energies*, 10(1):77, 2017.

[NZ20a] Michał Narajewski and Florian Ziel. Econometric modelling and forecasting of intraday electricity prices. *Journal of Commodity Markets*, 19:100107, 2020.

[NZ20b] Michał Narajewski and Florian Ziel. Ensemble forecasting for intraday electricity prices: Simulating trajectories. *Applied Energy*, 279:115801, 2020.

[NZ22] Michał Narajewski and Florian Ziel. Optimal bidding in hourly and quarter-hourly electricity price auctions: Trading large volumes of power with market impact and transaction costs. *Energy Economics*, 110:105974, 2022.

[OCM+23] Kin G Olivares, Cristian Challu, Grzegorz Marcjasz, Rafał Weron, and Artur Dubrawski. Neural basis expansion analysis with exogenous variables: Forecasting electricity prices with nbeatsx.

International Journal of Forecasting, 39(2):884–900, 2023.

[OCPV24a] Ciaran O'Connor, Joseph Collins, Steven Prestwich, and Andrea Visentin. Optimizing quantile-based trading strategies in electricity arbitrage, 2024.

[OCPV24b] Ciaran O'Connor, Joseph Collins, Steven Prestwich, and Andrea Visentin. Electricity price forecasting in the irish balancing market. *Energy Strategy Reviews*, 54:101436, 2024.

[OLSK+18] Gerardo J Osório, Mohamed Lotfi, Miadreza Shafie-Khah, Vasco MA Campos, and João PS Catalão. Hybrid forecasting model for short-term electricity market prices with renewable integration. *Sustainability*, 11(1):57, 2018.

[OT19] André Ortner and Gerhard Totschnig. The future relevance of electricity balancing markets in europe-a 2030 case study. *Energy Strategy Reviews*, 24:111–120, 2019.

[PB21] Heikki Peura and Derek W Bunn. Renewable power and electricity prices: The impact of forward markets. *Management Science*, 67(8):4772–4788, 2021.

[PDŁ15] Tomasz Popławski, Grzegorz Dudek, and Jacek Łyp. Forecasting methods for balancing energy market in poland. *International Journal of Electrical Power & Energy Systems*, 65:94–101, 2015.

[PdSN11] João Pinto, Jorge de Sousa, and Mário Ventim Neves. The value of a pumping-hydro generator in a system with increasing integration of wind power. In *2011 8th International Conference on the European Energy Market (EEM)*, pages 306–311. IEEE, 2011.

[Pea95] Karl Pearson. Vii. note on regression and inheritance in the case of two parents. *proceedings of the royal society of London*, 58(347-352):240–242, 1895.

[PLDV20] Ksenia Poplavskaya, Jesus Lago, and Laurens De Vries. Effect of market design on strategic bidding behavior: Model-based analysis of european electricity balancing markets. *Applied Energy*, 270:115130, 2020.

[PMM+24] Alireza Pourdaryaei, Mohammad Mohammadi, Hamza Mubarak, Abdallah Abdellatif, Mazaher Karimi, Elena Gryazina, and Vladimir Terzija. A new framework for electricity price forecasting via multi-head self-attention and cnn-based techniques in the competitive electricity market. *Expert Systems with Applications*, 235:121207, 2024.

[PS08] Anastasios Panagiotelis and Michael Smith. Bayesian density forecasting of intraday electricity prices using multivariate skew t distributions. *International Journal of Forecasting*, 24(4):710–727, 2008.

[PVG+11] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011.

[PWC+24] You Peng, Zhenyu Wang, Ivan Castillo, Gunnell LaGrande, and Shengli Jiang. A new modeling framework for real-time extreme electricity price forecasting. *IFAC-PapersOnLine*, 58(14):899–904, 2024.

[QWB+22] Zhao Qinghe, Xiang Wen, Huang Boyan, Wang Jong, and Fang Junlong. Optimised extreme gradient boosting model for short term electric load demand forecasting of regional grid system. *Scientific Reports*, 12(1):19282, 2022.

[RB90] CLEVELAND RB. Stl: A seasonal-trend decomposition procedure based on loess. *J Off Stat*, 6:3–73, 1990.

[RB18] Ray Byrne Richard Baxter. U.S. DOE Energy Storage Handbook — sandia.gov. `https://www.sandia.gov/ess/publications/doe-oe-resources/eshb`, 2018.

[RBA24] Yvet Renkema, Nico Brinkel, and Tarek Alskaif. Conformal prediction for stochastic decision-making of pv power in electricity markets. *arXiv preprint arXiv:2403.20149*, 2024.

[RM+21] Shafiul Hasan Rafi, Mohammad Mahruf Mahdi, et al. A short-term load forecasting technique using extreme gradient boosting algorithm. In *2021 IEEE PES Innovative Smart Grid Technologies-Asia (ISGT Asia)*, pages 1–5. IEEE, 2021.

[RRE22] Nafise Rezaei, Roozbeh Rajabi, and Abouzar Estebsari. Electricity price forecasting model based on gated recurrent units. In *2022 IEEE International Conference on Environment and Electrical Engineering and 2022 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe)*, pages 1–5. IEEE, 2022.

[RRF+22] G Rancilio, A Rossi, D Falabretti, A Galliani, and M Merlo. Ancillary services markets in europe: Evolution and regulatory trade-offs. *Renewable and Sustainable Energy Reviews*, 154:111850,

2022.

[RZW+23] Congjun Rao, Yue Zhang, Jianghui Wen, Xinping Xiao, and Mark Goh. Energy demand forecasting in china: A support vector regression-compositional data second exponential smoothing model. *Energy*, 263:125955, 2023.

[SA19] Priyanka Shinde and Mikael Amelin. A literature review of intraday electricity markets and prices. *2019 IEEE Milan PowerTech*, pages 1–6, 2019.

[SCEZ23] Selahattin Murat Sirin, Ercument Camadan, Ibrahim Etem Erten, and Alex Hongliang Zhang. Market failure or politics? understanding the motives behind regulatory actions to address surging electricity prices. *Energy Policy*, 180:113647, 2023.

[SDJW09] Ramteen Sioshansi, Paul Denholm, Thomas Jenkin, and Jurgen Weiss. Estimating the value of electricity storage in pjm: Arbitrage and some welfare effects. *Energy economics*, 31(2):269–277, 2009.

[SDS03] Damien C Sansom, Tom Downs, and Tapan K Saha. Evaluation of support vector machine based forecasting tool in electricity price forecasting for australian national electricity market participants. *Journal of Electrical & Electronics Engineering, Australia*, 22(3):227–233, 2003.

[SEM] SEMO. Markets and timelines.

[SEM23a] SEMO. Joining the Balancing Market — sem-o.com. `https://www.sem-o.com/markets/balancing-market-overview/joining-the-balancing-market/`, 2023.

[SEM23b] SEMO. SEMOpx Fees — semopx.com. `https://www.semopx.com/joining-the-market/fees/`, 2023.

[Ser11] Francesco Serinaldi. Distributional modeling and short-term forecasting of electricity prices by generalized additive models for location, scale and shape. *Energy Economics*, 33(6):1216–1226, 2011.

[SFF22] Md Jamal Ahmed Shohan, Md Omar Faruque, and Simon Y Foo. Forecasting of electric load using a hybrid lstm-neural prophet model. *Energies*, 15(6):2158, 2022.

[SFGJ20] David Salinas, Valentin Flunkert, Jan Gasthaus, and Tim Januschowski. Deepar: Probabilistic forecasting with autoregressive recurrent networks. *International journal of forecasting*,

36(3):1181–1191, 2020.

[SK08] Vaskar Sarkar and Shrikrishna A Khaparde. A comprehensive assessment of the evolution of financial transmission rights. *IEEE Transactions on Power Systems*, 23(4):1783–1795, 2008.

[SK22] Ali Shavandi and Majid Khedmati. A multi-agent deep reinforcement learning framework for algorithmic trading in financial markets. *Expert Systems with Applications*, 208:118124, 2022.

[Sne83] Ronald D Snee. Regression diagnostics: Identifying influential data and sources of collinearity, 1983.

[Spe61] Charles Spearman. The proof and measurement of association between two things. 1961.

[SPE22] Ana R Silva, HMI Pousinho, and Ana Estanqueiro. A multistage stochastic approach for the optimal bidding of variable renewable energy in the day-ahead, intraday and balancing markets. *Energy*, 258:124856, 2022.

[Spi14] Amanda Spisto. Market valuation of a hypothetical pump-hydro storage plant in the italian power system. In *11th International Conference on the European Energy Market (EEM14)*, pages 1–5. IEEE, 2014.

[SPYW22] Manu Suvarna, PS Pravin, Ken Shaun Yap, and Xiaonan Wang. Application of machine learning and big data for smart energy management in manufacturing. In *Computer Aided Chemical Engineering*, volume 49, pages 1699–1704. Elsevier, 2022.

[SR16] Iain Staffell and Mazda Rustomji. Maximising the value of electricity storage. *Journal of Energy Storage*, 8:212–225, 2016.

[SSMC23] Stefano Frizzo Stefenon, Laio Oriel Seman, Viviana Cocco Mariani, and Leandro dos Santos Coelho. Aggregating prophet and seasonal trend decomposition for time series forecasting of italian electricity spot prices. *Energies*, 16(3):1371, 2023.

[Sul15] Timothy John Sullivan. *Introduction to uncertainty quantification*, volume 63. Springer, 2015.

[SUW19] Tomasz Serafin, Bartosz Uniejewski, and Rafał Weron. Averaging predictive distributions across calibration windows for day-ahead electricity price forecasting. *Energies*, 12(13):2561, 2019.

[SV08] Glenn Shafer and Vladimir Vovk. A tutorial on conformal prediction. *Journal of Machine Learning Research*, 9(3), 2008.

[SW19] Simon Schnürch and Andreas Wagner. Machine learning on

epex order books: Insights and forecasts. *arXiv preprint arXiv:1906.06248*, 2019.

[SW24] Wei Shi and Yu Feng Wang. A robust electricity price forecasting framework based on heteroscedastic temporal convolutional network. *International Journal of Electrical Power & Energy Systems*, 161:110177, 2024.

[Tat18] Andrejus Ngujen Tat. Electricity price forecasting using monte carlo simulation: the case of lithuania. *Ekonomika*, 97(1):76–86, 2018.

[TFBCR19] Ryan J Tibshirani, Rina Foygel Barber, Emmanuel Candes, and Aaditya Ramdas. Conformal prediction under covariate shift. *Advances in neural information processing systems*, 32, 2019.

[TG19] Yaser Tohidi and Madeleine Gibescu. Stochastic optimisation for investment analysis of flow battery storage systems. *IET Renewable Power Generation*, 13(4):555–562, 2019.

[TJH18] Vahid Sohrabi Tabar, Mehdi Ahmadi Jirdehi, and Reza Hemmati. Sustainable planning of hybrid microgrid towards minimizing environmental pollution, operational cost and frequency fluctuations. *Journal of cleaner production*, 203:1187–1200, 2018.

[TMB+20] Jean-François Toubeau, Thomas Morstyn, Jérémie Bottieau, Kedi Zheng, Dimitra Apostolopoulou, Zacharie De Grève, Yi Wang, and François Vallée. Capturing spatio-temporal dependencies in the probabilistic forecasting of distribution locational marginal prices. *IEEE Transactions on Smart Grid*, 12(3):2663–2674, 2020.

[TPPR22] Léonard Tschora, Erwan Pierre, Marc Plantevit, and Céline Robardet. Electricity price forecasting on the day-ahead market using machine learning. *Applied Energy*, 313:118752, 2022.

[TSEEK17] Reza Tahmasebifar, Mohammad Kazem Sheikh-El-Eslami, and Reza Kheirollahi. Point and interval forecasting of real-time and day-ahead electricity prices by a novel hybrid approach. *IET Generation, Transmission & Distribution*, 11(9):2173–2183, 2017.

[TSU+24] Manish Sharma Timilsina, Subhadip Sen, Bibek Uprety, Vashishtha B Patel, Prateek Sharma, and Pratik N Sheth. Prediction of hhv of fuel by machine learning algorithm: Interpretability analysis using shapley additive explanations (shap). *Fuel*,

357:129573, 2024.

[UMW19a] Bartosz Uniejewski, Grzegorz Marcjasz, and Rafał Weron. On the importance of the long-term seasonal component in day-ahead electricity price forecasting: Part ii—probabilistic forecasting. *Energy Economics*, 79:171–182, 2019.

[UMW19b] Bartosz Uniejewski, Grzegorz Marcjasz, and Rafał Weron. Understanding intraday electricity markets: Variable selection and very short-term price forecasting using LASSO. *International Journal of Forecasting*, 35(4):1533–1547, 2019.

[Uni23a] Bartosz Uniejewski. Enhancing accuracy of probabilistic electricity price forecasting: A comparative study of novel quantile regression averaging generalization. In *2023 19th International Conference on the European Energy Market (EEM)*, pages 1–5. IEEE, 2023.

[Uni23b] Bartosz Uniejewski. Smoothing quantile regression averaging: A new approach to probabilistic forecasting of electricity prices. *arXiv preprint arXiv:2302.00411*, 2023.

[UNW16] Bartosz Uniejewski, Jakub Nowotarski, and Rafał Weron. Automated variable selection and shrinkage for day-ahead electricity price forecasting. *Energies*, 9(8):621, 2016.

[UOT18] Umut Ugurlu, Ilkay Oksuz, and Oktay Tas. Electricity price forecasting using recurrent neural networks. *Energies*, 11(5):1255, 2018.

[UTG18] Umut Ugurlu, Oktay Tas, and Umut Gunduz. Performance of electricity price forecasting models: Evidence from turkey. *Emerging Markets Finance and Trade*, 54(8):1720–1739, 2018.

[UW21] Bartosz Uniejewski and Rafał Weron. Regularized quantile regression averaging for probabilistic electricity price forecasting. *Energy Economics*, 95:105121, 2021.

[VDVAH12] Reinier AC Van Der Veen, Alireza Abbasy, and Rudi A Hakvoort. Agent-based analysis of the impact of the imbalance pricing mechanism on market behavior in electricity balancing markets. *Energy Economics*, 34(4):874–881, 2012.

[VDVDH17] RA Verzijlbergh, LJ De Vries, GPJ Dijkema, and PM Herder. Institutional challenges caused by the integration of renewable energy sources in the european electricity sector. *Renewable and Sustainable Energy Reviews*, 75:660–667, 2017.

[VdVH16] Reinier AC Van der Veen and Rudi A Hakvoort. The electricity balancing market: Exploring the design challenge. *Utilities Policy*, 43:186–194, 2016.

[VGS05] Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. *Algorithmic learning in a random world*, volume 29. Springer, 2005.

[VJ10] Vahid Vahidinasab and Shahram Jadid. Bayesian neural network model to predict day-ahead electricity prices. *European Transactions on Electrical Power*, 20(2):231–246, 2010.

[VS06] Sudhir Varma and Richard Simon. Bias in error estimation when using cross-validation for model selection. *BMC bioinformatics*, 7:1–8, 2006.

[WBKK18] IAG Wilson, Edward Barbour, T Ketelaer, and W Kuckshinrichs. An analysis of storage revenues from the time-shifting of electrical energy in germany and great britain from 2010 to 2016. *Journal of energy storage*, 17:446–456, 2018.

[Wer14] Rafał 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.

[WFH+23] Wei Wang, Bin Feng, Gang Huang, Chuangxin Guo, Wenlong Liao, and Zhe Chen. Conformal asymmetric multi-quantile generative transformer for day-ahead wind power interval prediction. *Applied Energy*, 333:120634, 2023.

[Whi80] Halbert White. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: journal of the Econometric Society*, pages 817–838, 1980.

[WHZ+20] Richard Wang, Shu-Chien Hsu, Saina Zheng, Jieh-Haur Chen, and Xuran Ivan Li. Renewable energy microgrids: Economic evaluation and decision making for government policies to contribute to affordable and clean energy. *Applied Energy*, 274:115287, 2020.

[WLS+22] Gerald Woo, Chenghao Liu, Doyen Sahoo, Akshat Kumar, and Steven Hoi. Etsformer: Exponential smoothing transformers for time-series forecasting. *arXiv preprint arXiv:2202.01381*, 2022.

[WM08] Rafał Weron and Adam Misiorek. Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models. *International journal of forecasting*, 24(4):744–763, 2008.

[WS10] Lei Wu and Mohammad Shahidehpour. A hybrid model for day-ahead price forecasting. *IEEE Transactions on Power Systems*, 25(3):1519–1530, 2010.

[WTGB17] Florian Wankmüller, Prakash R Thimmapuram, Kevin G Gallagher, and Audun Botterud. Impact of battery degradation on energy arbitrage revenue of grid-level energy storage. *Journal of Energy Storage*, 10:56–66, 2017.

[WWL+16] HZ Wang, GB Wang, GQ Li, JC Peng, and YT Liu. Deep belief network based deterministic and probabilistic wind speed forecasting approach. *Applied energy*, 182:80–93, 2016.

[WXD+22] Peng Wang, Keqi Xu, Zhaohao Ding, Yuling Du, Wenyu Liu, Beibei Sun, Zhizhong Zhu, and Huidi Tang. An online electricity market price forecasting method via random forest. *IEEE Transactions on Industry Applications*, 58(6):7013–7021, 2022.

[WXW+13] Can Wan, Zhao Xu, Yelei Wang, Zhao Yang Dong, and Kit Po Wong. A hybrid approach for probabilistic forecasting of electricity price. *IEEE Transactions on Smart Grid*, 5(1):463–470, 2013.

[WZ18] Hao Wang and Baosen Zhang. Energy storage arbitrage in real-time markets via reinforcement learning. In *2018 IEEE Power & Energy Society General Meeting (PESGM)*, pages 1–5. IEEE, 2018.

[XB19] Jian Xu and Ross Baldick. Day-ahead price forecasting in ercot market using neural network approaches. In *Proceedings of the Tenth ACM International Conference on Future Energy Systems*, pages 486–491, 2019.

[XCL+22] Hang Xie, Shijun Chen, Chunyang Lai, Guangwen Ma, and Weibin Huang. Forecasting the clearing price in the day-ahead spot market using extreme gradient boosting. *Electrical Engineering*, 104(3):1607–1621, 2022.

[XJX24] Chen Xu, Hanyang Jiang, and Yao Xie. Conformal prediction for multi-dimensional time series by ellipsoidal sets. *arXiv preprint arXiv:2403.03850*, 2024.

[XLWD24] Yan Xu, Jing Li, Honglu Wang, and Pei Du. A novel probabilistic forecasting system based on quantile combination in electricity price. *Computers & Industrial Engineering*, 187:109834, 2024.

[XQ23] Xiaoping Xiong and Guohua Qing. A hybrid day-ahead electricity price forecasting framework based on time series. *Energy*,

264:126099, 2023.

[XX21] Chen Xu and Yao Xie. Conformal prediction interval for dynamic time-series. In *International Conference on Machine Learning*, pages 11559–11569. PMLR, 2021.

[XX23] Chen Xu and Yao Xie. Sequential predictive conformal inference for time series. In *International Conference on Machine Learning*, pages 38707–38727. PMLR, 2023.

[YBC+22] Yoh Yasuda, Lori Bird, Enrico Maria Carlini, Peter Børre Eriksen, Ana Estanqueiro, Damian Flynn, Daniel Fraile, Emilio Gómez Lázaro, Sergio Martín-Martínez, Daisuke Hayashi, et al. Ce (curtailment–energy share) map: An objective and quantitative measure to evaluate wind and solar curtailment. *Renewable and Sustainable Energy Reviews*, 160:112212, 2022.

[YCL17] Zhang Yang, Li Ce, and Li Lian. Electricity price forecasting by a hybrid model, combining wavelet transform, arma and kernel-based extreme learning machine methods. *Applied Energy*, 190:291–305, 2017.

[YF17] Nanpeng Yu and Brandon Foggo. Stochastic valuation of energy storage in wholesale power markets. *Energy Economics*, 64:177–185, 2017.

[YGLZ24] Yifan Yang, Ju'e Guo, Yi Li, and Jiandong Zhou. Forecasting day-ahead electricity prices with spatial dependence. *International Journal of Forecasting*, 40(3):1255–1270, 2024.

[YKG+22] Sinan Yıldırım, Mohammad Khalafi, Tayyar Güzel, Halil Satık, and Murat Yılmaz. Supply curves in electricity markets: a framework for dynamic modeling and monte carlo forecasting. *IEEE Transactions on Power Systems*, 38(4):3056–3069, 2022.

[YS20] Li Yang and Abdallah Shami. On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, 415:295–316, 2020.

[YS22] Haolin Yang and Kristen R Schell. Qcae: A quadruple branch cnn autoencoder for real-time electricity price forecasting. *International Journal of Electrical Power & Energy Systems*, 141:108092, 2022.

[YTY+22] Yuyun Yang, Zhenfei Tan, Haitao Yang, Guangchun Ruan, Haiwang Zhong, and Fengkui Liu. Short-term electricity price forecasting based on graph convolution network and attention

mechanism. *IET Renewable Power Generation*, 16(12):2481–2492, 2022.

[Yul27] GU Yule. On a method of investigating periodicities in disturbed series with a special reference to wolfer s sunspot numbers, philosphical transactions, 1927, 226a. *Rs. crore IPFS Rs. crore IPFS*, 1927.

[YXS+17] Binbin Yong, Zijian Xu, Jun Shen, Huaming Chen, Yanshan Tian, and Qingguo Zhou. Neural network model with monte carlo algorithm for electricity demand forecasting in queensland. In *Proceedings of the Australasian computer science week multiconference*, pages 1–7, 2017.

[YY24] Firat Melih Yilmaz and Engin Yildiztepe. Statistical evaluation of deep learning models for stock return forecasting. *Computational Economics*, 63(1):221–244, 2024.

[YYT+23] Wang Yuanchen, Huang Yizhi, Guo Tong, Dai Lei, Wang Ying, and Xie Wenbin. Electricity price forecasting based on bayesian network and monte carlo simulation. In *2023 IEEE 6th International Conference on Automation, Electronics and Electrical Engineering (AUTEEE)*, pages 542–546. IEEE, 2023.

[ZAJ+19] Maheen Zahid, Fahad Ahmed, Nadeem Javaid, Raza Abid Abbasi, Hafiza Syeda Zainab Kazmi, Atia Javaid, Muhammad Bilal, Mariam Akbar, and Manzoor Ilahi. Electricity price and load forecasting using enhanced CNN and enhanced SVR in smart grids. *Electronics*, 8(2):122, 2019.

[ZCBD16] Dimitrios Zafirakis, Konstantinos J Chalvatzis, Giovanni Baiocchi, and Georgios Daskalakis. The value of arbitrage for energy storage: Evidence from european electricity markets. *Applied energy*, 184:971–986, 2016.

[ZCDG21] Behnam Zakeri, Samuel Cross, Paul E Dodds, and Giorgio Castagneto Gissey. Policy options for enhancing economic profitability of residential solar photovoltaic with battery energy storage. *Applied Energy*, 290:116697, 2021.

[ZCS18] Wenjie Zhang, Farwa Cheema, and Dipti Srinivasan. Forecasting of electricity prices using deep learning networks. In *2018 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, pages 451–456. IEEE, 2018.

[ZDL+18] Yongli Zhu, Renchang Dai, Guangyi Liu, Zhiwei Wang, and

Songtao Lu. Power market price forecasting via deep learning. In *IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society*, pages 4935–4939. IEEE, 2018.

[ZF23] Chenxu Zhang and Yong Fu. Probabilistic electricity price forecast with optimal prediction interval. *IEEE Transactions on Power Systems*, 39(1):442–452, 2023.

[ZFB22] Fan Zhang, Hasan Fleyeh, and Chris Bales. A hybrid model based on bidirectional LSTM neural network and catboost for short-term electricity spot price forecasting. *Journal of the Operational Research Society*, 73(2):301–325, 2022.

[ZFG+22] Margaux Zaffran, Olivier Féron, Yannig Goude, Julie Josse, and Aymeric Dieuleveut. Adaptive conformal predictions for time series. In *International Conference on Machine Learning*, pages 25834–25866. PMLR, 2022.

[ZHH+23] G Zachmann, L Hirth, C Heussaff, I Schlecht, J Mühlenpfordt, and A Eicke. The design of the european electricity market–current proposals and ways ahead. *European Parliament: Luxembourg*, 2023.

[ZRB24] Shuo Zhang, Emma Robinson, and Malabika Basu. Wind power forecasting based on a novel gated recurrent neural network model. *Wind Energy and Engineering Research*, 1:100004, 2024.

[ZS18] Florian Ziel and Rick Steinert. Probabilistic mid-and long-term electricity price forecasting. *Renewable and Sustainable Energy Reviews*, 94:251–266, 2018.

[ZSMM24] Grzegorz Zakrzewski, Kacper Skonieczka, Mikołaj Małkiński, and Jacek Mańdziuk. Remodels: Quantile regression averaging models. *arXiv preprint arXiv:2405.11372*, 2024.

[ZTW+22] Tingting Zhang, Zhenpeng Tang, Junchuan Wu, Xiaoxu Du, and Kaijie Chen. Short term electricity price forecasting using a new hybrid model based on two-layer decomposition technique and ensemble learning. *Electric Power Systems Research*, 205:107762, 2022.

[ZWS19] Changfei Zhao, Can Wan, and Yonghua Song. An adaptive bilevel programming model for nonparametric prediction intervals of wind power generation. *IEEE Transactions on Power Systems*, 35(1):424–439, 2019.

[ZYN+06] Ming Zhou, Z Yan, YX Ni, Gengyin Li, and Yi Nie. Electricity

price forecasting with confidence-interval estimation through an extended ARIMA approach. *IEEE Proceedings-Generation, Transmission and Distribution*, 153(2):187–195, 2006.

[ZZM+19] Siyu Zhou, Lin Zhou, Mingxuan Mao, Heng-Ming Tai, and Yihao Wan. An optimized heterogeneous structure LSTM network for electricity price forecasting. *IEEE Access*, 7:108161–108173, 2019.