FINAL REPORT

PREDICTIVE ANALYTICS

ENHANCING LOAD, WIND AND SOLAR GENERATION FOR DAY-AHEAD FORECASTING OF ELECTRICITY PRICES

Authors: Joanna Stawik, Adriana Naumczuk

Date of Submission: 07.01.2024

Table of Contents

1 INTRODUCTION

1.1 BACKGROUND OF THE TOPIC

Energy is an indispensable force that fuels our modern way of life. It powers our homes, businesses, and industries, underpinning every aspect of our daily routines. Yet, the accessibility and affordability of energy have remained pivotal concerns. The price of energy can significantly impact both individual households and the broader economy. The cost of electricity, in particular, has a direct bearing on the well-being of societies and their economic prospects. The price of electricity is not merely determined by the costs of generating power but is subject to a complex interplay of factors. These factors encompass everything from fuel prices, infrastructure maintenance, and government policies to global geopolitics and environmental concerns. As a result, the electricity market has historically been a realm of tightly regulated monopolies, where consumers had little choice and prices were often opaque. However, in recent decades, there has been a global shift towards the liberalization and deregulation of electricity markets. This transformation has brought about the establishment of electricity power exchanges like Nord Pool, EEX, PJM, and NEM, which have become the bedrock of competitive electricity trading. Day-ahead markets, where electricity offers are placed around noon on the day preceding delivery, have emerged as prominent platforms for transactions within these reformed markets [1].

This report aims to replicate and extend the main findings of the research article titled "Enhancing load, wind and solar generation for day-ahead forecasting" written by Katarzyna Maciejowska, Weronika Nitka and Tomasz Weron. The original study investigated the bias present in forecasts of fundamental variables, such as load, wind generation, and solar generation, published by Transmission System Operators (TSOs). It proposed the use of simple regression models to improve these forecasts and explored their impact on predicting spot and intraday prices in the German electricity market [1].

The replicated study will focus on validating the results of the original research by employing the same methodologies and datasets. Additionally, efforts will be made to extend the findings by incorporating additional methods, models, or alternative datasets. By doing so, we aim to contribute to the existing body of knowledge and further enhance the understanding of forecasting fundamental variables and their influence on electricity market prices.

The report is structured as follows: Chapter 2 presents and discusses the data used for the analysis. Chapter 3 introduces and describes the models employed in the replication and extension of the study. Chapter 4 presents the replicated results, while Chapter 5 offers an analysis and interpretation of the findings. Finally, Chapter 6 concludes the report by summarizing the key insights and suggesting potential avenues for future research.

By replicating and extending the original research, this report seeks to provide valuable insights into the forecasting of fundamental variables and its implications for day-ahead electricity market predictions.

1.2 SUMMARY OF RESULTS

The original research focuses on key fundamental variables such as total load, wind generation, solar generation which are collected for Germany and supplemented by forecasted temperatures for German cities (Hamburg and Munich). Additionally, it incorporates day-ahead and intraday market prices within specific bidding zones (Austria + Germany + Luxembourg before 1 October 2018, Germany + Luxembourg after 1 October 2018). It mentions the creation of electricity power exchanges and the importance of day-ahead markets in electricity trading. It also discusses the statistical properties of load, wind, and solar generation, including their dependence on factors such as the day of the week and yearly seasonality [1].

The study took data from October 1, 2015, to September 30, 2019, spanning four years. This data included prices for electricity in both the day-ahead and intraday markets across different areas. It also contained information about the actual levels of fundamental variables and forecasts made by the system. The research assessed how well these forecasts predicted the actual German market prices and calculated the financial benefits achieved through this approach [1].

The study begins by employing ARX-type models for forecasting fundamental variables. To determine the optimal length of the calibration window - combining forecasts from both short and long window sizes was used. The results reveal a significant enhancement in load forecasts over the TSO benchmark, indicating that the system operator may not fully leverage available market information. However, the ARX models (a combination of an autoregressive model and an exogenous input model) only slightly improved wind and solar predictions, indicating that these variables require a specific modeling approach, possibly involving nonlinearity or additional exogenous variables [1].

The improved forecasts of fundamental variables were put to use in predicting both day-ahead and intraday electricity prices. Both of these price predictions were made to help with better decision-making in the German market. Two types of intraday models were analyzed: one similar to the day-ahead market and another checking dependences on day-ahead prices and fundamental forecast errors. The results indicate that knowing about the actual levels of electricity generation and their structure does not significantly impact forecasting day-ahead prices. However, using Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) as accuracy measures, the research compared the forecasts for three key fundamentals (load, wind, and solar generation) over the years 2016 to 2018. The results were tested for statistical significance using the Diebold-Mariano (DM) test. The findings revealed that load forecasts could be significantly improved over the Transmission System Operator (TSO) benchmark, suggesting that the TSO might not utilize all available market information when making forecasts, with MAE reduced by 31.6%, 26.8%, and 37.9% in the respective years. RMSE also decreased by more than 20% in all years [1].

The improved forecasts of fundamentals and electricity prices were applied in the decision process. The utility had to decide where to sell 1 MWh of electricity: in the day-ahead or intraday market. The decision was data-driven and compared with a benchmark (selling in the day-ahead market). The gains from enhancing predictions were measured by additional yearly revenue. Correcting fundamental forecasts resulted in a significant income increase, with extra revenue from the market choice rising by 13.5% from 4075 EUR to 4627 EUR per year per 1 MWh. Furthermore, if the actual values of fundamentals were known, the revenue could potentially reach 16,333 EUR, providing strong motivation for further research in this field [1].

In summary, this research shows the potential to improve forecasting accuracy in the German electricity market, especially for fundamental variables. It emphasizes the importance of utilizing enhanced predictions for more informed decision-making and highlights financial benefits that can be achieved through such improvements. The findings provide a compelling foundation for future research and advancements in this critical field [1].

2 DATA DESCRIPTION AND ANALYSIS

2.1 DATASET DIVISION

The dataset covers hourly records from October 1, 2015, to September 30, 2019, divided into four years for structured analysis. The breakdown is as follows:

1. **Year 1 (October 2015–September 2016):** This initial year serves as the calibration period for models employed in forecasting fundamental variables. It lays the foundation for understanding the intricacies of the variables under consideration.

2. **Year 2 (October 2016–September 2017):** The second year is dedicated to the collection and evaluation of predictions for fundamental variables. These predictions form crucial inputs for subsequent models predicting electricity prices.

3. **Years 3 and 4 (October 2017–September 2019):** These final two years focus on assessing the performance of price forecasts generated through the proposed approach. Financial gains resulting from the application of this methodology are computed during this period.

Table 2.1 provides a concise summary of the notation conventions and the division of the dataset across the specified years.

Table 2.1. Dataset division and notation [1].

Notation	Start date	End date
2015	1 October 2015	30 September 2016
2016	1 October 2016	30 September 2017
2017	1 October 2017	30 September 2018
2018	1 October 2018	30 September 2019

2.2 VARIABLES

The dataset is composed of day-ahead ($DA_{h,t}$) and intraday ($ID_{h,t}$) market prices, specifically for the bidding zones of Austria, Germany, and Luxembourg before October 1, 2018, and Germany plus Luxembourg after that date. In this research, intraday prices are based on ID3 indexes, representing volume-weighted prices from the last 3 hours of trade [1].

The analysis incorporates actual levels and system forecasts of fundamental variables. These include:

1. **Total Load ($L_{h,t}$):** Serving as a proxy for demand, it reflects the overall electricity consumption.

2. **Renewable Energy Sources (RES):**

   - **Wind Generation ($W_{h,t}$):** Represents the contribution of wind energy.
   - **Solar Generation ($S_{h,t}$):** Represents the contribution of solar energy.

Fundamental variables are collected specifically for Germany and further enriched with the forecasted temperatures for two German cities: Hamburg and Munich ($FT_{h,t}$). The hour is denoted by 'h,' and the day number is denoted by 't'. Table 2.2 provides a concise summary of data sources, units, and notation conventions.

Table 2.2. Units and sources of data [1].

Data	Notation	Units	Source
Day-ahead prices	DA	EUR/MWh	EPEX SPOT, http://www.epexspot.com
Intraday prices	ID	EUR/MWh	EPEX SPOT, http://www.epexspot.com
Load	L	GWh	https://transparency.entsoe.eu
Wind generation	W	GWh	https://transparency.entsoe.eu
PV generation	S	GWh	https://transparency.entsoe.eu
Forecasted load	FL	GWh	https://transparency.entsoe.eu
Forecasted wind generation	FW	GWh	https://transparency.entsoe.eu
Forecasted PV generation	FS	GWh	https://transparency.entsoe.eu
Forecasted temperature	FT	°C	https://api.meteo.pl

The data was obtained from the research authors. Each dataset was specified for 15-minute intervals. The research authors aggregated the values for each hour. To obtain accurate results, the sum of the values from four 15-minute intervals should be divided by 4. This approach was implemented consistently across all datasets.

2.3 PRICE DYNAMICS

The time paths of both day-ahead and intraday prices are visually represented in Figure 2.1, illustrating time plots for two illustrative hours (4 a.m. and 6 p.m.). Specifically, two hours, h=4 and h=18, are highlighted to showcase the peak and off-peak periods within a day. Notably:

- **Peak vs. Off-Peak:** Peak prices, observed during hours of high demand, are visibly higher than off-peak prices in both day-ahead and intraday markets.

- **Temporal Variability:** The variability of prices changes over time, displaying a tendency for clustering. This suggests certain patterns or trends in market behavior.

- **Market Synchronization:** Prices in different markets exhibit a co-movement. Positive and negative spikes occur synchronously in both day-ahead and intraday markets, with the magnitude of spikes more pronounced in the intraday prices.

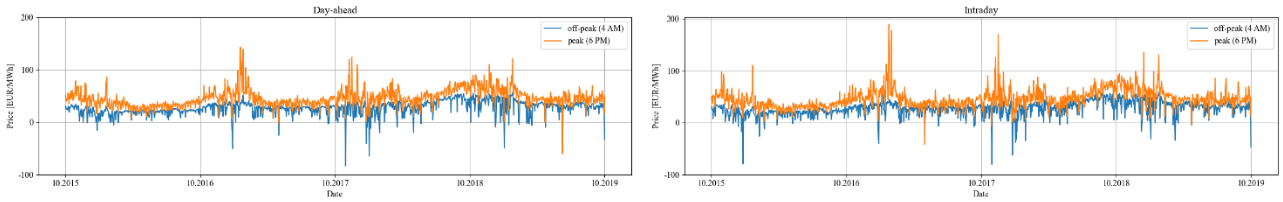


Figure 2.1. Time Plots of Day-Ahead and Intraday Prices.

In summary, the replication results are identical to the original analysis in terms of peak-off-peak differentials, temporal variability, and market synchronization.

2.4 STATISTICAL ANALYSIS OF FUNDAMENTAL VARIABLES

When considering fundamental variables, namely load, wind, and solar generations, distinct statistical properties emerge. Daily averages of these variables, along with their Transmission System Operator (TSO) forecast errors, are illustrated in Figure 2.2. Key observations include:

- **Load Dynamics:**
  - Strong dependence on the day of the week.
  - Follows a yearly seasonality, peaking during winter for heating needs and decreasing in warmer temperatures.
  - Slight increase during summer, attributed to air-conditioning usage.

- **RES Generation:**
  - Wind Generation: Rises in winter, with slight drops in summer. Exhibits significant variation within short periods.
  - Solar Generation: Peaks in summer, approaching zero in winter due to shorter days and insufficient sunlight.

- **Forecast Errors:**
  - Unlike fundamentals, forecast errors show no clear weekly or yearly patterns.
  - Solar forecast errors are larger in spring, coinciding with increased photovoltaic generation.

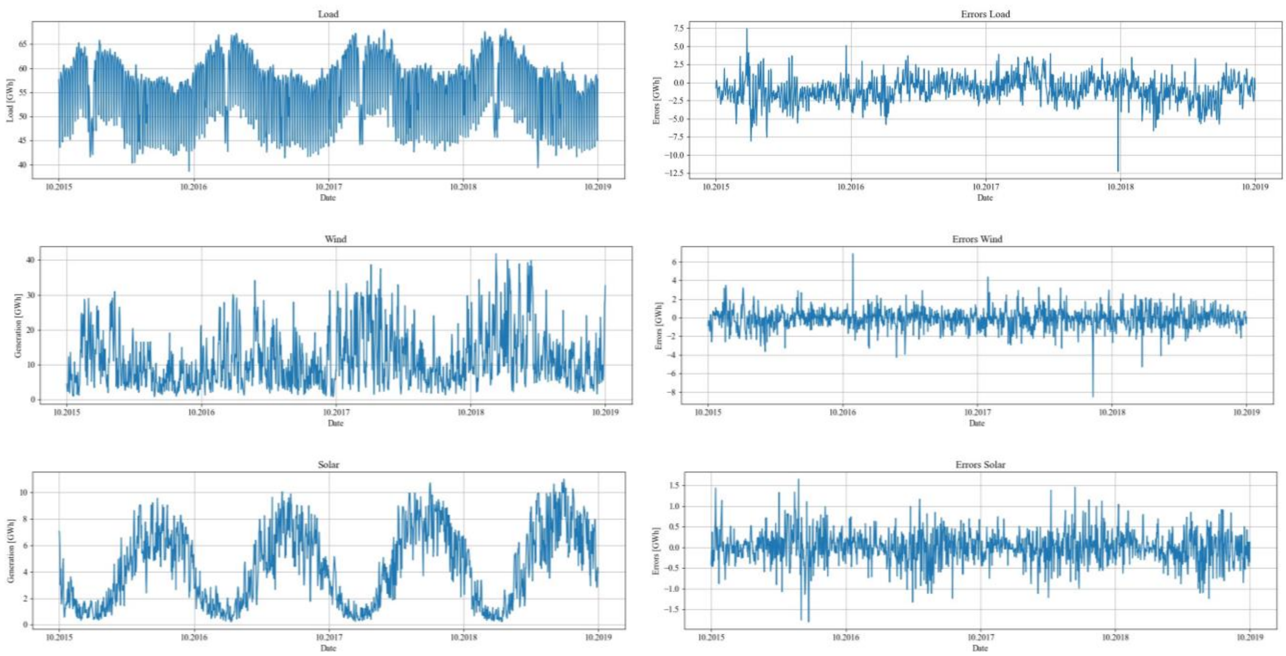


Figure 2.2. Time Plots of Day-Ahead and Intraday Prices.

In summary, the replication study successfully reproduces the essential findings related to load, wind, and solar generations, highlighting the robustness and reliability of the observed statistical properties and forecast error patterns across both analyses.

2.5 ANALYSIS OF TSO FORECAST ERRORS

The fundamental examination of Transmission System Operator (TSO) forecast errors is encapsulated in Table 2.3, shedding light on key statistical properties:

- **Mean Errors:**
    - Substantial bias across all fundamental variables.
    - Load exhibits systematic underestimation, with a mean error of -1.022 GWh during peak hours and -0.915 GWh in off-peak hours.
- **Standard Deviations:**
    - Varied biases from zero for both peak and off-peak hours for load, wind, and solar generation.
- **Autocorrelation Assessment:**
    - Ljung-Box (LB) test reveals strong autocorrelation of forecast errors.
    - Limited rejection of the null hypothesis, indicating autocorrelation, particularly evident in solar and wind generation during specific hours.

Table 2.3. TSO forecast error statistics.

Statistics	**Peak**			**Off-peak**		
	Load	Wind	Solar	Load	Wind	Solar
Mean	-1.022	-0.041	0.009	-0.915	-0.139	0.002
STD	0.017	0.012	0.008	0.014	0.012	0.001
LB test	12	11	12	12	12	5

In essence, the replicated examination reaffirms the original study's depiction of TSO forecast errors, emphasizing their systematic bias, varied deviations from zero, and a notable autocorrelation structure.

3 MODEL DESCRIPTION AND METHODOLOGY

3.1 RESEARCH FRAMEWORK

This study comprises three main components [1]:

1. Calibration of ARX-type models for load, wind, and solar generation, leading to day-ahead forecasts.
2. Evaluation of the prediction accuracy of fundamental variables.
3. Assessment of financial gains resulting from improved fundamental forecasts, measured by enhanced electricity price predictions and additional revenues from price-driven trading strategies.

The assessment includes two measures: enhancement of load, wind, and solar predictions on electricity price forecasts, and additional revenues compared to strategies based solely on TSO information.

3.1.1 Methodology

- PyCharm 2021.2.2 (Community Edition) and Matlab R2021b are the software tools used for implementation.
- For fundamental forecasts, a rolling window approach with a combination of short and long windows is employed.
- Electricity price models use a one-year observation window to capture the direct impact of enhanced fundamental forecasts.

3.2 FUNDAMENTALS FORECASTING

To forecast fundamental variables, ARX-type models are employed, incorporating information from both system forecasts and past realizations. This research employs three distinct model specifications.

First, the total load is represented by the following model [1]:

$$L_{t,h} = \alpha_h^L D_t^L + \underbrace{\theta_{h,1}^L L_{t-1.h}^* + \sum_{p\in\{2,7\}} \theta_{h,p}^L L_{t-p,h}}_{AR\ components} + \underbrace{\beta_{h,1}^L FL_{t,h} + \beta_{h,2}^L FW_{t,h} + \beta_{h,3}^L FS_{t,h}}_{Forecast\ of\ fundamentals} + \underbrace{\beta_{h,4}^L FL_{t,ave} + \beta_{h,5}^L FL_{t,max} + \beta_{h,6}^L FL_{t,min}}_{Daily\ statistics} + \underbrace{\beta_{h,7}^L FT_{t,h}}_{Weather\ forecasts} + \varepsilon_{t,h}^L \quad (1)$$

Where:

D_t^L – (4×1) vector of deterministic variables consisting of a constant and three dummy variables for Mondays, Saturdays and Sundays/Holidays;

$FL_{t,h}$, $FW_{t,h}$, $FS_{t,h}$ – TSO forecasts for all 3 fundamental variables for the current day and hour, as defined in Table 2.2;

$FL_{t,ave}$, $FL_{t,max}$, $FL_{t,min}$ – daily statistics computed as the mean, maximum and minimum of the TSO load forecast for the day t over 24 hours;

$FT_{t,h}$ – weather forecast vector includes predicted temperature for two cities: Hamburg and Munich);

$p \in \{2,7\}$ – lags which corresponds to lags used in price forecasting. This lag structure encompasses both short-term dependencies and weekly seasonality;

$L_{t-1.h}^*$ – variable which replaces missing observations with their corresponding Transmission System Operator (TSO) forecasts, because for some hours ($h > 10$) there is no information on the actual generation available at the time of forecasts [1]:

$$L_{t-1.h}^* = \begin{cases} L_{t,h} & if \quad h \leq 10, \\ FL_{t,h} & if \quad h > 10 \end{cases} \tag{2}$$

The model for wind generation is less complex than (1) and is expressed as [1]:

$$W_{t,h} = \alpha_h^W D_t^W + \theta_{h,1}^W W_{t-1.h}^* + \beta_{h,1}^W FW_{t,h} + \beta_{h,2}^W FW_{t,h-1} + \beta_{h,3}^W FW_{t,h+1} + \varepsilon_{t,h}^W \tag{3}$$

Where:

D_t^W – deterministic variable, which includes only a constant because wind does not follow a weekly seasonality;

$W_{t-1.h}^*$ – variable that controls for missing information and is defined as (2);

Equation (3) additionally incorporates data on forecasted wind generation in two adjacent hours: $h - 1$ and $h + 1$ (when accessible). The assumption is made that wind generation is independent of other fundamental variables, such as load or solar, and its autoregressive structure involves only one lag.

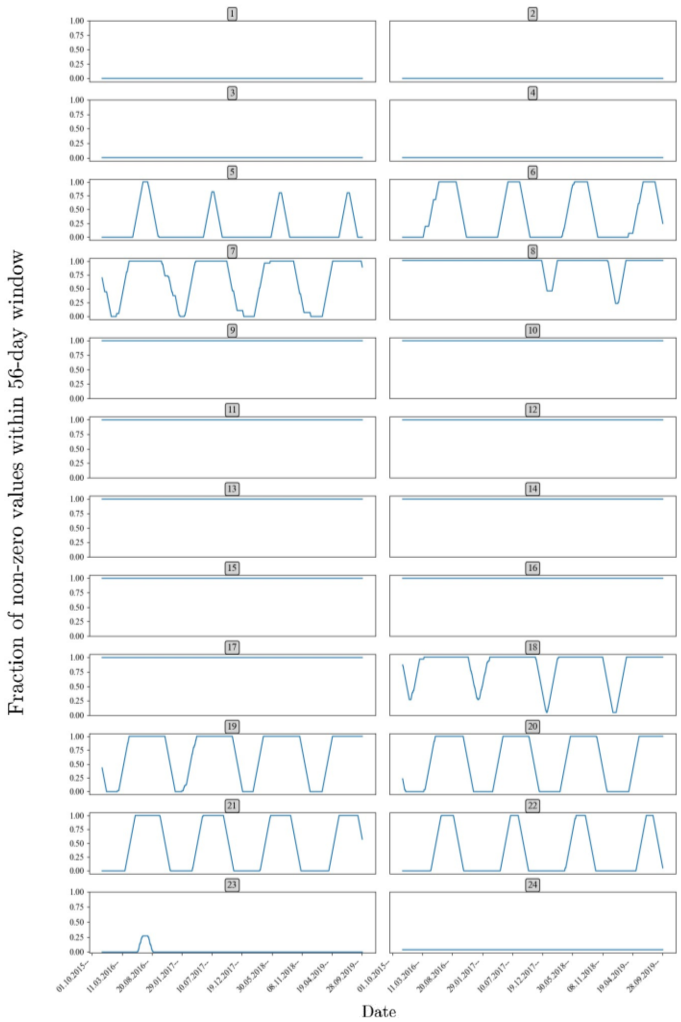


Figure 2.3. Fraction of days within a rolling 56-day period where forecasted solar energy values exceed 0 MWh for each hour of the day.

The model for solar generation follows a similar structure and is defined by the following equation [1]:

$$S_{t,h} = \alpha_h^S D_t^S + \theta_{h,1}^S S_{t-1.h}^* + \beta_{h,1}^S FS_{t,h} + \beta_{h,2}^S FS_{t,h-1} + \beta_{h,3}^S FS_{t,h+1} + \varepsilon_{t,h}^S \tag{4}$$

Where:

D_t^S– containing an intercept and the number of sun hours within a day to approximate yearly seasonality. Similar to equation (3), a straightforward autoregressive structure with one lag is employed.

Models (1), (3), and (4) are estimated using various calibration window lengths (τ). Short windows include τ values of {56, 84, 112} (equivalent to 8, 12, and 16 weeks), while long windows have τ values of {351, 358, 365} for a balanced short-term effect. Predictions are obtained through a simple average of individual forecasts.

Notably, model (4) is applicable only to hours where some TSO forecasts differ from zero in each calibration window.

Figure 2.3 illustrates the proportion of nonzero TSO solar predictions ($FS_{t,h}$) in consecutive 56-day-long calibration windows. The analysis considers only hours 8–17 meeting the nonzero condition for solar generation prediction.

3.3 FORECASTING ELECTRICITY PRICES

To generate day-ahead predictions for electricity prices, we employ autoregressive models incorporating external factors. Specifically, the $DA_{t,h}$ price for day t and hour h is formulated as [1]:

$$DA_{t,h} = \alpha_h D_t + \underbrace{\sum_{p\in\{1,2,7\}} \theta_{h,p} DA_{t-p,h}}_{AR\ components} + \underbrace{\beta_{h,4} DA_{t-1,ave} + \beta_{h,5} DA_{t-1,min} + \beta_{h,6} DA_{t-1,max}}_{Daily\ quantities} + \underbrace{\beta_{h,7} DA_{t-1,24}}_{Last\ known\ price} + \underbrace{\theta_h \hat{X}_{t,h}}_{Fundamentals} + \varepsilon_{t,h} \tag{5}$$

Where:

$DA_{t-1,ave}$, $DA_{t-1,min}$, $DA_{t-1,max}$ – average, the minimum and the maximum of prices from the preceding day;

$DA_{t-1,24}$ – the last known price;

D_t – (4 × 1) vector of deterministic variables consisting of a constant and three dummy variables for Mondays, Saturdays and Sundays/Holidays;

$\hat{X}_{t,h} = (\hat{L}_{t,h}, \hat{X}_{t,h}, \hat{X}_{t,h})$ – vector of forecasts of fundamental variables, which are based either on TSO predictions (then $\hat{X}_{t,h} = (FL_{t,h}, FW_{t,h}, FS_{t,h})'$) or results from models described in the previous section.

Please take note that the model incorporates solar generation predictions exclusively for hours 8 to 17, encompassing the time period when solar radiation is significant.

In this study, two distinct models for intraday prices are examined. The initial model mirrors the methodology used for day-ahead price forecasting and is expressed as follows [1]:

$$ID_{t,h} = \alpha_h D_t + \theta_{h,p} ID_{t-1,h}^* \underbrace{\sum_{p\in\{2,7\}} \theta_{h,p} ID_{t-p,h}}_{AR\ components} + \underbrace{\beta_{h,4} DA_{t-1,ave} + \beta_{h,5} DA_{t-1,min} + \beta_{h,6} DA_{t-1,max}}_{Daily\ day-ahead\ quantities} + \underbrace{\beta_{h,7} DA_{t-1,24}}_{Last\ known\ price} + \underbrace{\theta_h \hat{X}_{t,h}}_{Fundamentals} + \varepsilon_{t,h} \tag{6}$$

Where:

$ID_{t-1.h}^*$ – variable which replaces missing observations with their corresponding Transmission System Operator (TSO) forecasts, because for some hours (h > 10) there is no information on the actual generation available at the time of forecasts [1]:

$$ID_{t-1.h}^* = \begin{cases} ID_{t,h} & if \quad h \le 10, \\ DA_{t,h} & if \quad h > 10 \end{cases} \tag{7}$$

The second model is conditioning intraday prices on day-ahead prices and forecast errors of fundamental variables. This model extends to encompass lagged prices alongside current fundamental predictions. To approximate the forecast errors of fundamentals, the model relies on the variance between model-based forecasts and TSO forecasts. The equation representing this model is as follows [1]:

$$\begin{aligned} ID_{t,h} &= \alpha_h D_t + \beta_{h,1} \widehat{D}A_{t,h} + \beta_{h,2} ID_{t-1,h}^* + \theta_{h,1} \hat{X}_{t,h} \\ &+ \underbrace{\theta_{h,2} \left(\hat{X}_{t,h} - FX_{t,h}\right) + \theta_{h,3} (X_{t-1,h}^* - FX_{t-1,h})}_{forecast\ errors\ of\ fundamentals} + \varepsilon_{t,h} \end{aligned} \tag{8}$$

Where:

$\hat{X}_{t,h} = (FL_{t,h}, FW_{t,h}, FS_{t,h})'$ – is a (3 × 1) vector of summarizing TSO forecasts and the difference $\left(\hat{X}_{t,h} - FX_{t,h}\right)$ approximates the forecast error of fundamental variables;

$X_{t,h}^*$– variable defined similar to (2), where $X_{t,h}^* = \hat{X}_{t,h}$ for $h > 10$.

It's important to note that when estimating model (8) solely with TSO-provided information, the forecast error $(\hat{X}_{t,h} - FX_{t,h})$ equals zero and is consequently omitted from the equation. Moreover, during the calculation of $ID_{t,h}$, there's a lack of information regarding $DA_{t,h}$. Thus, the model utilizes their forecasts $\widehat{D}A_{t,h}$ obtained through model (5), in place of the actual day-ahead price levels.

It's crucial to add that when we forecast Intraday price for *None* fundamental we don't erase the forecast error $(\hat{X}_{t,h} - FX_{t,h})$, but we add $FX_{t,h}$ part instead of it.

In our approach, in model (8), we use the day-ahead forecast $(\widehat{D}A_{t,h})$ calculated for the same fundamental variable for which we aim to compute the intraday price forecast. Respectively:

- For $ID^{Enh}{}_{t,h}$ we use $\widehat{D}A^{Enh}{}_{t,h}$ part;
- For $ID^{Real}{}_{t,h}$ we use $\widehat{D}A^{Real}{}_{t,h}$ part;
- For $ID^{TSO}{}_{t,h}$ we use $\widehat{D}A^{TSO}{}_{t,h}$ part;
- For $ID^{None}{}_{t,h}$ we use $\widehat{D}A^{None}{}_{t,h}$ part.

Additionally we want to check if Intraday price forecast will provide us with better results when for each fundamental we use the same $\widehat{D}A^{Enh}{}_{t,h}$ forecast computed for *Enhanced* fundamental (9).

$$\begin{aligned} ID_{t,h} &= \alpha_h D_t + \beta_{h,1}\widehat{D}A^{Enh}{}_{t,h} + \beta_{h,2} ID^*_{t-1,h} + \theta_{h,1}\hat{X}_{t,h} \\ &+ \underbrace{\theta_{h,2}(\hat{X}_{t,h} - FX_{t,h}) + \theta_{h,3}(X^*_{t-1,h} - FX_{t-1,h})}_{forecast\ errors\ of\ fundamentals} + \varepsilon_{t,h} \end{aligned} \tag{9}$$

3.4 FORECASTING THE SIGN OF THE PRICE SPREAD

When we model the market choice, we use a similar approach to the one outlined in the original study. We define a binary decision variable $Y_{t,h}$, which represents whether the electricity produced for day *t* hour *h* is sold in the intraday market (set as 1), or not (set as 0). In this scenario, a benchmark strategy, termed the *"naïve day-ahead strategy"*, assumes that all generated electricity is sold in the day-ahead market, setting $Y_{t,h}$ as 0 for all *t* and *h*. This benchmark strategy is compared with an alternative data-driven approach that bases the decision on the relationship between day-ahead and intraday prices [1]:

$$Y_{t,h} = \begin{cases} 1 & if \quad ID_{t,h} - DA_{t,h} > 0, \\ 0 & if \quad ID_{t,h} - DA_{t,h} \le 0. \end{cases} \tag{10}$$

The utility's decision-making process relies on forecasted price differences, specifically the forecasted spread $\Delta\hat{P}_{t,h} = \hat{I}D_{t,h} - \widehat{D}A_{t,h}$ as the actual price difference $\Delta P_{t,h} = ID_{t,h} - DA_{t,h}$ isn't known at the time of decision-making. Consequently, the electricity is sold in the intraday market if the predicted spread is positive, and in the day-ahead market otherwise.

4 REPLICATION OF RESULTS

4.1 FUNDAMENTALS FORECASTING

The study assesses the potential enhancement of fundamental predictions compared to Transmission System Operator (TSO) forecasts, analyzing the outcomes through Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) measures. The models (1), (3), and (4) are pitted against TSO forecasts from 2016 to 2018. Table 2.4 presents the MAE and RMSE values for load, wind, and solar predictions during these three years.

Table 2.4. Fundamentals variables forecast accuracy.

Variable	Load			Wind			Solar		
Year	2016	2017	2018	2016	2017	2018	2016	2017	2018
MAE									
TSO	1.527	1.539	1.938	1.002	1.178	1.160	0.741	0.678	0.718
Enhanced	1.110	1.150	1.258	1.002	1.153	1.136	0.745	0.677	0.710
RMSE									
TSO	1.960	2.026	2.473	1.544	1.760	1.580	1.077	1.003	1.014
Enhanced	1.565	1.641	1.662	1.525	1.734	1.529	1.080	0.999	1.011

*The last hour of the year 2018 cannot be predicted because our formula (3) involves a future hour for which we don't have data.

4.2 FORECASTING ELECTRICITY PRICES

In our analysis, we explore the influence of enhanced load, wind, and solar generation predictions on electricity price forecasting using four different model configurations. The first setup, considered as a benchmark, incorporates TSO forecasts into specific models (5), (6), and (8). Contrarily, alternative models exclude fundamental variables from these regressions, notably simplifying model (8) to include only the predicted day-ahead prices and lagged intraday prices. Additionally, we evaluate the performance of models utilizing predictions from models (1), (3), and (4). Lastly, we investigate a scenario where researchers have access to perfect forecasts of fundamental variables, allowing them to know the real values of load, wind, and solar generations before calculating price forecasts.

When examining electricity prices, our models are fine-tuned using a year's worth of observations. This narrowed focus to a single window length allows us to distinctly capture how improved fundamental forecasts directly influence price predictions.

We present the accuracy of price forecasts, measured by MAE and RMSE, collectively for the years 2017–2018 in Table 2.5.

Table 2.5. MAE and RMSE of price forecasts (Day-Ahead) for the years 2017–2018.

Variable	**Model**	**Measure**	**Fundamentals**			
			TSO	**None**	**Enhanced**	**Real**
DA	(5)	MAE	6.050	7.252	5.983	5.939
		RMSE	8.591	10.951	8.245	8.637
ID	(6)	MAE	7.753	8.642	7.683	7.187
		RMSE	11.319	13.09	10.453	10.599
	(8)	MAE	7.630	8.362	7.314	6.800
		RMSE	10.937	11.797	10.047	10.524
	(9)	MAE	7.548	7.015	7.314	6.423
		RMSE	10.177	9.747	10.047	9.086

4.3 FORECASTING THE SIGN OF THE PRICE SPREAD

The market choice is determined by the price spread sign, with $Y_{t,h} \in \{0,1\}$, as specified in (10). If $ID_{t,h} > DA_{t,h}$, $Y_{t,h}$ is set to 1; otherwise, it's 0. As the actual prices aren't available, we predict the spread sign using day-ahead price forecasts: $\hat{Y}_{t,h} = 1$, when $\widehat{ID}_{t,h} > \widehat{DA}_{t,h}$, and $\hat{Y}_{t,h} = 0$, when $\widehat{ID}_{t,h} \leq \widehat{DA}_{t,h}$. The accuracy of forecasts, $\hat{Y}_{t,h}$, is evaluated using two measures: the correct prediction ratio (p) and additional revenues (π). The ratio of correct predictions is calculated as follows [1]:

$$p = \frac{\#(Y_{t,h} = \hat{Y}_{t,h})}{\#Y_{t,h}} \tag{11}$$

Where:

$\#(Y_{t,h} = \hat{Y}_{t,h})$ – number of correctly predicted spread signs;

$\#Y_{t,h}$ – size of the evaluation sample.

The extra revenue is derived from selling 1 MWh based on the forecasted decision variable, $\hat{Y}_{t,h}$, against the day-ahead benchmark. Therefore, the total daily additional revenue, π_t, is obtained as [1]:

$$\pi_t = \sum_{h=1}^{24} (\hat{Y}_{t,h} ID_{t,h} + (1 - \hat{Y}_{t,h}) DA_{t,h} - DA_{t,h}) = \sum_{h=1}^{24} \hat{Y}_{t,h} \, \Delta P_{t,h} \tag{12}$$

The one-year cumulative revenue, π, tallies the daily revenues from October 1, 2017, to September 30, 2018. It's crucial to note that financial metrics, often more critical in evaluating forecast accuracy than conventional statistical measures like p, guide our assessment. These results are detailed in Table 2.6.

Table 2.6. Market Classifications and Additional Revenues (2017–2018)

Models		**Correct classifications, *p (%)***			**Revenues, *π (EUR)***		
DA	**ID**	**TSO**	**Enhanced**	**Real**	**TSO**	**Enhanced**	**Real**
(5)	**(6)**	50.71	50.93	51.86	2045	2078	2632
(5)	**(8)**	50.63	51.55	64.2	2419	6912	10960
(5)	**(9)**	56.97	51.55	61.20	6443	6912	8715

5 ANALYSIS AND INTERPRETATION OF FINDINGS

5.1 FUNDAMENTALS FORECASTING

This section interprets the results of our assessment on TSO and enhanced forecasts, specifically examining the MAE and RMSE for hours with observed solar generation, varying across different seasons.

5.1.1 TSO forecasts

First we focused on calculating RMSE and MAE in Load, Wind and Solar forecasts. As expected, our results closely resemble the outcomes of the original article, which can be shown in Figure 2.4.

In our study, we specifically assess TSO Solar forecasts for hours displaying observable generation, which vary depending on the seasonal changes, specifically focusing on hours 8 to 17.

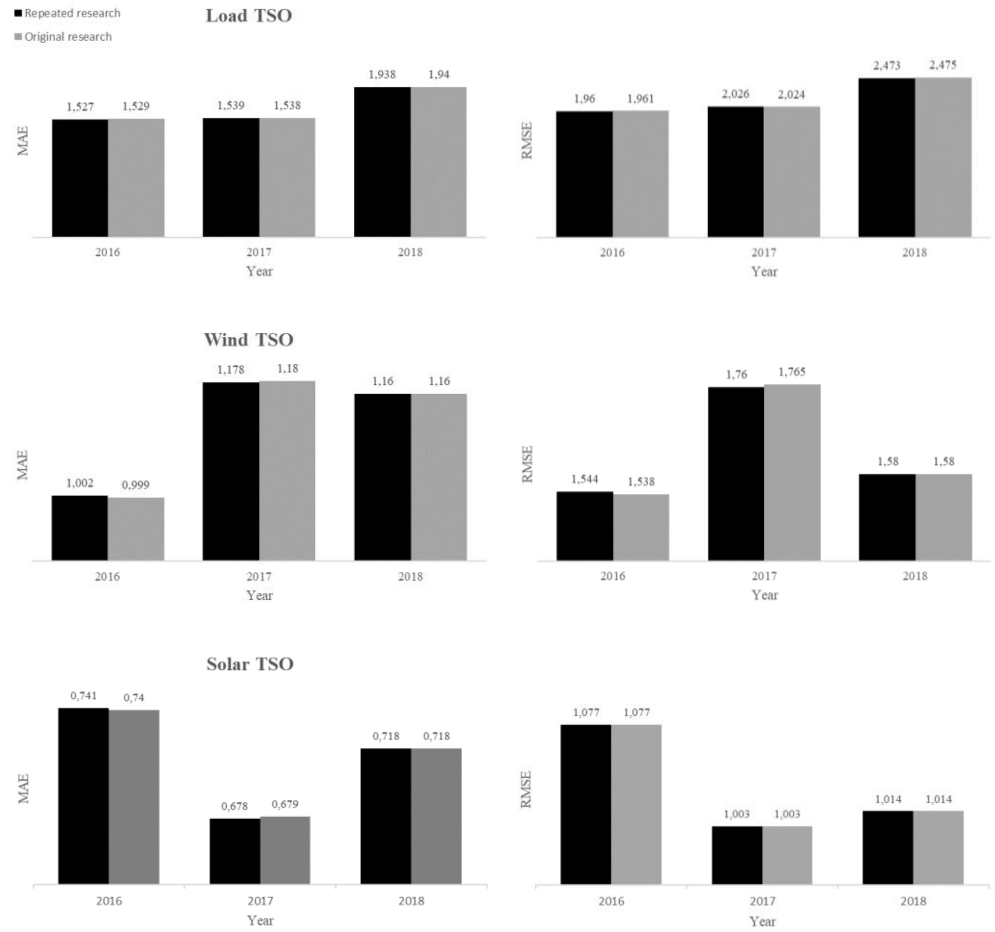
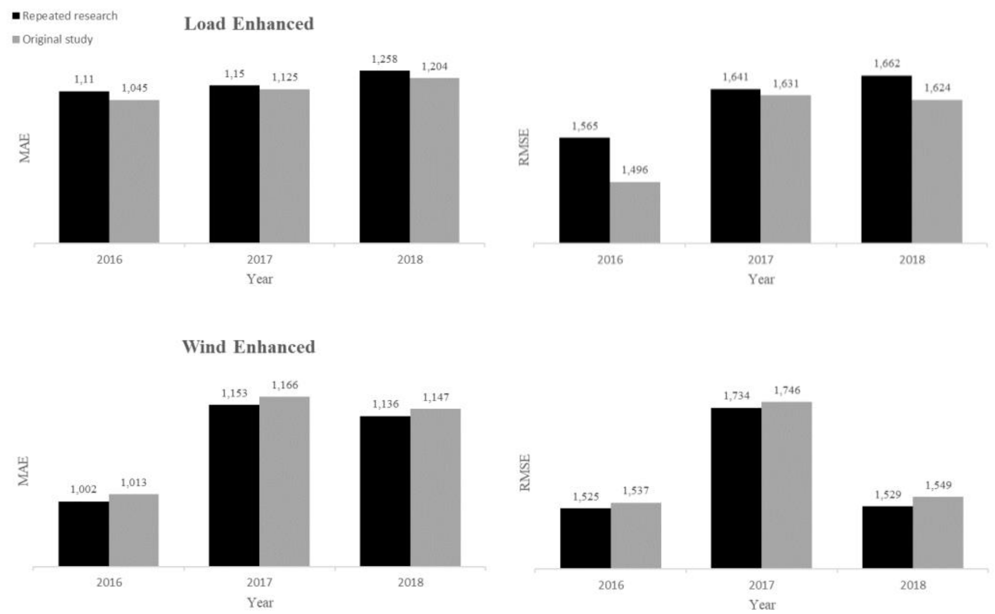


Figure 2.4. Comparison of RMSE and MAE in Load, Wind and Solar TSO Forecasts: Original vs. Repeated Study.

5.1.2 Enhanced forecasts

In our approach, we exclude the weather forecast component from model (1) since the original researchers found that this variable doesn't enhance the model and is considered irrelevant. As expected, our results resemble the outcomes of the original article, which can be shown in Figure 2.5. Little differences might be visible due to different software used or modified algorithms.



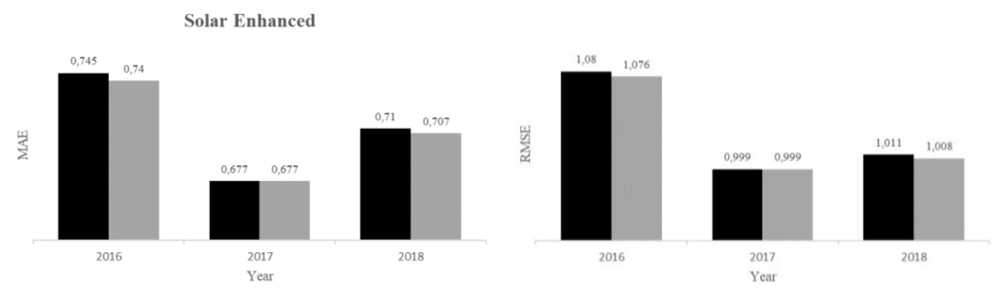


Figure 2.5. Comparison of RMSE and MAE in Load, Wind and Solar Enhanced Forecasts: Original vs. Repeated Study.

The comparison between the original research and our repeated study reveals noteworthy variations in forecasting performance for load, wind, and solar variables:

1. Load Forecast:

- **Original:** MAE reductions of 31.6%, 26.8%, and 37.9% in the years 2016 to 2018, with significant RMSE decreases by more than 20% in all years.
- **Repeated:** MAE improvements of 27%, 25%, and 35% for the years 2016 to 2018, with RMSE improvements of 20%, 19%, and 33%. The repeated study achieves comparable or slightly lower MAE and RMSE reductions than the original research for load forecasts. However, it maintains consistency in improving load forecasting accuracy over the studied years.

2. Wind Forecast:

- **Original:** Modest improvements in MAE and RMSE for wind forecasts.
- **Repeated:** Slightly better results than the original, with improvements in MAE of 0%, 2%, and 2%, and RMSE of 1%, 1%, and 3%.

3. Solar Forecast:

- Both original and repeated research find solar prediction challenging due to limitations in capturing its dynamic structure.
- We excluded negative forecasted Solar generations from further consideration and from forecasting electricity prices.

Overall, the repeated research largely aligns with the original findings in load and solar forecasting, while demonstrating slight improvements in wind prediction accuracy. The challenges in improving solar forecasts remain a consistent factor in both studies, highlighting the complexity of capturing solar generation dynamics accurately.

5.2 FORECASTING ELECTRICITY PRICES

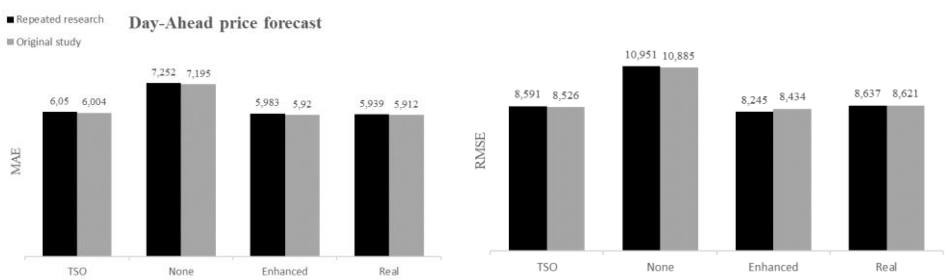
In the original study when evaluating day-ahead prices, there's a discrepancy in the results. The MAE suggests that refining fundamental forecasts (both Enhanced and Real columns) leads to more precise price predictions. However, when examining RMSE, forecasts incorporating the real generation structure show no notable difference compared to predictions from the benchmark model. This indicates that having precise knowledge of forthcoming load, wind, and solar generation levels doesn't significantly aid in forecasting day-ahead prices [1].

As expected, our results resemble the outcomes of the original article, which can be shown in Figure 2.6. Little differences might be visible due to different software used or modified algorithms.

It’s clearly visible that price forecasts for both models (Day-Ahead and Intraday model (6)) using enhanced forecasts lead to better MAE/RMSE scores. Original research leads to improving MAE and RMSE scores by 17,52% and 22,52% for Day-Ahead model using enhanced fundamentals, and 11,77% and 14,84% for Intraday model, respectively.

In our research we improve Day-Ahead model with none fundamentals by 17,5% in MAE and by 24,71% in RMSE using enhanced fundamentals. Intraday model is improved by 11,1% and 20,15%, respectively.

In summary in our repeated study we improved both RMSE scores slightly better than in original research.



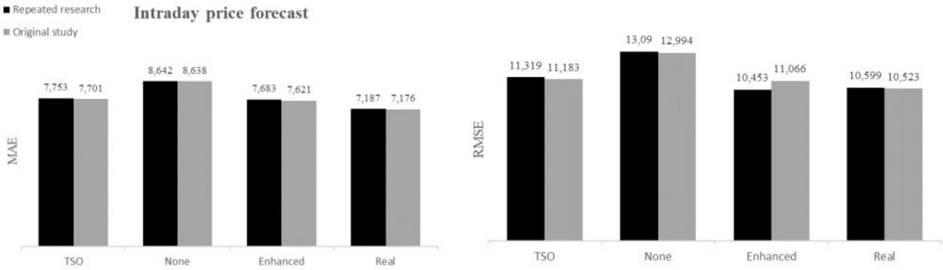


Figure 2.6. Comparison of RMSE and MAE in electricity price forecasts – Day-Ahead (5) and Intraday model (6): Original vs. Repeated Study.

We replicated the outcomes for the Intraday price forecast using model (8), similar to the original study, presented in Figure 2.7. Our findings indicate improved results across nearly all fundamentals. Concentrating on enhanced forecasts, we achieve a 12.53% enhancement in MAE and a 14.83% reduction in RMSE compared to the outcomes derived without using fundamentals.

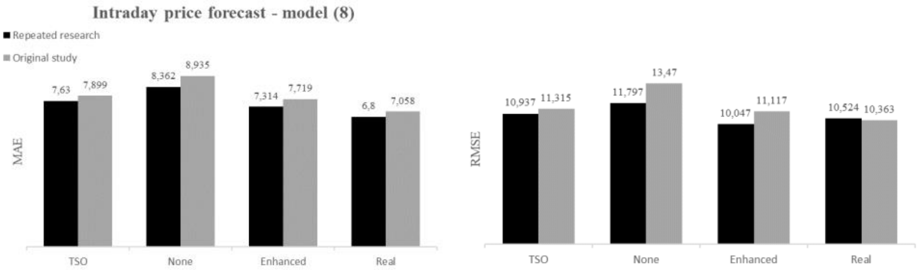


Figure 2.7. Comparison of RMSE and MAE in electricity price forecasts – Intraday model (8): Original vs. Repeated Study.

5.2.1 Extended research

In our replicated study, we sought further enhancement of the Intraday model. As previously outlined, we introduced model (9), resulting in superior outcomes, as depicted in Figure 2.8. This figure provides a comparison between two models: model (8) as implemented by the original authors and our proposed model (9).

It's evident that our model outperforms the original approach across almost all fundamental variables we forecasted. Although both models yield identical MAE and RMSE values for Enhanced fundamental, it signifies that the underlying approach remains consistent. Nevertheless, our model showcases an improvement over the results obtained in the original research.

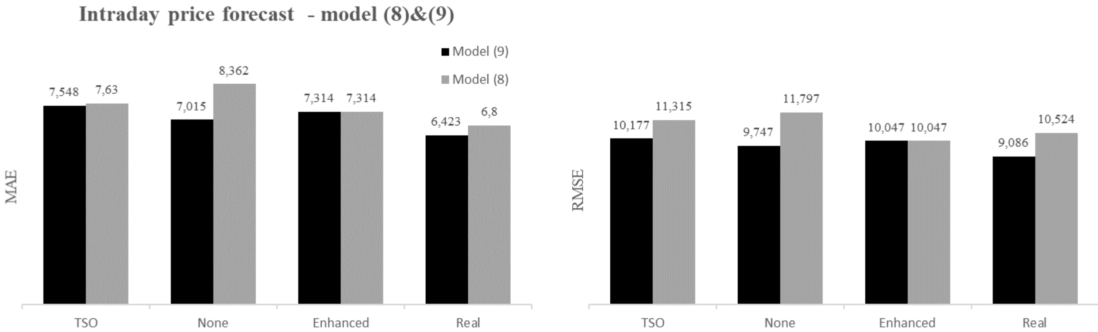


Figure 2.8. Comparison of RMSE and MAE in electricity price forecasts – Intraday model (8) vs model (9).

5.3 FORECASTING THE SIGN OF THE PRICE SPREAD

In terms of market classification accuracy, models relying on TSO-based forecasts in original article correctly predict higher-priced markets in 49.6% to 52.0% of instances. With enhanced forecasts, these ratios range from 50.8% to 52.1%, peaking with the intraday model (6). Enhanced forecasts improve market classification accuracy, reaching up to 58.9% with perfect fundamental knowledge. This surpasses the 52.0% accuracy with TSO-based forecasts. Perfect foresight on fundamentals could yield over 16,000 EUR compared to 4075–5705 EUR using TSO predictions, showing a 30% improvement. Using enhanced fundamental forecasts boosts revenues by 4627–5908 EUR. The models with enhanced forecasts consistently outperform TSO-based approaches, with model (8) generating 4627 EUR in revenues, a 552 EUR increase over the benchmark. Despite the gains, the perfect information scenario suggests further room for improvement, validating the fundamental approach [1].

In our research replication, TSO-based forecast models improved slightly, correctly predicting markets in 50.71% to 56.97% of instances. Enhanced forecasts ranged from 50.93% to 51.55%, peaking with intraday models (8) and (9). Real-based forecasts showed the highest accuracy, ranging from 51.86% to 64.2% in market predictions. Perfect foresight on Real-based forecasts indicated potential revenues of 10,960 EUR, notably higher than the 2,045-6,443 EUR range using TSO predictions. Employing enhanced fundamental forecasts increased revenues to 2,078-6,912 EUR. Models utilizing enhanced forecasts consistently outperformed TSO-based approaches, with model (8) and (9) generating 6,912 EUR in revenues—a 469 EUR and 4,494 EUR increase, respectively, over their respective benchmarks.

6 CONCLUSION AND FUTURE RESEARCH

The original study delved into key fundamental variables for the German electricity market—load, wind, and solar generation—showing enhancements in load forecasts over TSO benchmarks while wind and solar predictions improved marginally. Our replicated findings closely aligned, showcasing similar improvements in load and slight advancements in wind prediction accuracy. However, solar forecasting remained challenging in both studies, indicating its complexity.

The replicated study reinforced the original research's insights on load forecasting improvements, showcasing a 27%, 25%, and 35% enhancement in MAE over the years 2016 to 2018. These advancements, while slightly lower than the original findings, consistently highlighted a significant edge over TSO-based predictions. However, wind predictions, while displaying better accuracy than the original research by 0%, 2%, and 2%, still presented only modest enhancements.

Solar forecasting remained a challenge in both studies due to the complexities involved in capturing its dynamic structure. Excluding negatively forecasted solar generations and focusing solely on observable solar generation for forecasting electricity prices became a critical pivot in our approach. This exclusion aimed to refine the predictions and eliminate inconsistencies stemming from unreliable forecasts.

In exploring electricity price forecasting, particularly in the day-ahead and intraday markets, our study echoed the original outcomes with minor differences, attributable to varying software or modified algorithms. Intriguingly, the refined models using enhanced fundamental forecasts consistently showed superior performance, mirroring the original research's observations but with slightly better improvements in RMSE scores.

The introduction of our innovative model (9) presented a substantial leap in accuracy, showcasing better results across nearly all fundamental variables. Although the MAE and RMSE values for Enhanced fundamentals matched those of the original model (8), indicating a consistent underlying approach, our model demonstrated an overall enhancement over the original research.

Examining market classification accuracy, our replication showed slightly improved models reliant on TSO-based forecasts, correctly predicting markets in 50.71% to 56.97% of instances. Models using enhanced forecasts ranged from 50.93% to 51.55%, peaking with intraday models (8) and (9). Notably, Real-based forecasts displayed the highest accuracy, ranging from 51.86% to 64.2% in market predictions, marking a significant improvement over both TSO and Enhanced forecasts.

The potential financial gains from perfect Real-based forecasts—10,960 EUR—far exceeded the 2,045-6,443 EUR range using TSO predictions. However, employing enhanced fundamental forecasts increased revenues to 2,078-6,912 EUR, consistently outperforming TSO-based approaches. Model (8) and (9) generated 6,912 EUR in revenues—a 469 EUR and 4,494 EUR increase, respectively, over their respective benchmarks, emphasizing the lucrative potential of enhanced forecasts in decision-making within the electricity market. Looking ahead, future research could enhance models (8) and (9) further, or expand datasets to bolster accuracy.

In summary, our replicated study validated the importance of enhanced fundamental forecasts for refining electricity market predictions. It emphasized the significance of using improved forecasts for informed decision-making and financial gains, paving the way for promising future research in this crucial domain.

7 REFERENCES

[1] K. Maciejowska, W. Nitka i T. Weron, „Enhancing load, wind and solar generation for day-ahead forecasting of electricity prices,” 7 2021. [Online].