Natalia Czyżyk 252614, Weronika Urbańczyk 254330

Final Report for Predictive Analytics Project

based on the article:

"Averaging Predictive Distributions Across Calibration Windows for Day-Ahead Electricity Price Forecasting"

by Tomasz Serafin, Bartosz Uniejewski and Rafał Weron

Submitted on: 08.01.2024

1. Introduction

Electricity price forecasting is a critical aspect of the energy sector, aimed at accurately predicting both current and future prices in wholesale electricity markets. This forecasting task is highly complex and specific due to the unique nature of electricity as a commodity. Unlike other goods, electricity cannot be stored, making it essential to maintain a constant balance between production and consumption. This balance is influenced by a multitude of factors, including natural elements such as temperature, precipitation, and wind speed, as well as various human and business-related factors like peak-hours, weekdays versus weekends, and seasonal variations [1]. Given its pivotal role, accurate electricity price forecasting has become an essential component in the decision-making processes of energy companies. Consequently, there is a growing need for ongoing research and development in this area to enhance the precision and reliability of such forecasts.

In the last 25 years there has been a lot of focus on point and probabilistic predictions, but not as much attention has been paid to the critical aspect of determining and selecting the optimal length of calibration windows for these predictions. This aspect has a substantial impact on the performance of predictive models and the quality of forecast outcomes. A typical approach in this matter is choosing a window that given researchers consider as 'fit' or 'long enough', oftentimes only one or two specific time periods. However, as demonstrated by Hubicka et al. in their 2019 research [2], averaging results across multiple calibration windows has yielded much better results than basing the forecast on one chosen window.

Taking all of this into consideration, we have chosen to base our Predictive Analytics project on the following article: "Averaging Predictive Distributions Across Calibration Windows for

Day-Ahead Electricity Price Forecasting" by Tomasz Serafin, Bartosz Uniejewski and Rafał Weron [3]. This article bases on work demonstrated by Hubicka et al. and uses a well performing model described by Marcjasz et. al [4]; an ARX2 autoregressive model, fitted to asinh-transformed day-ahead prices, using six suggested calibration windows lengths. Serafin et al. then introduced two extensions to this model: Quantile Regression Averaging (QRA), which applies regression to a set of point forecasts, and Quantile Regression Machine (QRM), which first averages the point forecasts, and then applies quantile regression to the average.

The research was conducted on two datasets from two major power markets. The first is the Nord Pool. It is dominated by water and has strong seasonal fluctuations. The second is PJM, the world's largest competitive wholesale power market with a balanced mix of coal, gas and nuclear generation.The objective of this project is to build the same model using the same datasets, explore and apply the QRA and QRM techniques, attempt to replicate the results and evaluate them.

2. Results of the article

The research in the article showed that for both data sets used in this paper, it is QRM that has more accurate probabilistic predictions with shorter calibration windows. The larger the calibration window, the closer the results are to each other, but QRM still outperforms QRA. However, the authors note that this may be due to the fact that QRM has 3.5 times more parameters to estimate than QRA. Probabilistic forecasts for windows shorter than 14 days are underperforming and thus are not discussed in this paper.

In the case of QRA for both datasets, the best of the window calibration combinations for probabilistic electricity price forecasting is QRA(14:7:364). For QRM, QRM(14:7:28, 308:28:364) and QRM(14:28:70, 308:28:364) are the best and most computationally efficient. Therefore, they recommend using QRM, which is almost 3 times faster. Furthermore, QRM (14:7:28,308:28:364) outperforms QRM (14:28:70,308:28:364) on one of the dataset. It is also easy to see that the combination of three short and three long windows gives the best results.

The significance of the above results was verified by testing the conditional predictive ability (CPA) using the Diebold-Mariano test. The results were measured by the aggregate pinball score.

3. Description of the empirical research

3.1. Datasets

Our report utilizes the identical datasets as the original article. These consist of datasets from two dominant energy markets:

The Nord Pool (Northern Europe), which is characterized by hydropower domination and considerable seasonal variability. The datasets encompass hourly system prices in EUR/MWh and next-day consumption forecasts for four Nordic nations, namely, Denmark, Finland, Norway, and Sweden. The Nord Pool dataset covers 1,674 days ranging from December 31, 2013, to July 31, 2018; however, only records from December 27, 2016, to July 31, 2018 (582 days) were utilized to appraise probabilistic predictions.

The second dataset entails information from the largest competitive wholesale market for electricity in the world, the PJM Interconnection, located in the Northeastern region of the United States. It contains a well-distributed combination of coal, gas, and nuclear power. The PJM dataset encompasses hourly prices and forecasts of load for the following day in the Commonwealth Edison (COMED) domain. The PJM dataset spans 1820 days, ranging from April 9, 2013 to April 2, 2018. For the assessment of probabilistic forecasts, April 5, 2016 to April 2, 2018, a period of 728 days, is utilized.

3.2. Software used

Programming language: Python

IDE software: PyCharm and Spyder

Additional tools: Microsoft Excel

3.3. Algorithm

Python algorithm for computing point forecasts:

- Iterate over different window lengths (56, 84, 112, 714, 721, 728).
- For each window length, iterate over each hour of the day.

- Extract relevant historical data for the given window length and hour.
- Use the expert$_{DoW,nl}$ Ziel and Weron model [5] to compute point forecasts:

$$X_{d,h} = \beta_{h,1}X_{d-1,h} + \beta_{h,2}X_{d-2,h} + \beta_{h,3}X_{d-7,h} + \beta_{h,4}X_{d-1,min} + \beta_{h,5}X_{d-1,max} + \beta_{h,6}X_{d-1,24} + \beta_{h,7}C_{d,h} + \sum_{i=1}^{7} \beta_{h,7+i}D_i + \varepsilon_{d,h}$$

- Calculate the regression coefficients (betas) using least squares regression.
- Use the coefficients to make a forecast for the next time point.
- Store forecasts and absolute errors for each forecast.

Python algorithm for computing probabilistic forecasts:

- For the QRM method use an averaged vector of predictions from all six point forecasts windows and for the QRA method use a matrix of all predictions.
- For each probabilistic window length, iterate over days and hours.
- For each hour, extract the historical data for the probabilistic window length and the corresponding point forecasts.
- Apply Quantile Regression for each quantile, obtaining a distribution of predictions.
- Store the predicted quantiles.

Python algorithm for calculating pinball score and aggregate pinball score:

- Iterate over the days and hours in the stored probabilistic forecasts.
- Iterate over the quantiles in each day and hour in the stored probabilistic forecasts.
- Compare each forecast to the real value and calculate the pinball score according to the below formula

$$PS\left(\hat{Q}_q(P_{d,h}), P_{d,h}, q\right) = \begin{cases} (1-q)\left(\hat{Q}_q(P_{d,h}) - P_{d,h}\right) & \text{for } P_{d,h} < \hat{Q}_q(P_{d,h}) \\ q\left(P_{d,h} - \hat{Q}_q(P_{d,h})\right) & \text{for } P_{d,h} \geq \hat{Q}_q(P_{d,h}) \end{cases}$$

- Aggregate the pinball scores over the window lengths and combine certain scores to compare them.

4. Obtained results

Figure 1 displays a sample of point forecast results for a rolling calibration window of length 728, juxtaposed with the actual NP2018 dataset. The values are generally of the same scale, with minor variations.

	real value	forecasted value
day 1 hour 1	19,612	21,359
day 1 hour 2	16,416	21,228
day 1 hour 3	16,522	20,178
...	...	...
day 728 hour 22	23,797	27,337
day 728 hour 23	21,369	22,758
day 728 hour 24	19,525	20,509

Figure 1. Point forecasts for rolling calibration window of length 728 vs. real data for the NP2018 model.

The subsequent table, depicted in Figure 2, communicates the Mean Absolute Error (MAE) values for both datasets across six distinct calibration window lengths. In the PJM data, the MAE values increase as the length of the calibration window increases. The Nord Pool dataset shows the opposite, with MAE values decreasing as the calibration window length increases.

Rolling calibration windows of lengths:	**MAE (PJM)**	**MAE (NP2018)**
56	3,799	2,437
84	3,789	2,338
112	3,732	2,308
714	4,156	2,060
721	4,173	2,057
728	4,180	2,053

Figure 2. MAE values of point forecasts for each rolling calibration window for PJM and NP2018 model.

Figure 3 presents a subset of results depicting probabilistic predictions for a rolling calibration window of 14 days and 99 quantiles for the PJM dataset. A noticeable trend is observed, wherein larger quantiles correspond to larger prediction values.

	0,01	0,02	0,03	...	0,97	0,98	0,99
day 1 hour 1	13,86	15,25	16,48	...	25,56	27,04	27,21
day 1 hour 2	13,03	14,76	16,00	...	24,88	26,43	26,59
day 1 hour 3	12,40	13,97	15,22	...	23,81	25,46	25,60
day 1 hour 4	11,98	13,61	14,86	...	23,31	25,00	25,14
day 1 hour 5	12,92	14,42	15,67	...	24,43	26,02	26,17
...	...	...	...	...	...	...	...
day 728 hour 20	21,12	21,58	22,71	...	34,20	34,91	35,17
day 728 hour 21	22,85	23,09	24,20	...	36,27	36,78	37,07
day 728 hour 22	20,08	20,67	21,82	...	32,96	33,78	34,03
day 728 hour 23	15,27	16,48	17,69	...	27,23	28,57	28,75
day 728 hour 24	13,33	14,78	16,02	...	24,92	26,46	26,62

Figure 3. A preview of probabilistic forecasts for all hours of 728 days for 99 quantiles for the PJM dataset.

In Figures 4-7, individual probabilistic calibration window lengths are represented by filled circles, while lines with symbols indicate the window lengths chosen for averaging forecasts. In every instance, the combination (14:7:28) outperforms (14:28:70).The best result of these two combinations can be observed (in Figure 4) for Nord Pool data for QRM(14:7:28).

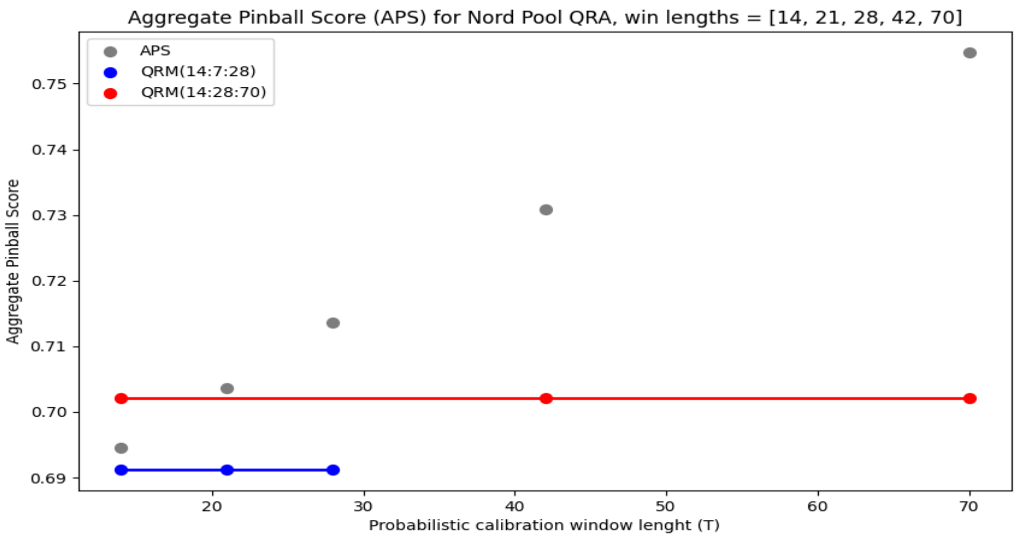
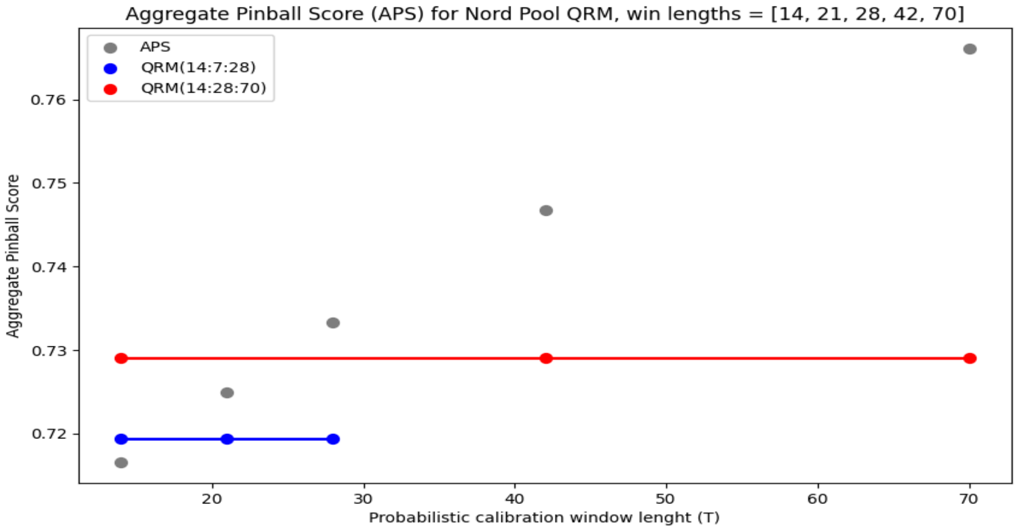


Figure 4. Aggregate pinball scores (APS) for probabilistic QRA forecasts for the Nord Pool dataset.



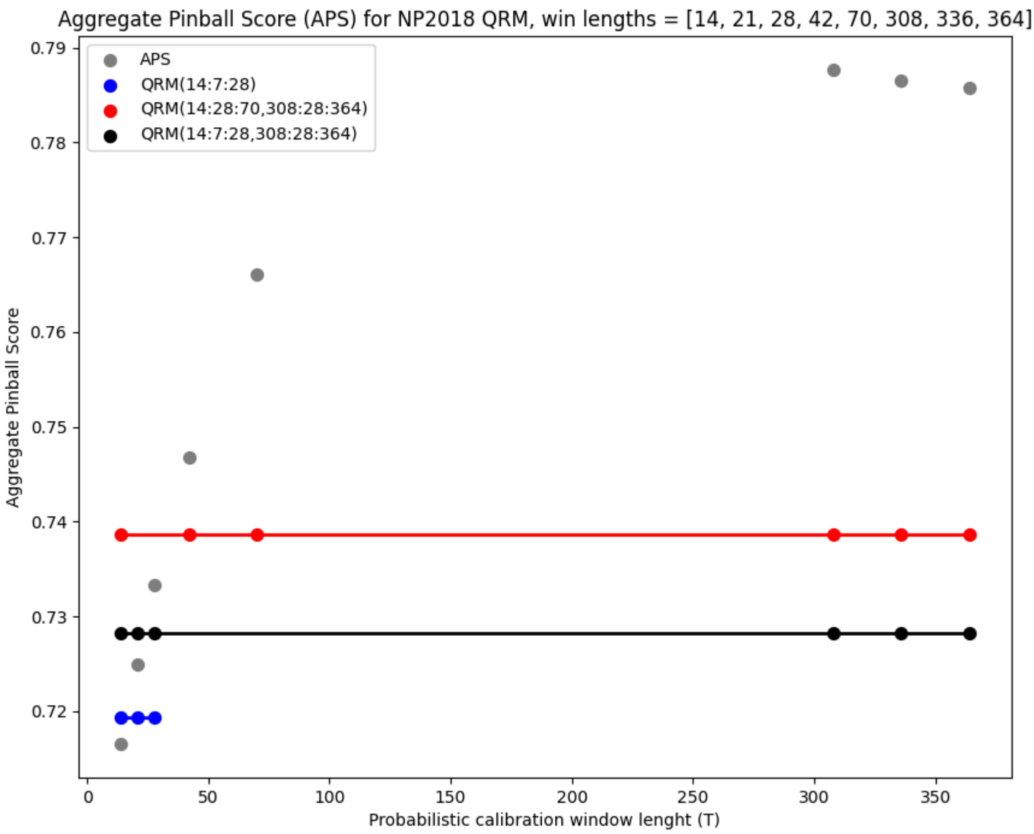


Figure 5. Aggregate pinball scores (APS) for probabilistic QRM forecasts for the Nord Pool dataset.

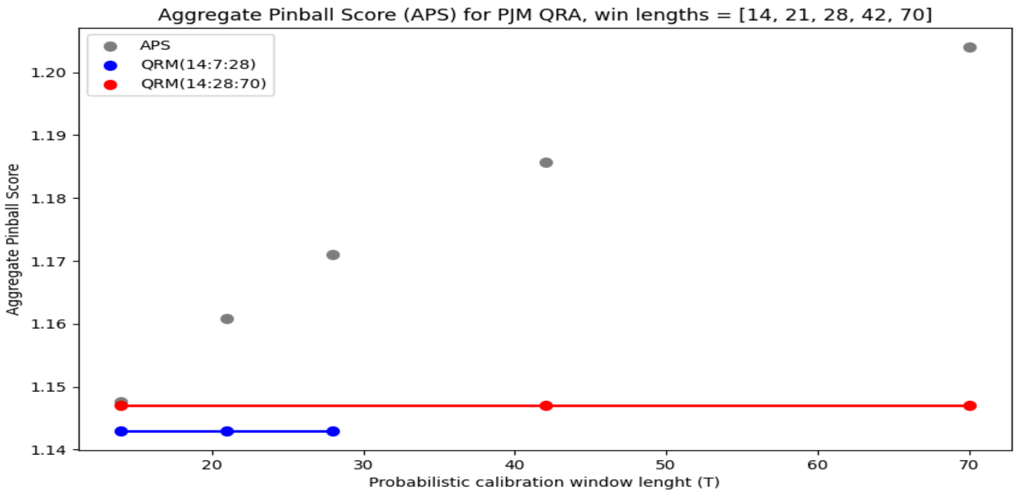


Figure 6. Aggregate pinball scores (APS) for probabilistic QRA forecasts for the PJM dataset.

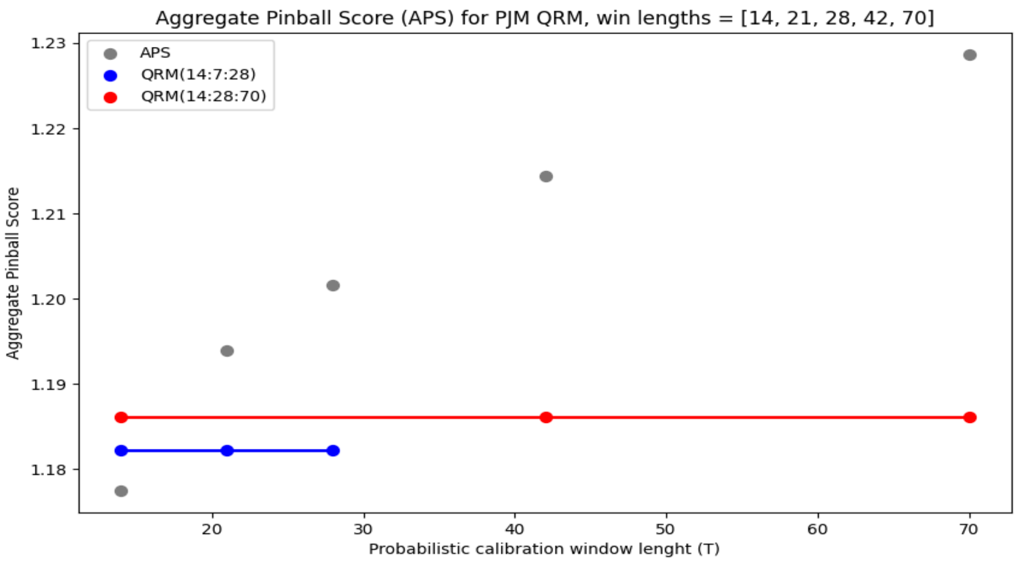


Figure 7. Aggregate pinball scores (APS) for probabilistic QRM forecasts for thePJM dataset.

5. Conclusions

We have successfully replicated a significant portion of the work presented in the article "Averaging Predictive Distributions Across Calibration Windows for Day-Ahead Electricity Price Forecasting" by Serafin, Uniejewski, and Weron. Our efforts involved fixing, preparing, and adjusting the NP2018 and PJM datasets. Subsequently, we applied the expert$_{DoW,nl}$ Ziel and Weron model [5] to generate point forecasts using six different rolling window lengths.

Our findings indicate a clear inverse relationship between the calibration window length and Mean Absolute Error (MAE) values of the point forecasts. Longer calibration windows resulted in smaller MAE values, suggesting that utilizing extended data periods enhances the reliability of beta estimates.

Subsequently, we utilized the point forecasts to generate probabilistic forecasts employing two methods for each dataset: Quantile Regression Machine (QRM) and Quantile Regression Averages (QRA). QRM proved to be more time-efficient, though both methods were computationally demanding. Given our attempt to calculate probabilistic forecasts for 350 rolling window lengths (ranging from 14 to 364 days) across 582 (NP2018) and 728 (PJM) days, considering hourly data for 99 quantiles under both QRA and QRM, our computational resources were strained, preventing the completion of all calculations.

Nevertheless, we managed to compute five probabilistic predictions for each dataset and method on days 14, 21, 28, 42, and 70. Subsequently, we assessed predictive performance using Pinball Scores and Aggregated Pinball Scores.

The obtained results were compared with the original article, and a high degree of similarity was observed. Mean Absolute Errors were computed for all predictions, and their values closely matched. Any slight discrepancies could be attributed to a difference in the methods used for computation; for example, for betas estimation we employed the numpy.linalg.lstsq function instead of the np.dot method. Finally, our project, unlike the original article, has been prepared using Python which can also influence the final results.

6. References

[1]. Weron, R. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*. **2014**, 30, 1030-1081.

[2]. Hubicka, K.; Marcjasz, G.; Weron, R. A note on averaging day-ahead electricity price forecasts across calibration windows. *IEEE Trans. Sustain. Energy.* **2019**, 10, 321-323.

[3]. Serafin, T.; Uniejewski, B.; Weron, R. Averaging Predictive Distributions Across Calibration Windows for Day-Ahead Electricity Price Forecasting. *Energies.* **2019**, 12, 2561.

[4]. Marcjasz, G.; Serafin, T.; Weron, R. Selection of calibration windows for day-ahead electricity price forecasting. *Energies* **2018**, 11, 2364.

[5]. Ziel, F.; Weron, R. Day-ahead electricity price forecasting with high-dimensional structures: Univariate vs. multivariate modeling frameworks. *Energy Econ.* **2018**, 70, 396–420.