Warsaw, 4 December 2024

Prof. dr hab. Michał Rubaszek
Head of Financial Markets Modelling Unit
Warsaw School of Economics
michal.rubaszek@sgh.waw.pl

DOCTORAL DISSERTATION REVIEW

Title:	**Decision support for electricity market participants: point and probabilistic forecasting using resampling methods and statistical learning**
Author:	**mgr Weronika Nitka**
Supervisor:	**prof. dr hab. Rafał Weron**
Assistant supervisor:	**dr hab. Katarzyna Maciejowska**
Discipline of science:	**Management and Quality Studies**

The review is based on a letter from prof. Rafał Weron, the Chairman of the Scientific Council of the Management and Quality Studies Discipline at Wrocław University of Science and Technology.

Research objective

The main objective of the thesis is to develop automated and data-driven electricity price forecasting methods with low to medium computational complexity, which can be used by small and medium size enterprises in the decision-making processes.

There are four detailed research objectives:

O1. Conduct a critical study of calibration sample selection for automation of electricity price forecasting.

O2. Use resampling methods to generate predictive distributions of electricity prices and better assess uncertainty.

O3. Utilize renewable generation and load forecasts to design trading strategies in day ahead and intraday markets.

O4. Develop decision support methods for day-ahead bidding that use combinations of predictive distributions.

Structure and content of the dissertation

The thesis is divided into two parts. The first one consists of the introduction, the description of electricity markets, and a summary of the key findings of the dissertation. The second part wraps five articles that constitute the main achievement to be evaluated in this report.

A1. W Nitka, T Serafin, D Sotiros, 2021, Forecasting electricity prices: Autoregressive hybrid nearest neighbors (ARHNN) method, International Conference on Computational Science, 312-325.

Article **A1** proposes a new method of sample weighting scheme in estimating ARX model, which is based on k-nearest neighbors (kNN) algorithm applied to explanatory variables. This method is designed to estimate ARX model parameters on a subset of historical data that are most relevant to the current situation. Next, on the basis of data for the German electricity market (EPEX) from 2015-2020, the PhD Student reports that the proposed method delivers better electricity point forecasts compared to rolling window ARX benchmarks. This article addresses objective **O1**.

A2. J Nasiadka, W Nitka, R Weron, 2022. Calibration window selection based on change-point detection for forecasting electricity prices, International Conference on Computational Science, 278-284

A2 constitutes an extension of **A1.** It also addresses objective **O1,** is based on the same EPEX data, and explores ARX models of the same specification. The main differences between **A1** and **A2** relate to the choice of the sample weighting method and the transformation of the input variables. In particular, **A2** applies the narrowest-over-threshold (NOT) algorithm (instead of kNN) to select subsample of observations with the same distribution as the current observation. As regards transformation, **A2** explores the advantages of $asin$ transformation. The main results is that (i) $asin$ transformation allows to reduce RMSE (by about 10%), and (ii) that NOT sample selection does not significantly improve the accuracy of point forecasts.

A3. K Maciejowska, W Nitka, 2024. Multiple split approach--multidimensional probabilistic forecasting of electricity markets, DOI: 10.48550/arXiv.2407.07795

A3 addresses objectives **O2** and **O3.** It proposes a Multiple Split method to produce probabilistic forecasts for four variables describing the EPEX market: day-ahead price, intraday price, total load, and RES generation. The quality of this model is assessed using both statistical and economic criteria. Firstly, it is showed that the Multiple Split model performs better than the historical simulation and quantile regression alternatives in terms of interval quality forecasts. Secondly, forecasts from the Multiple Split method are applied to generate trading strategies for a wind farm that needs to decide how to sell production on the day-ahead and intraday market. The PhD Student reports that the proposed method allow to augment the average selling price in comparison to a naïve strategy.

A4. K Maciejowska, W Nitka, T Weron, 2021. Enhancing load, wind and solar generation for day-ahead forecasting of electricity prices, Energy Economics 99, 105273

A4 also applies the ARX framework to describe the dynamics of the German day-ahead and intraday market. In this case the focus is on assumptions for the explanatory variables (load, wind and solar generation), the values for which are taken in three variants: (i) a forecast from the Transmission System Operator, (ii) a forecast from the ARX model and (iii) realized values (unknown at time of forecast formulation). It is showed that (i) load forecasts from the ARX model are significantly better than those provided by TSO, (ii) this can be used to improve the quality of point forecasts for electricity prices, (iii) and to develop a market neutral trading strategy exploiting the forecasted spread between day-ahead and intraday markets. In the period from Oct. 1, 2017 to Sept. 30, 2019 this strategy delivered around 5k EUR of profit per each traded MWh.

A5. W. Nitka, R Weron, 2023. Combining predictive distributions of electricity prices. Does minimizing the CRPS lead to optimal decisions in day-ahead bidding?, Operations Research and Decisions 33, 105-118.

A5 addresses objective **O4** by investigating on how ensemble forecasts perform in terms of forecast accuracy and trading strategies, once again using the EPEX data. The PhD Student considers a pool of distributional forecasts from twelve models (eight deep distributional neural networks and four quantile regressions based on AR models). These twelve methods are combined by using weights maximizing past performance measured by CRPS, were the weights are chosen separately for each quantile. The main results are twofold. Firstly, it is showed that for point and distributional forecasts, the worst ensemble forecast is more accurate than the best individual forecast. Second, it is reported that ensemble forecasts can be used to develop a profitable trading strategy for an owner of a rechargeable battery. By using methods developed in **A5**, such an owner could have gained as much as 12 225EUR over the period 27 June 2019 – 31 December 2020.

Comments

In my opinion, there are several reasons to consider the dissertation as valuable contribution to the discipline of Management and Quality Sciences. The most important ones are:

1. It addresses an important and up-to-date topic of electricity price forecasting.
2. It delivers new results on the usefulness of ARX models in point and density electricity price forecasting for day-ahead and intraday markets.
3. It proposes a set of guidelines on how to use ARX models in forecasting electricity prices.
4. It shows that ARX-based model forecasts are not only accurate in statistical sense, but also deliver economic value as they can be used to develop trading strategies.

Apart from that, it should be appreciated that the thesis explicitly tackles up-to-date challenges in energy forecasting, namely those related to the increasing importance of renewables and storage capacity. This makes the work timely and significant in areas that needs decision support tools for handling variable power sources.

In general, I very like and highly evaluate the dissertation. I don't have any major comments, especially as four among five evaluated articles have been accepted in top academic journals, hence they underwent tough review process. However, I would like to point to three issues. First of all, the thesis uses data from the German EPEX market. It is correct, but raises a question on whether the key findings are representative for other electricity markets. Secondly, the main goal of the thesis is to develop automated and data-driven electricity price forecasting methods with low to medium computational complexity that would be helpful for small enterprises. In my opinion some methods, especially those in **A5**, are too complex for small firms with limited analytical resources. If the presented methods are supposed to be adopted in practice, a toolbox with an extensive manual would be needed. Third, electricity markets are subject to continuous structural change, which raises a question on how the proposed methods could be adapted effectively in a small company so that they were responsive enough to the changing environment.

I have also formulated four questions, which I would be happy to discuss during thesis defense.

Question 1. My first question is related to the method proposed in **A1**. Is it possible to replicate the method presented in **A1** with a weighted least squares estimation, in which the weights are given by applying the kNN algorithm? If yes, what would be the pattern of these weights (w_s, where s is the distance from the last observation)? Would it resemble exponentially decaying weights as in the exponentially weighted regression model?

Question 2. To what extent the forecast performance improvement presented in Table 1 of **A1** is driven by the application of kNN extension with a constant k, and to what extent by the averaging procedure described in equation (2)? A more specific version of this question would be: what is the difference in the forecasting performance if we use k last observations instead of k observations selected by kNN? This could be done if Table 1 would provide the results for ARHNN(364), which could be compared to Win(364), etc.

Question 3. In subsection 3.1 of **A3** there are separate ARX models for variables that are linked through an identity, namely:

- Total load (L, eq. 1), renewable generation (RES, eq. 4) and residual load ($RL = L - RES$, eq. 5).
- Day-ahead prices (DA, eq. 6), intraday process (ID, eq. 7) and their spread ($SP = DA - ID$, eq. 8).

Given that these variables are linked through an identity, if model specification for two variables is correctly specified, the third model is mis-specified (the dependent variable is implied by the identity and the remaining two equations). In the empirical application of the Multiple Split method the set of variables was limited to DA, ID, L and RES. My question is if the results of the **A3** dependent on the choice of variables -- would they differ for the system consisting of DA, SP, RES and RL? A more general form of this question is:

- Why in a system of two price indices it was chosen to use two price indices (DA and ID) instead of the headline index and the spread (DA and SP)?
- Why in a system of a quantity that consists of two subcomponents, is was chosen to use a total and one subcomponent (L and RES) instead of two subcomponents (RES and RL)?

Question 4. The last question is related to the main objective of the thesis (to develop electricity price forecasting methods that can be used by small and medium size enterprises in the decision-making processes). What kind of SME could be interested in the use the results from **A3, A4** and **A5**? Moreover, could these results be useful in developing energy transition policy on a country level, related to investment in renewables and storage capacity?

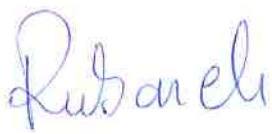
Additional information

Weronika Nitka has authored or co-authored many scientific publications, which is outstanding at this stage of scientific career. The Google Scholar profile shows 11 items (some published in top field forecasting or energy journals) that have accumulated 265 citations and h-index of 7 (date 4 December 2024).

Conclusion

The research carried out by Weronika Nitka shows that she has profound understanding of time-series methods, the functioning of electricity markets and forecasting literature. Moreover, she shows good skills in terms of processing large and sophisticated databases. Without doubts, the presented dissertation constitutes valuable contribution to the electricity price forecasting literature. Consequently, it meets the requirements for doctoral theses as described in the Act of 20 July 2018 on Scientific Degrees and Academic Title. I therefore recommend to admit the thesis to the next stages of the doctoral dissertation proceedings. Additionally, I propose awarding the thesis in accordance with the regulations of the University.

Prof. dr hab. Michał Rubaszek



KIEROWNIK
Zakład Modelowania Rynków Finansowych

prof. dr hab. Michał Rubaszek

www.sgh.waw.pl Instytut Ekonometrii, Szkoła Główna Handlowa w Warszawie, al. Niepodległości 162, 02-554 Warszawa tel.: +48 22 564 92 56, ie@sgh.waw.pl, www.sgh.waw.pl/ie