HSC Research Report

HSC/14/03

Merging quantile regression with forecast averaging to obtain more accurate interval forecasts of Nord Pool spot prices

Jakub Nowotarski¹
Rafał Weron¹

¹Institute of Organization and Management, Wrocław
University of Technology, Poland

Hugo Steinhaus Center
Wrocław University of Technology
Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland
<http://www.im.pwr.wroc.pl/~hugo/>

Merging quantile regression with forecast averaging to obtain more accurate interval forecasts of Nord Pool spot prices

Jakub Nowotarski

Institute of Organization and Management
Wrocław University of Technology
Wrocław, Poland
jakub.nowotarski@pwr.wroc.pl

Rafał Weron

Institute of Organization and Management
Wrocław University of Technology
Wrocław, Poland
rafal.weron@pwr.wroc.pl

Abstract—We evaluate a recently proposed method for constructing prediction intervals, which utilizes the concept of quantile regression (QR) and a pool of point forecasts of different time series models. We find that in terms of interval forecasting of Nord Pool day-ahead prices the new QR-based approach significantly outperforms prediction intervals obtained from standard, as well as, semi-parametric autoregressive time series models.

Index Terms—Electricity spot price; Prediction interval; Quantile regression; Forecasts combination

I. INTRODUCTION

As of today the electricity spot price forecasting literature is relatively rich and includes over 200 Scopus-indexed articles and nearly 300 Scopus-indexed conference papers. Statistical/econometric approaches (like multiple regressions, AR, ARIMA, AR-GARCH, jump-diffusions, factor models and regime-switching models) and computational intelligence techniques (like neural networks, fuzzy techniques and support vector machines) constitute the two main streams of models. Yet, despite this abundance of models, two issues remain open. First, it is impossible to select one, most reliable approach [1], [8], [19]. This fact is a good motivation for considering combining electricity spot price forecasts obtained for different predictors (methods). Surprisingly, this approach has not been considered in the electricity spot price forecasting literature until very recently [4], [14], [16]. Second, while there are a variety of empirical studies on point forecasts (i.e. the ‘best guess’ or expected value of the spot price), the more valuable from a risk management perspective interval and density forecasts have not been investigated very extensively to date (for a comprehensive review see [20]).

Hence, the main objective of this study is to construct prediction intervals (i.e. interval forecasts) with respect to the two issues mentioned above. We should emphasize here that some authors use the term *confidence interval* instead of *prediction interval* (PI). However, a PI is associated with a random variable (e.g. electricity price) yet to be observed, while a confidence interval is associated with a parameter of a model, see [9] for a discussion. In most forecasting applica-

tions we are interested in PI, i.e. intervals which contain the true values of future observations with specified probability, not in confidence intervals.

From a practical point of view, PI provide additional information on price forecasts. High volatility and uncertainty of electricity price forecasts may frequently deviate from the true price levels. In fact, possible errors in point predictions always occur regardless of the proposed framework, i.e. for any statistical model. Chatfield [6] admits that the majority of forecasters are aware of interval forecasts’ importance and lists the following reasons of that: (i) assessment of future uncertainty, (ii) ability to plan different strategies for the range of possible outcomes indicated by the interval forecast, and (iii) possibility of more thorough forecasts comparison. Interestingly, as Amjady and Hemmati [1] note, also electrical engineers are aware that high-quality market clearing price PI would help utilities submit effective bids with low risks.

Although the idea of combining interval forecasts is not new by itself [17], [18], this concept has not been utilized in the context of electricity spot prices, except for a recent working paper [15]. It is exactly our aim here to evaluate the method proposed in the latter paper in the context of Nord Pool spot prices. The article is structured as follows. In Section II we discuss the models and combining techniques used in our study. In Section III we introduce the dataset and in Section IV we evaluate the interval forecasting performance of the three tested models in terms of unconditional and conditional coverage. Finally, in Section V we wrap up the results and conclude.

II. THE MODELS

As mentioned above, our goal in this study is to construct short-term prediction intervals (PI) of electricity spot prices. Recall that a one step-ahead prediction interval constructed at time t with given confidence level $(1 - \alpha)$, where $\alpha \in [0, 1]$, is an interval $[L_{t+1|t}, U_{t+1|t}]$ such that

$$P(p_{t+1} \in [L_{t+1|t}, U_{t+1|t}]) = 1 - \alpha, \quad (1)$$

where p_{t+1} is the future (i.e. time $t+1$) spot price. Hence, prediction intervals tell us what is the range of future values that the true spot price is predicted to lie within, with probability $(1 - \alpha)$.

Let us now explain the methodology used in this study. The construction of PI is a multi-step procedure and begins with obtaining a set of point forecasts (so-called *individual forecasts*). Then, we use them in a forecast combination scheme utilizing the concept of quantile regression. Unlike in a linear regression model, here the output is a quantile of the spot price. With a pair of those, say 5% and 95%, we obtain a prediction interval, i.e. the 90% PI. A detailed description of the models and techniques used in this paper is given below.

A. Individual point forecasts

The first step of all empirical applications using forecast averaging is the specific choice of individual models. An independence of the model implied forecasts – while desired from a statistical point of view – is generally impossible to achieve since the future prices are driven by certain factors, in particular the current spot price, so all the models depend on the same or a similar set of variables. Yet, we can try to choose individual models in a way that would minimize the potential co-dependence of their forecasts, e.g. by selecting different modeling frameworks. In this study we follow [20] and consider six ‘price-only’ models: AR, TAR, SNAR, MRJD (all four models as in [21]), NAR (a non-linear AR) and a multivariate 3-factor model (FM). The former five models split the original time series into 24 separate subsets, one each for hour of a day, and essentially yield 24 ‘independent’ models. The factor model, on the other hand, extracts the information from all hourly prices.

The autoregressive model (AR) is defined as follows:

$$p_t = \phi_1 p_{t-24} + \phi_2 p_{t-48} + \phi_3 p_{t-168} + \phi_4 m p_t + d_1 D_{Mon} + d_2 D_{Sat} + d_3 D_{Sun} + \varepsilon_t, \quad (2)$$

where p_t represents the logarithm of the Nord Pool spot price at time t , $m p_t$ creates a link between bidding and price signals from the entire previous day (it is the minimum of the previous day’s 24 hourly log-prices, see [13]) and the three dummy variables D_{Mon} , D_{Sat} , D_{Sun} account for weekly seasonality. This model is also used as a benchmark model for interval forecasts. The model’s (empirical) PI are computed as sample quantiles of the empirical distribution of the one step ahead prediction errors.

The threshold AR (TAR) model is a natural extension of the AR model to two regimes. We assume that the price is governed by the value of an observable threshold variable v_t (equal to the difference in mean prices for yesterday and eight days ago) which switches between the two regimes. The smoothed nonparametric AR model (SNAR) retains the functional form of the AR model, see eqn. (2), but the parameter estimates are obtained from a numerical maximization of the empirical likelihood as suggested by Cao et al. [5], and in the context of electricity price forecasting by Weron and Misiolek [21]. The model’s prediction intervals are constructed

differently than for AR – here we assume a nonparametric distribution of the innovations. Hence, interval forecasts are taken as quantiles of the kernel estimator of the error density. The SNAR model was found to be successful in interval forecasting and we use it here as the second benchmark.

The the mean-reverting jump diffusion (MRJD) model used in this study is given by the following formula:

$$p_t = \phi_1 p_{t-24} + d_1 D_{Mon} + d_2 D_{Sat} + d_3 D_{Sun} + \varepsilon_{t,i}, \quad (3)$$

where the subscript i takes the value of 1 when there is no jump or 2 if there is a jump, $\varepsilon_{t,1} \sim N(0, \sigma^2)$ and $\varepsilon_{t,2} \sim N(\mu, \sigma^2 + \gamma^2)$. In discrete time (and this is our setup), the MRJD model is equivalent to a set of two AR(1) processes with different noise terms. The second, ‘jump’ AR(1) process is chosen with probability equal to intensity λ of the Poisson component.

The non-linear AR (NAR) model is a recurrent network with a three-layer feed-forward architecture, sigmoid activation functions in the hidden layer, linear activation functions in the output layer and delay lines to store previous values of the predicted time series. Despite the fact that the NAR architecture is a natural extension of the most popular benchmark time series model (i.e. AR), except for [2], [20], it has not been applied in the electricity price forecasting literature. In our study, the NAR model has the same inputs as the AR model in eqn. (2) and is estimated using the Levenberg-Marquardt algorithm.

Finally, following [12], [20] we use a factor model (FM) of hourly prices p_{kt} , where $k = 1, \dots, 24$ represents the hour of the day and t – the day in the sample. The main assumption of the model is that all hourly prices co-move and depend on a small set of common factors $\mathbf{F}_t = [F_{1t}, \dots, F_{Nt}]'$. The individual series p_{kt} can be modeled as a linear function of the principal components $\mathbf{F}_t$ and stochastic residuals ν_{kt} :

$$p_{kt} = \mathbf{\Lambda}_k \mathbf{F}_t + \nu_{kt}, \quad (4)$$

where the *loads* (or *loadings*) $\mathbf{\Lambda}_k = [\Lambda_{k1}, \dots, \Lambda_{kN}]$ describe the relation between the factors $\mathbf{F}_t$ and the panel variables p_{kt} . Note, that these loads are not ‘power system loads’, but model parameters, as in [3]. The number of common factors can be chosen on the basis of information criteria or the fraction of total variability explained. In this study we use $N = 3$ factors.

In order to predict future values of p_{kt} , we need to forecast both, the common factors F_{nt} and the idiosyncratic components ν_{kt} . Although the factors are contemporaneously orthogonal, due to normalization assumptions, they may be still inter-temporally correlated. Hence, it seems reasonable to model them jointly. Moreover, they may depend on some other variables, such as the deterministic day-of-the-week variables ($\mathbf{D}_t$). At the same time, the idiosyncratic components can be only weakly correlated across periods and therefore can be modeled separately, for each hour. Moreover, they cannot have the same seasonal pattern because all the co-movement between hours is captured by the factors. It is natural to assume that the common factors follow a vector autoregressive model,

here VAR(7):

$$\mathbf{F}_t = \Phi \mathbf{D}_t + \sum_{i=1}^7 \Theta_i \mathbf{F}_{t-i} + \zeta_t, \quad (5)$$

where Φ denotes a $N \times M$ matrix of deterministic coefficients, M is the number of deterministic variables, and Θ_i are $N \times N$ matrices of autoregressive parameters. To describe and forecast the idiosyncratic components ν_{kt} we use AR(7) models, independently for each k . As a result of this procedure we obtain the whole vector of 24 hourly price forecasts at once. For a recent treatment and an interesting application of VAR and structural VAR models in the context of electricity spot prices we refer to [11].

B. Combining individual forecasts and prediction intervals

Following a recent idea proposed in [15], we apply quantile regression to point forecasts of the individual models. More specifically, we use the individual point forecasts as independent variables and the corresponding observed spot price as the dependent variable in the quantile regression model [10]. This method – denoted in the text as **QR** – yields an interval forecast of the spot price, but does not use the PI of the individual methods. This is an important point, since as [18] remarks: *combining intervals directly will not in general give an interval with the correct probability*.

The averaging problem is given by:

$$F_{p_t}^{-1}(q|\hat{\mathbf{p}}_t) = \hat{\mathbf{p}}_t \mathbf{w}_t + F_{\varepsilon_t}^{-1}(q|\hat{\mathbf{p}}_t), \quad (6)$$

where ε_t is a model's residual at time t and F_{ε_t} is its cumulative distribution function. The weights are estimated by minimizing the loss function for a particular q -th quantile:

$$\min_{\mathbf{w}_t} \left[\sum_{\{t: p_t \geq \hat{\mathbf{p}}_t \mathbf{w}_t\}} q|p_t - \hat{\mathbf{p}}_t \mathbf{w}_t| + \sum_{\{t: p_t < \hat{\mathbf{p}}_t \mathbf{w}_t\}} (1-q)|p_t - \hat{\mathbf{p}}_t \mathbf{w}_t| \right] = \min_{\mathbf{w}_t} \left[\sum_t (q - \mathbb{1}_{p_t < \hat{\mathbf{p}}_t \mathbf{w}_t})(p_t - \hat{\mathbf{p}}_t \mathbf{w}_t) \right]. \quad (7)$$

Note that the proposed model is a natural extension of combining point forecasts. Indeed, if we decide to use the least absolute deviation (LAD) regression for estimating weights of the forecasts of the individual methods, as in [14], the problem is equivalent to the above QR model with $q = \frac{1}{2}$.

III. THE DATA

We consider hourly electricity system prices from the Nord Pool market in the period 8.8.2012–31.12.2013, see Figure 1. The first day (8.8.2012) marks the start of the calibration period for the 6 individual models (for model definitions see Section II.A). The first prediction of these models is made for 8.5.2013. Then the window is expanded by one day, the individual models are recalibrated and spot price predictions are made for 9.5.2013, etc. Initially eight weeks are used to obtain PI using the QR model, then the calibration period is expanded by one day, etc. Hence, the out-of-sample 26-week (i.e. 182-day) test period starts on 4.7.2013 and lasts until 31.12.2013. Compared to the empirical study in [15], here we

TABLE I
UNCONDITIONAL COVERAGE OF THE 50% AND 90% TWO-SIDED DAY-AHEAD PREDICTION INTERVALS (PI) BY THE ACTUAL SPOT PRICE FOR THE AR, SNAR AND QR MODELS. DESCRIPTIVE STATISTICS OF THE PI WIDTH ARE GIVEN IN THE CENTER AND BOTTOM ROWS.

PI	AR	SNAR	QR
<i>Unconditional coverage</i>			
50%	77.50	61.93	49.77
90%	97.53	96.41	89.33
<i>Mean (standard deviation) of length</i>			
50%	4.55 (1.34)	2.76 (0.61)	2.23 (0.81)
90%	11.14 (3.31)	9.33 (2.45)	6.78 (2.20)
<i>Median (IQR)</i>			
50%	4.38 (1.62)	2.70 (0.80)	2.15 (0.85)
90%	10.71 (3.99)	9.16 (2.85)	6.38 (2.37)

start with a shorter (8 vs. 32 weeks) calibration window for the QR averaging and as a result compare the PI over a longer (182 vs. 114 days) out-of-sample test period. Naturally, the datasets also differ (Nord Pool vs. PJM).

IV. THE RESULTS

We evaluate the quality of the prediction intervals (PI) by comparing both the unconditional and the conditional coverage. For each of the three considered models – AR, SNAR and QR – we first calculate the PI and determine the coverage of the 50% and 90% two sided day-ahead intervals by the actual spot price. If the model implied interval forecasts are accurate then the coverage will match the nominal values (i.e. 50% and 90%, respectively). For each model, 24×182 hourly values are determined and compared to the actual spot price. As can be seen in the upper part of Table I, SNAR and especially AR has a substantially higher coverage than nominal. On the other hand, QR's prediction intervals almost exactly match given probability levels. The obtained results, i.e. 49.77% and 89.33% for the 50% and 90% PI, respectively, are highly appealing. Also in terms of the width of the PI the QR model yields the best interval forecasts, see the center and bottom rows in Table I. On the other hand, the AR-implied interval forecasts are the worst, both in terms of unconditional coverage and interval width.

Next, we apply the approach of Christoffersen [7] to test the unconditional and conditional coverage. This model independent approach is designed to overcome the clustering effect. With the same notation as in eqn. (1), the idea behind the test is to operate on variables indicating whether the true price lies in the pre-constructed prediction interval. In other words, we analyze the sequence:

$$I_t = \begin{cases} 1 & p_t \in [L_{t|t-1}, U_{t|t-1}], \\ 0 & p_t \notin [L_{t|t-1}, U_{t|t-1}]. \end{cases}$$

The tests are carried out in the likelihood ratio (LR) framework. Three LR statistics are calculated: for unconditional coverage, independence and conditional coverage. The former two are distributed asymptotically as $\chi^2(1)$ and the latter as $\chi^2(2)$. If we condition on the first observation, then the

Fig. 1. Nord Pool hourly system spot price in the period 8.8.2012-31.12.2013. The out-of-sample half-year (26-week) test period for constructed prediction intervals (PI) is indicated by a rectangle. The vertical dotted line represents the beginning of the individual models' forecasts and the calibration window for the quantile regression averaging routine (four weeks prior to the test period).

conditional coverage LR test statistics is the sum of the other two. The conditional and unconditional coverage LR statistics are plotted in Figure 2. We follow the approach of [21] and conduct the tests separately for the 24 hourly time series. It would not make sense to compute the statistics jointly for all hours, since, by construction, the forecasts for consecutive hours are correlated – predictions for all 24 hours of the next day are made at the same time using the same information set.

According to Christoffersen's test statistics, the QR-based PI are the best of the three tested approaches. Except for only one hour, the QR-based PI are not rejected by the conditional coverage test at the 5% level for 50% PI. For the 90% probability level, the obtained PI pass the test for all 24 hours even at the 5% significance level. The test for unconditional coverage gives similar yet slightly worse results – here the 50% intervals are rejected only once (three times) at the 1% (5%) significance level. Again, QR-based 90% intervals pass the test for all 24 hours at the 5% level. On the other hand, the AR model turns out to be the worst of the analyzed methods. The null hypothesis of the conditional coverage test is rejected at the 1% level for all hours both for the 50% and 90% PI. The second benchmark model, SNAR, yields better PI than AR, but still much worse than QR.

Comparing with the results reported in [15], we can observe that while both benchmark models perform worse, quantile regression performs almost perfectly. The latter may be the result of a different spot price characteristics (for instance there are no spikes in the Nord Pool market in the QR calibration window, contrary to the PJM market) in the out-of-sample test period, a shorter calibration period of the QR model (at least for the first weeks in the test period) or a combination of both. This issue has to be further investigated.

V. CONCLUSIONS

This paper revisits the problem of computing prediction intervals (PI) for electricity spot prices. It can be considered

as an extension of the empirical studies [15], [20] and [21]. Here we use four out of six 'price-only' models (AR, TAR, SNAR and MRJD) evaluated in a point and interval forecasting exercise in [21]. The smoothed nonparametric autoregressive (SNAR) model, that was found in [21] to be the best model in terms of PI, is used here as a benchmark. Further, we complement this set of four models with a nonlinear AR (NAR) and a factor model (FM). In [20] the same six individual models are used as building blocks for three approaches to combining point (not interval) forecasts. Finally, the used in this study 'QR averaging' approach is proposed in [15]. It utilizes the concept of quantile regression (QR) and a pool of point forecasts of individual (i.e. not combined) models.

We find that QR-based PI are more accurate than those of the semiparametric SNAR and the standard AR models. Compared to the empirical study [15] utilizing less spiky (at least in the out-of-sample test period) PJM spot price data, here we start with a shorter (8 vs. 32 weeks) calibration window for the QR averaging and as a result compare the PI over a longer (182 vs. 114 days) out-of-sample test period. While the performance of both benchmark methods is worse than in [15], the QR-based intervals are more accurate here, both in terms of unconditional and conditional coverage. That said, we can recommend constructing prediction intervals with the 'QR averaging' method, though this should to be further confirmed using other datasets.

ACKNOWLEDGMENTS

We are grateful to Pierre Pinson, Eran Raviv and Stefan Trück for inspiring discussions on forecast averaging and electricity spot price forecasting. This paper has also benefited from conversations with the participants of the Applicable Semiparametrics Conference (2013), the 10th Conference 'European Energy Market' (EEM13), the Conference on Energy Finance (EF2013) and the Energy Finance Christmas Workshop (EFC13). This work was supported by funds from the

Fig. 2. The conditional coverage (left panels) and unconditional coverage (right panels) LR statistics for the PI obtained from the three considered forecasting models: AR, SNAR and QR. The solid (dashed) horizontal lines represent the 5% (1%) significance level of the appropriate χ^2 distribution. All test values exceeding 20 are set to 20.

National Science Centre (NCN, Poland) through grant no. 2011/01/B/HS4/01077.

REFERENCES

- [1] Amjady, N., Hemmati, M. (2006) Energy price forecasting, IEEE Power & Energy Magazine, March/April, 20-29.
- [2] Andalib, A., Atry, F. (2009) Multi-step ahead forecasts for electricity prices using NARX: A new approach, a critical analysis of one-step ahead forecasts, Energy Conversion and Management 50, 739-747.
- [3] Bai, J. (2003) Inferential theory for factor models of large dimensions, Econometrica 71(1), 135-171.
- [4] Bordignon, S., Bunn, D.W., Lisi, F., Nan, F. (2013) Combining day-ahead forecasts for British electricity prices, Energy Economics 35, 88-103.
- [5] Cao, R., Hart, J.D., Saavedra, A. (2003) Nonparametric maximum likelihood estimators for AR and MA time series, Journal of Statistical Computation and Simulation 73(5), 347-360.
- [6] Chatfield, C. (2000) *Time-Series Forecasting*, Chapman & Hall/CRC, Boca Raton, Florida.
- [7] Christoffersen, P. (1998) Evaluating interval forecasts, International Economic Review 39(4), 841-862.
- [8] Hong, T. (2014) Energy forecasting: Past, present, and future, Foresight, Winter, 43-48.
- [9] Hyndman, R. (2013) The difference between prediction intervals and confidence intervals, Hyndsight Blog (13 March 2013), <http://robjhyndman.com/hyndsight/intervals>.
- [10] Koenker, R. (2005) *Quantile Regression*, Cambridge University Press.
- [11] Maciejowska, K. (2014) Fundamental and speculative shocks, what drives electricity prices? IEEE Conference Proceedings, 11th International Conference on the European Energy Market (EEM'14). A working paper version is available from RePEc: <http://ideas.repec.org/p/wuu/wpaper/hsc1405.html>.
- [12] Maciejowska, K., Weron, R. (2013) Forecasting of daily electricity spot prices by incorporating intra-day relationships: Evidence from the UK power market. IEEE Conference Proceedings, 10th International Conference on the European Energy Market (EEM'13), art. no. 6607314 (DOI 10.1109/EEM.2013.6607314).
- [13] Misiorek, A., Trück, S., Weron, R. (2006) Point and interval forecasting of spot electricity prices: Linear vs. non-linear time series models. Studies in Nonlinear Dynamics and Econometrics 10(3), Article 2.
- [14] Nowotarski, J., Raviv, E., Trück, S., Weron, R. (2013) An empirical comparison of alternate schemes for combining electricity spot price forecasts. Working Paper, <http://ideas.repec.org/p/wuu/wpaper/hsc1307.html>.
- [15] Nowotarski, J., Weron, R. (2014) Computing electricity spot price prediction intervals using quantile regression and forecast averaging. Working Paper, <http://ideas.repec.org/p/wuu/wpaper/hsc1312.html>.
- [16] Raviv, E., Bouwman, K.E., van Dijk, D. (2013) Forecasting day-ahead electricity prices: Utilizing hourly prices. Tinbergen Institute Discussion Paper 13-068/III. Available at SSRN: <http://dx.doi.org/10.2139/ssrn.2266312>.
- [17] Timmermann, A.G. (2006) Forecast combinations. In: Elliott, G., Granger, C.W., and Timmermann, A., eds., *Handbook of Economic Forecasting*, Elsevier, 135-196.
- [18] Wallis, K.F. (2005) Combining density and interval forecasts: A modest proposal, Oxford Bulletin of Economics and Statistics 67, 983-994.
- [19] Weron, R. (2006) *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*, Wiley, Chichester.
- [20] Weron, R. (2014) A review of electricity price forecasting: The past, the present and the future, International Journal of Forecasting, forthcoming. A working paper version is available from RePEc: <http://ideas.repec.org/p/wuu/wpaper/hsc1402.html>.
- [21] Weron, R., Misiorek, A. (2008) Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models, International Journal of Forecasting 24, 744-763.

HSC Research Report Series 2014

For a complete list please visit <http://ideas.repec.org/s/wuu/wpaper.html>

- 01 *Diffusion and adoption of dynamic electricity tariffs: An agent-based modeling approach* by Anna Kowalska-Pyzalska, Katarzyna Maciejowska, Katarzyna Sznajd-Weron and Rafał Weron
- 02 *A review of electricity price forecasting: The past, the present and the future* by Rafał Weron
- 03 *Merging quantile regression with forecast averaging to obtain more accurate interval forecasts of Nord Pool spot prices* by Jakub Nowotarski and Rafał Weron