Probabilistic Forecasting of Electricity Prices using Kernel Regression

Grzegorz Dudek

Electrical Engineering Department
Czestochowa University of Technology
Czestochowa, Poland
dudek@el.pcz.czest.pl

***Abstract*—Electricity price forecasting has become crucial for energy companies due to its fundamental importance for decision making processes and operational management. Electricity price time series exhibit variable means, significant volatility and spikes, which places high demands on forecasting models. Moreover, in recent years researchers and practitioners have come to understand the limitations of point forecasts and require models to generate probabilistic forecasts. In contrast to point forecasts, the probabilistic forecasts takes the form of a predictive probability distribution over future quantities or events of interest. In the paper the probabilistic forecasting model based on Nadaraya-Watson estimator is proposed. The model generates the point forecasts as 24-component vectors representing day-ahead electricity prices. The probabilistic forecasts are calculated as quantiles based on the residual distribution for historical data forecasts. The performance of the proposed model is validated by testing on data from the Polish electricity market.**

***Index Terms*-- Electricity price forecasting, Kernel regression, Probabilistic forecasting, Quantile forecasts.**

I. Introduction

In electricity markets, each intra-day trading period is characterized by a rather distinct price profile reflecting the daily variation of demand, supply, costs and operational constraints. Also many shocks in fuel prices and institutional involvement such as carbon dioxide prices affect the price profile. Another important factor affecting electricity prices, which is gaining more and more importance in recent years, are smart grids development and renewable energy sources (since the beginning of the 21st century, global investments in renewable energy sources have been growing exponentially). Moreover, the electricity price drivers are expected to vary across time periods and markets, depending on local specificities, such as the generation mix, the degree of market power, and the market design [1]. All these factors make the forecasting of electricity prices a big challenge for energy supply companies for which price forecasts have become essential in decision making processes, such as the procurement strategies optimization.

The electricity price forecasting literature has focused on point forecasting, where the expected values of future prices are forecasted [2]. The flexible forecasting methods are searched which can reflect actual market outcomes in the dynamic environment as well as capture price spikes (often abrupt and unanticipated) and complex relationships between variables.

Probabilistic electricity price forecasting provides additional information on the variability and uncertainty of future price values. In the probabilistic approaches we are interested in the prediction intervals which contain the true values of future electricity prices with a specified probability. To obtain prediction intervals different methods are used, e.g. bootstrapping [3]. Another example is [4] where a set of individual forecasting models of different types generate point forecasts. Using principle component analysis these point forecasts are projected onto a set of principal components, which are treated as input variables in a quantile regression producing interval forecasts. More difficult task is to forecast the entire forecast density instead of the single prediction interval. In [5] a semi-parametric methodology for generating such densities is presented. It is based on a time-adaptive quantile regression model and a description of the distribution tails with exponential distributions. More literature examples on recent advances in electricity price forecasting and tutorial review of probabilistic forecasting can be found in [6]. The authors of this paper present winners of the Global Energy Forecasting Competition (GEFCom2014) price track and existing approaches for probabilistic electricity price forecasting based on statistical and computational intelligence tools such as density forecasts, bootstrapped prediction intervals, factor models and spike occurrence forecasting.

In this work a simple tool for probabilistic electricity price forecasting is proposed. The Nadaraya-Watson estimator is used for point forecasting hourly prices for the next day. Input data include only price profile from the preceding day. The model generates entire daily price profile at once (as 24-component vector) taking into account cross-dependencies between hourly prices. Then, a set of 99 quantiles as discretization of the full predictive densities of price distribution is calculated from the point forecasts and the empirical distribution of errors determined on the historical data.

II. DATA

Two datasets are used in this study. The first one comprises imbalance settlement hourly prices (p1) obtained from the Polish Balancing Market (source: https://www.pse.pl/). The second one comprises day-ahead marked hourly prices (p2) obtained from the Polish Power Exchange (source: https://www.tge.pl/). Both price time series from the period covering 2015 are shown in Fig. 1.

To detect components of the p1 and p2 time series, they are decomposed using STL method, i.e. Seasonal and Trend decomposition using Loess (local regression) [7]. Fig. 2 shows results. The grey bars to the right of each panel show the relative scales of the components. Each bar represents the same length. The smallest bars in the bottom panels show that the variation in the remainder components are greater compared to the variation in the seasonal and trend components. Thus, the random component in both price time series plays the largest role. It is larger in p1 time series than in p2 time series: average absolute value of the remainder is 17.73 for p1 vs. 10.19 for p2. As we can see from Fig. 2, p2 time series has greater seasonal component than p1. The scope of this component is around 90 for p2 and 58 for p1. It is worth nothing also flexible trends expressing weekly seasonality.

The features of the price time series described above: dominant random component, multiple seasonal patterns, flexible trend, and also spikes make them hard to predict and put high demands on the forecasting model.

III. FORECASTING MODEL

The forecasting model consist of the Nadaraya-Watson model (N-WM) which produces point forecasts and the procedure of extending point forecasts into quantile forecasts. N-WM generates forecasts for the next day $\mathbf{y}$ (24 price values at once for our hourly granularity data) on the basis of the daily profile of the previous day $\mathbf{x}$. Thus, the input data when forecasts for the day k are generated is a vector $\mathbf{x}_k = [p_{k-1,1}\ p_{k-1,2}\ \dots\ p_{k-1,24}]^T \in X = \mathbb{R}^{24}$, and the output data is a vector $\mathbf{y}_k = [p_{k,1}\ p_{k,2}\ \dots\ p_{k,24}]^T \in Y = \mathbb{R}^{24}$, where $p_{k,j}$ is an electricity price at hour j of the day k. No data processing has been applied.

A. Nadaraya-Watson Estimator

Nadaraya-Watson estimator belongs to the class of nonparametric kernel regression. The model estimates a nonlinear relationship between random variables as a locally weighted average, using kernels as weighting functions. When both input and output data are vectors and kernels are expressed using multidimensional product kernels, N-WM is defined as follows [8]:

$$m(\mathbf{x}) = \frac{\sum_{k=1}^{N} \prod_{j=1}^{n} K\left(\frac{x_j - x_{k,j}}{h_j}\right) \mathbf{y}_k}{\sum_{k=1}^{N} \prod_{j=1}^{n} K\left(\frac{x_j - x_{k,j}}{h_j}\right)} \qquad (2)$$

where N is a number of construction samples, n is a number of components in $\mathbf{x}$, $K(.)$ is a kernel function and h is its bandwidth.

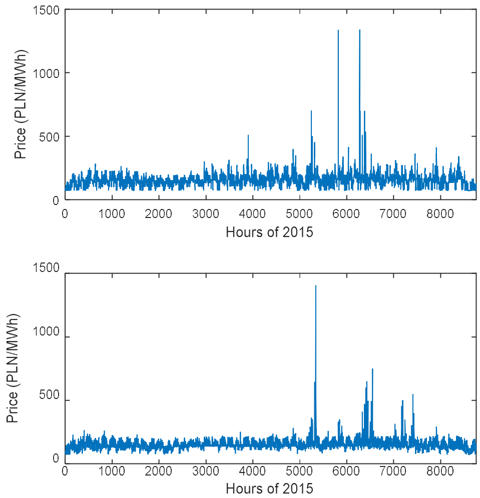

Figure 1. Imbalance settlement hourly prices (upper chart) and day-ahead marked prices (bottom chart).

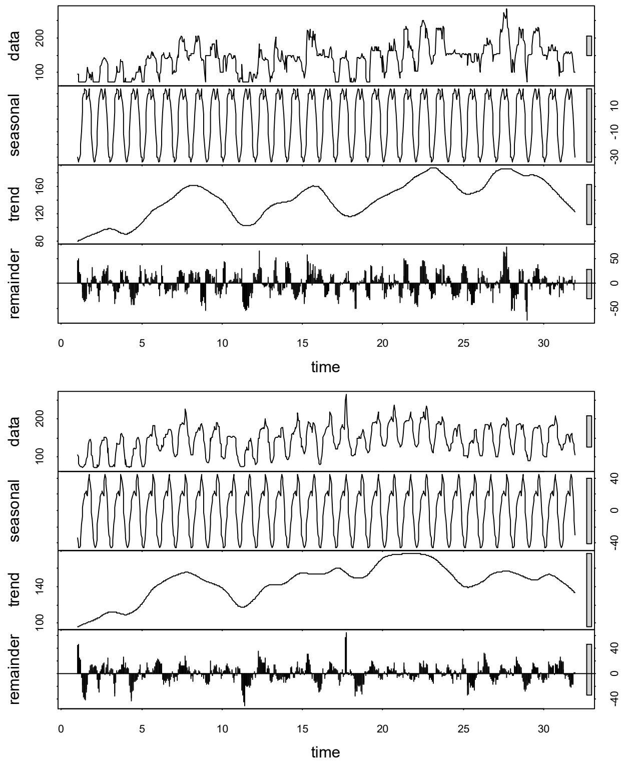

Figure 2. STL decomposition of imbalance settlement price (p1) time series (upper chart) and day-ahead marked price (p2) time series (bottom chart).

Construction samples are pairs of profiles ($\mathbf{x}_k$, $\mathbf{y}_k$) from history representing the same days of the week as the current input profile $\mathbf{x}$ and the forecasted profile $\mathbf{y}$, respectively. For example, when we forecast price profile for Tuesday, the pairs of construction profiles are selected from history, where x-profiles represent Mondays and y-profiles represent Tuesdays. This decomposition of the forecasting problem into separate days of the week results from the existence of the weekly season in the price time series and different profile shapes for working and weekend days as well.

When the most popular Gaussian kernel is used, N-WM takes the form:

$$m(\mathbf{x}) = \frac{\sum_{k=1}^{N} \exp\left(-\sum_{j=1}^{n} \frac{(x_j - x_{jk,j})^2}{2h_j^2}\right) \mathbf{y}_k}{\sum_{k=1}^{N} \exp\left(-\sum_{j=1}^{n} \frac{(x_t - x_{k,j})^2}{2h_j^2}\right)} \tag{3}$$

This equation expresses a linear combination of vectors $\mathbf{y}_k$ weighted by the normalized kernels, which nonlinearly map the distance between patterns $\mathbf{x}$ and $\mathbf{x}_k$. The distance is parameterized by the bandwidths. Thus, the bandwidth h_j controls the share of the j-th component of $\mathbf{x}$ in the distance. The bandwidth values decide about the bias-variance tradeoff of the estimator. When they are too small the estimator tends to undersmoothing, and when they are too large it tends to oversmoothing. Thus the selection of the bandwidth values is a crucial issue. To find them we calculate their initial values using formula proposed by Scott [9] for the normal product density estimators:

$$h_j^S = \hat{\sigma}_j N^{-\frac{1}{n+4}} \tag{4}$$

where $\hat{\sigma}_j$ is the standard deviation of the j-th component of $\mathbf{x}$ estimated from the sample.

The bandwidths are searched using iterative process where vectors $\mathbf{h}$ are generated one by one according to:

$$\mathbf{h} = a\mathbf{h}^S \tag{5}$$

where $a = a_{\min}$, $a_{\min} + \Delta a$, $a_{\min} + 2\Delta a$, ..., $a_{\max}$.

The optimal a value is selected taking into account the model performance on historical data (see Section IV).

B. Quantile forecasts

The N-WM generates point forecasts. To get probabilistic forecasts in quantiles, the model is used to forecast historical data. The residuals distribution for historical data is determined and its quantiles are calculated. We assume that for new forecasting task the error distribution will not change and will be similar to the error distribution for historical period. In the experimental part of the work we use historical period covering 2014 for residual quantiles estimation when constructing the forecasting model for data from 2015. The histograms of residuals for 2014 in Fig. 3 are shown. As we can see from this figure for p2 much bigger residuals are observed than for p1.

The quantiles of residuals for probabilities p = 0.01, 0.02, ..., 0.99 are calculated and added to the point forecasts. This gives quantiles of the forecast distribution. The point forecast is the median of this distribution. A prediction interval is a range of specified coverage probability under that distribution. For example 95% prediction interval is defined by the 0.025 and 0.975 quantiles of the forecast distribution.

Fig. 4 shows an example of the forecast distribution in quantiles. The point forecast generated as 24-component vector by N-WM is expanded to 99 quantiles. The bottom line represents 0.01 quantile and the top line represent 0.99 quantile. The range between them is the 98% prediction interval.

The proposed method of generation of the quantile forecasts is different from that proposed in [10], where the standard deviation of errors for historical data was calculated. Having the point forecast and standard deviation, the inverse cumulative distribution function was used to expand the forecasts to quantiles. The forecast distribution obtained in this way is symmetrical about the point forecast. The proposed approach gives more realistic asymmetric distribution. The quantiles for probabilities close to one are much larger than for the normal distribution due to the occurrence of many spikes (see Fig. 1).

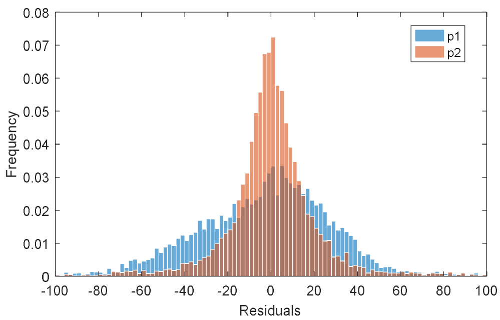

Figure 3. Histograms of residuals for 2014.

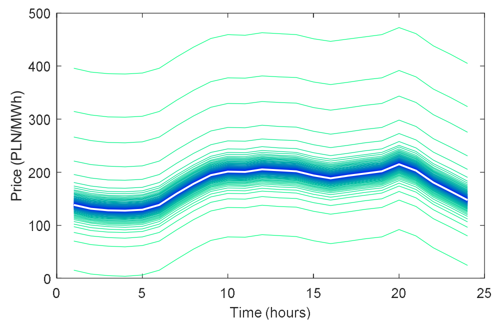

Figure 4. An example of the forecast distribution in quantiles.

IV. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed approach using N-WM, the method was examined on two price time series described in Section II:

- p1: time series of the imbalance settlement hourly prices obtained from the Polish Balancing Market over the period 2012–2015,
- p2: time series of the day-ahead marked hourly prices obtained from the Polish Power Exchange over the period 2012–2015.

Data form 2015 are treated as the test data. The forecasting model is used separately for each day of this period. The forecasting task for the model is to forecast electricity prices for the next day. For each forecasting task (test sample) the construction samples are selected individually from the period from January 1st, 2012 to the day preceding the forecasted day.

To select the best values of the parameters h_j in (3) we use iterative procedure described in Section III.A, changing a from 0.25 to 4.00 with a step of 0.01. For each a value the model is tested on each day of 2014 using construction samples from history. An average forecast error (MAPE) is used as a model performance measure indicating the optimal value of a. Using this value the model is applied for data from 2015 to get point forecasts.

Residual quantiles are estimated on data from 2014 using optimal N-WM. The cumulative distribution functions corresponding to the residual quantiles in Fig. 5 are shown. From this figure we can read the correction Δp for the point forecast to get its quantile. For example, the 0.2 quantile is about $\hat{p}_1 - 28$ for p_1 time series and $\hat{p}_2 - 11$ for p_2 time series, whilst 0.8 quantile is about $\hat{p}_1 + 25$ for p_1 and $\hat{p}_2 + 14$ for p_2. In the proposed simplified approach, for getting the given quantile forecast the same corrections Δp are used for each point forecast.

Fig. 6 shows forecast for the first week of 2015. Prediction intervals were determined on the basis of quantiles. In this figure point forecasts for ARIMA and exponential smoothing (EST) are also shown. The ARIMA and EST parameters were estimated in the stepwise procedures for traversing the model spaces implemented in the `forecast` package for the R environment for statistical computing [11]. These automatic procedures return the optimal models with the lowest Akaike information criterion value. The ARIMA and EST models were built for each forecasting task (each day of 2015) independently using 3-week price time series fragments immediately preceding the forecasted day.

The average errors of point forecasts: mean absolute percentage error (MAPE) and root mean squared error (RSME) in Tables I and II are shown. Interquartile ranges (IQR) of the percentage errors are also shown in these tables as a measure of error dispersion (IQR is usually used instead of the standard deviation as a more robust dispersion measure when analyzing asymmetric distributions like in our case). For comparison forecast results for the naïve model are also shown in these tables. This approach simply takes the forecasted electricity price of day k exactly equivalent to the electricity price of day $k - 7$.

As can be seen from Tables I and II, for both price time series the best results were achieved for the proposed N-WM. The second best results were provided by EST. Note that the naïve model was only a little worse than the more sophisticated models. This confirms the difficult nature of the analyzed price time series. As expected, the accuracy of forecasting the p1 time series is much worse than the accuracy for p2 time series. More detailed results, the histograms of residuals for 2015 in Fig. 7 are shown. For both datasets negative distribution skewness for N-WM is observed. It means that the forecasted prices are on average higher than actual ones.

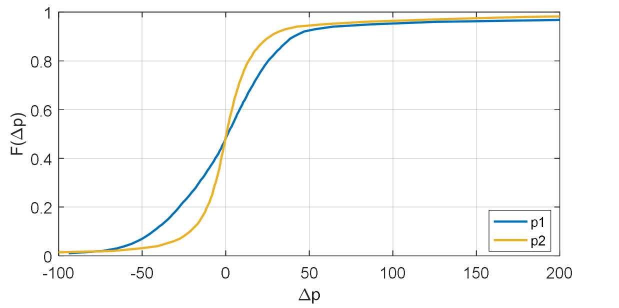

Figure 5. Cumulative distribution functions for calculating the quantiles of electricity prices determined using N-WM.

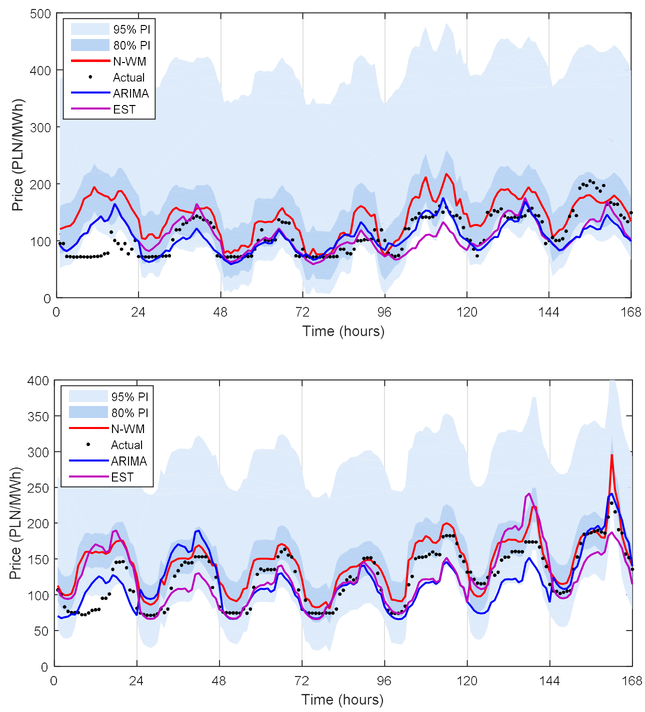

Figure 6. Forecasts for the first week of 2015: settlement hourly prices (upper chart) and day-ahead marked prices (bottom chart).

For evaluating the full predictive densities composed by the quantile forecasts the pinball loss is often used [6]:

$$L_{\tau}(\Delta) = \begin{cases} \tau\Delta, & \text{if } \Delta \geq 0 \\ (\tau - 1)\Delta, & \text{if } \Delta < 0 \end{cases} \tag{6}$$

where $\Delta = p - \hat{q}_{\tau}$, $\hat{q}_{\tau}$ is the price forecast at the τ-th quantile and p is the actual price.

Average values of the pinball loss over all target quantiles for data from 2015 were: 10.45 for p1 and 5.97 for p2.

TABLE I. FORECAST RESULTS FOR P1 DATA

Model	MAPE	IQR	RMSE
N-WM	20.32	30.47	45.57
ARIMA	24.80	37.68	53.58
EST	21.02	30.72	49.10
Naive	26.18	35.72	64.11

TABLE II. FORECAST RESULTS FOR P2 DATA

Model	MAPE	IQR	RMSE
N-WM	9.38	13.22	34.17
ARIMA	13.52	20.01	44.04
EST	10.36	15.38	40.39
Naive	12.37	16.43	54.74

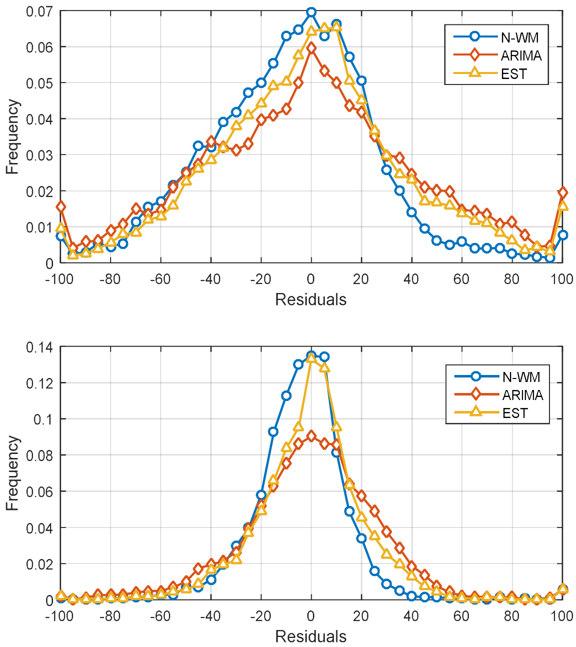

Figure 7. Histograms of residuals for settlement hourly prices (upper chart) and day-ahead marked prices (bottom chart).

V. CONCLUSIONS

The aim of this work was to develop a simple model based on Nadaraya-Watson estimator for probabilistic forecasting hourly electricity prices at day-ahead markets. The model generates the point forecasts as a daily price profile using the most similar profiles to the current profile from the history. The construction profiles forming the forecasted profile by weighted averaging, are selected from entire available data period. Note the difference comparing to ARIMA and exponential smoothing, where the forecasts are constructed on the basis of a time series fragment preceding the forecasted day (limited to 3 weeks in our case).

The proposed method constructs the full predictive densities composed by 0.01,0.02, …, 0.99 quantiles, which are estimated using empirical distribution of residuals determined on historical data. The roughness of the quantile estimation caused by the adoption of the same residual distribution for the entire forecast period will be improved in future studies. It is planned to estimate the distribution of residuals individually for each forecasting task. This should bring more accurate probabilistic forecasts.

REFERENCES

[1] N.V. Karakatsani, D.W. Bunn, "Forecasting electricity prices: The impact of fundamentals and time-varying coefficients," *International Journal of Forecasting*, vol. 24(4), pp. 764-785, 2008.

[2] R. Weron, "Electricity price forecasting: A review of the state-of-the-art with a look into the future," *International Journal of Forecasting*, vol. 30(4), pp. 1030-1081, 2014.

[3] X. Chen, Z.Y. Dong, K. Meng, Y. Xu, K.P. Wong, and H.W. Ngan, "Electricity price forecasting with extreme learning machine and bootstrapping," *IEEE Transactions on Power Systems*, vol. 27(4), pp. 2055–2062, 2012.

[4] K. Maciejowska, J. Nowotarski, R. Weron, "Probabilistic forecasting of electricity spot prices using Factor Quantile Regression Averaging," *International Journal of Forecasting*, vol. 32(3), pp. 957–965, 2016.

[5] T. Jonsson, P. Pinson, H. Madsen, and H.A. Nielsen, "Predictive densities for day-ahead electricity prices using time-adaptive quantile regression," *Energies*, vol. 7(9), pp. 5523–5547, 2014.

[6] J. Nowotarski, and R. Weron, "Recent advances in electricity price forecasting: A review of probabilistic forecasting," *Renewable and Sustainable Energy Reviews*, vol. 81, part 1, pp. 1548-1568, 2018.

[7] R.B. Cleveland, W.S. Cleveland, J.E. McRae and I.J. Terpenning, "STL: A seasonal-trend decomposition procedure based on loess," *Journal of Official Statistics*, vol. 6(1), pp. 3–73, 1990.

[8] G. Dudek, "Pattern similarity-based methods for short-term load forecasting – Part 2: Models," *Applied Soft Computing*, vol. 36, pp. 422-441, 2015.

[9] D.W. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley, 1992.

[10] G. Dudek, "Multilayer perceptron for GEFCom2014 probabilistic electricity price forecasting," *International Journal of Forecasting*, vol. 32, pp. 1057-1060, 2016.

[11] R.J. Hyndman, Y. Khandakar, "Automatic time series forecasting: The forecast package for R," *Journal of Statistical Software*, vol. 27(3), pp. 1–22, 2008.