Forecasting upward and downward jumps in Nord Pool electricity prices by means of the generalised ordered logistic regression

Jadwiga Kostrzewska[1], Maciej Kostrzewski[2]

Abstract

The paper analyses the impact of selected variables, including an electricity consumption forecast, a wind power forecast, a water level in hydroelectric power plants, power plant outages, and seasonality, on the probability of upward or downward electricity price jumps. Using the rolling window scheme, we detect jumps by means of the quantile analysis, Tukey criterion, and a method based on the adjusted boxplot. Next, we apply the generalised ordered logistic regression model in order to forecast jump occurrences. The analysis is conducted on the basis of hourly electricity prices from the day-ahead Nord Pool market.

Keywords: electricity prices, forecasting jumps, generalised ordered logistic regression model

JEL Classification: C53, C63, Q47

1. Introduction

Electricity prices are extremely volatile, and their volatility is strongly linked with still existing problems with electricity storage which accompany the requirement to constantly balance production and consumption. This volatility is manifested by frequent occurrences of sudden price changes, called jumps. A precise estimation of the probability of an upward and downward jump, as well as no jump state is crucial in forecasting electricity prices. Forecasts of prices in consecutive hours of a day are important from the risk management perspective and may influence the strategy adopted by producers and consumers of electricity.

The literature on electricity price forecasting is abundant (see Weron, 2014 for an overview). Furthermore, a number of papers are devoted to modelling and/or forecasting occurrences of price jumps (see e.g. Christensen *et al.*, 2012; Eichler *et al.*, 2012, 2014; Hellström *et al.*, 2012; Janczura *et al.*, 2013; Kostrzewski, 2012; Kostrzewski, 2019; Kostrzewska and Kostrzewski, 2018). For example, in order to forecast upward jumps/spikes of electricity prices (0-1 dependent variable), Christensen, Hurn and Lindsay (2012), Eichler

[1] Corresponding author: Cracow University of Economics, Department of Statistics, Rakowicka 27, 31-510 Cracow, Poland, jadwiga.kostrzewska@uek.krakow.pl.

[2] Cracow University of Economics, Department of Econometrics and Operations Research, Rakowicka 27, 31-510 Cracow, Poland, maciej.kostrzewski@uek.krakow.pl.

et al. (2012, 2014) employ the dynamic logit models, the ACH models and their modifications, while Kostrzewska and Kostrzewski (2018) use the logit model, and Hellström *et al.* (2012) consider modelling positive (upward) and negative (downward) jumps of daily electricity prices by means of the ordered probit model. Our contribution to the existing literature lies in employing the generalized ordered logistic regression model in order to forecast the probability of three ordered states: a downward jump, no jump, and an upward jump of electricity prices in hourly resolution. The model allows for taking into account a different impact of explanatory variables on the probability of an upward jump and a different impact on the probability of a downward jump.

The paper aims at forecasting an upward jump, a downward jump, and no jump occurrence of electricity prices and at analysing factors which influence the probability of jump occurrences. We consider electricity prices on Nord Pool, which is an example of the market with an energy mix based mostly on renewable energy. The rest of the paper is organized as follows. Section 2 introduces data and methods used in the study, Section 3 discusses empirical results, and Conclusions Section ends the paper.

2. Data and methods

The paper analyses electricity prices from the Nord Pool market in the period between December 29, 2014 and July 2, 2017. We apply a rolling window scheme with a window length equal to 364 days, i.e. containing 24x364 hourly observations. We estimate each model based on 364 days and then, due to the nature of the day-ahead market, we forecast electricity prices for 24 hours of a next day. The first in-sample period spans from December 29, 2014 to December 27, 2015. In the study, we estimate 553 models and obtain 24-hourly forecasts for each of the models.

Before applying a jump detection technique, we pre-process the data by removing the long- and short-term trend and seasonal component by means of the Hodrick-Prescott filter (see Weron and Zator, 2015) and a median filter moving over 168-hours (one week). These procedures are conducted for each window separately. The basic problem is how to define a jump. We employ three outlier detection techniques to identify jumps in electricity prices. The methods are not equivalent and do not detect the same jumps. Jumps detected under each of the methods are called the ‘observed’ ones and determine the categories of the dependent variable in the generalised ordered logistic regression model.

The method based on the quantiles (Janczura *et al.*, 2013; Kostrzewska *et al.*, 2016; Kostrzewska and Kostrzewski, 2018) marks 2.5% of the lowest and 2.5% of the highest values as downward and upward jumps, respectively. Let Q_1 and Q_3 be the lower and upper quartiles, and $IQR = Q_3 - Q_1$ be the interquartile range. Within the method based on Tukey criterion (a standard boxplot), the values outside the range $[Q_1 - 1.5 \cdot IQR; Q_3 + 1.5 \cdot IQR]$ are marked as downward and upward jumps (see Tukey, 1977; Pawełek *et al.*, 2015; Kostrzewska *et al.*, 2016). However, electricity prices have a skewed distribution (see Fig. 1), thus we also employ the method based on an adjusted box-plot (Hubert and Vandervieren, 2008). Within this method, the values outside the range $[Q_1 - 1.5 \cdot e^{-4MC} \cdot IQR; Q_3 + 1.5 \cdot e^{3MC} \cdot IQR]$ are marked as downward and upward jumps, respectively. The robust measure of the skewness *medcouple* (*MC*) was introduced by Brys *et al.* (2004) (see also Hubert and Vandervieren, 2008) and is defined as $MC = \underset{x_i \le Q_2 \le x_j}{\text{median}} \frac{(x_j - Q_2) - (Q_2 - x_i)}{x_j - x_i}$, where $x_1, \dots, x_n$ – a data set, Q_2 – a sample median of the set $x_1, \dots, x_n$, i.e. *MC* is calculated as a median of values $\frac{(x_j - Q_2) - (Q_2 - x_i)}{x_j - x_i}$ calculated for all $\{(i, j): x_i \le Q_2 \le x_j\}$. In the paper, we use the abbreviations *ADJ*, *QUA* and *TUK* for the jump detection methods based on the adjusted boxplot, the quantiles and Tukey's criterion, respectively.

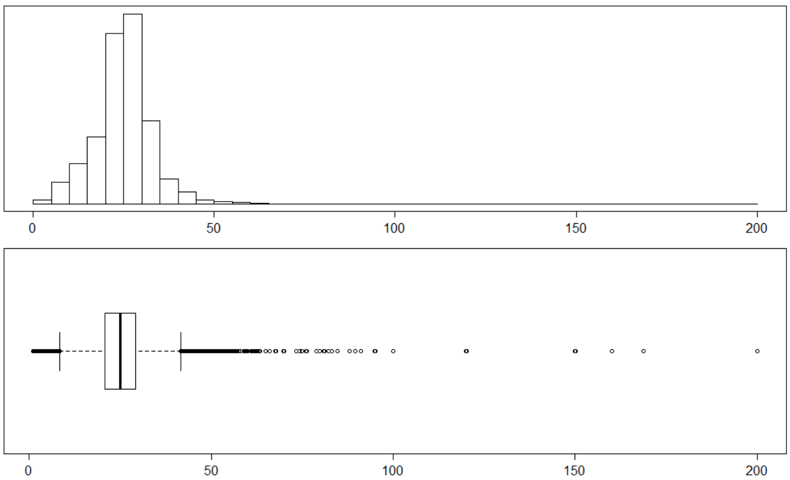

Figure 1. Histogram (top) and boxplot (bottom) of the hourly day-ahead electricity prices in the period between December 29, 2014 and July 2, 2017

The aim of our study is to model and forecast upward and downward jumps of electricity prices, as well as no jump state. In the study, we employ the generalised ordered logistic regression model to deal with the probability estimation of an occurrence of ordered outcomes. The variable describing jumps takes three values:

$$y_t = \begin{cases} -1 & \text{in case of a downward jump,} \\ 0 & \text{in case of no jump,} \\ 1 & \text{in case of an upward jump.} \end{cases} \quad (1)$$

The generalised ordered logistic regression model for M categories of the dependent variable takes the form (Williams 2006):

$$P(y_t > j|X_t) = \frac{\exp(\alpha_j + \beta_j X_t)}{1 + \exp(\alpha_j + \beta_j X_t)} \text{ for } j = 1, \dots, M-1. \quad (2)$$

where α_j are thresholds, $X = (X_1, \dots, X_k)$ – k explanatory variables, and $\beta_j = (\beta_{j1}, \dots, \beta_{jk})$ is a vector of parameters (without a constant term). In our study, the model consists of two equations: the first one describes the probability $P(y_t > -1|X_t)$, while the second one describes the probability $P(y_t > 0|X_t)$. In each of these equations different parameters β_j are allowed, thus different impact of the explanatory variables on the probabilities under consideration can be taken into account. We consider three methods of jump detection, thus we have three generalised ordered logistic regression models. In the next section, we compare the results obtained in these three cases.

In the study, we apply the rolling window scheme. For each window with a length of 24x364 hours, we pre-process the electricity prices by removing the long- and short-term trend and seasonal component, then we define jumps by means of one of three jump detection methods. Next, we apply the generalised ordered logistic regression model to describe jumps depending on a set of explanatory variables and to forecast a downward jump, no jump, or an upward jump occurrence for 24 hours ahead. For each window we assess (in-sample) the goodness of fit of models by means of McFadden's pseudo-R^2 and obtain the out-of-sample classification accuracy as a proportion of correctly forecasted occurrences: a downward jump, no jump, an upward jump.

3. Empirical results

As mentioned above, we apply a rolling window scheme, thus we obtain the generalised ordered logistic regression models for 553 in-sample periods separately for each jump detection methods *ADJ*, *QUA* and *TUK*. As a preliminary set of explanatory variables, we consider:

- electricity prices lagged by 24 hours (*Lagprice*),
- consumption forecasts (*Consumption*),
- wind power generation forecasts (*Wind*),
- a level of water in hydroelectric power plants (*Water*),
- information on outages of power plants lagged by 48 hours (*Outages*),

- a dummy for wintertime (*Winter*),
- dummies for days of a week (*Mon-Sun* without Wednesday),
- a dummy for peak hours, i.e. #8 – #20 (*Peak*).

Table 1. The frequency with which a variable has a statistically significant impact on the dependent variable within 553 models obtained for different jump detection methods (*ADJ*, *QUA*, *TUK*)

Explanatory Variable	Sign	*ADJ* $P(Y > -1)$	*ADJ* $P(Y > 0)$	*QUA* $P(Y > -1)$	*QUA* $P(Y > 0)$	*TUK* $P(Y > -1)$	*TUK* $P(Y > 0)$
Lagprice	+	100.0%	98.4%	100.0%	98.0%	100.0%	99.3%
	–	0.0%	1.6%	0.0%	2.0%	0.0%	0.7%
Consumption	+	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
	–	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Wind	+	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	–	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Water	+	31.4%	17.3%	40.3%	32.3%	29.8%	9.4%
	–	48.4%	62.5%	37.0%	44.9%	51.8%	72.2%
Outages	+	6.7%	0.0%	14.8%	0.0%	9.0%	0.0%
	–	73.8%	80.5%	64.1%	79.4%	77.1%	86.1%
Winter	+	83.9%	22.0%	85.4%	59.7%	92.2%	27.1%
	–	4.9%	66.8%	4.9%	31.0%	4.7%	69.9%
Mon	+	13.9%	13.7%	60.5%	68.1%	39.5%	45.3%
	–	1.1%	1.3%	5.1%	0.7%	6.7%	0.9%
Tue	+	28.2%	19.9%	25.8%	23.3%	27.1%	31.4%
	–	34.3%	42.6%	12.1%	14.8%	25.1%	20.8%
Thu	+	1.8%	6.1%	7.4%	50.0%	20.4%	30.0%
	–	20.0%	15.7%	56.9%	21.3%	12.8%	3.2%
Fri	+	34.3%	0.0%	22.2%	1.4%	32.7%	0.9%
	–	57.8%	92.1%	38.1%	60.6%	42.6%	74.4%
Sat	+	0.0%	0.0%	0.2%	0.0%	0.2%	0.0%
	–	97.3%	97.3%	95.5%	97.5%	69.1%	69.3%
Sun	+	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	–	100.0%	100.0%	98.4%	99.8%	99.1%	99.1%
Peak	+	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
	–	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Note: A positive/negative impact of a variable is denoted by a sign '+'/'–'. The green colour means the frequency is higher than 75%, the red colour means the frequency is lower than 25%, and the yellow colour means the frequency is between 25% and 75%.

In each model, we use the backward stepwise method to determine a set of explanatory variables that remain in the model. Thus, the set of these variables may differ in each of the 553 estimated models. Table 1 reports the frequencies of variables in 553 sets of explanatory variables with an indication of a direction of its impact (a positive or negative sign). If the frequencies of the positive and negative impact of a variable sum up to 100%, this variable has a statistically significant impact in all 553 models. Otherwise, as in case of the *Water* variable, the variable has a statistically insignificant impact in some of the models under consideration.

The variables describing the electricity prices lagged by 24 hours, the forecasted consumption, and the dummy variable describing peak hours (#8 – #20) have a statistically significant positive impact on the probabilities $P(Y > -1)$ and $P(Y > 0)$ in almost all data sets (98% – 100%) and the jump detection methods. We can conclude that both higher prices 24 hours before and the forecasted consumption increase the probability of an upward jump or no jump occurrence. On the other hand, the forecasted wind power (100%), dummy for Sundays (98.4% – 100%) and Saturdays (to a lesser extent) almost always have a statistically significant negative impact on these probabilities. The impact directions of some other variables are not clear. The level of water in hydroelectric power plants has a positive, negative or insignificant impact on the probabilities in some of 553 models. In case of outages, a negative impact on both probabilities is predominant (64.1% – 86.1%), which is rather surprising. Lower frequencies of dummy variables representing Mondays, Tuesdays and Thursdays may indicate that the probability $P(Y > -1)$ is almost the same on these days as on Wednesdays. We can conclude the same for the probability $P(Y > 0)$. Generally, the frequencies of the explanatory variables are quite high with the exception of working days.

Table 2. The percentage of correctly forecasted: downward jump, no jump, upward jump occurrences in the out-of-sample period (on the left), and the mean, maximum and minimum of pseudo-R^2 measure calculated on the basis on 553 in-sample periods (on the right)

	Percentage of correctly forecasted occurrences (out-of-sample)				Pseudo-R^2 (in-sample)		
Method	**Downward jump**	**No jump**	**Upward jump**	**All**	**Mean**	**Min**	**Max**
ADJ	57.5%	98.2%	51.3%	95.0%	0.623	0.541	0.726
QUA	57.6%	98.5%	43.5%	95.6%	0.648	0.548	0.754
TUK	69.7%	98.4%	51.0%	95.6%	0.619	0.509	0.720

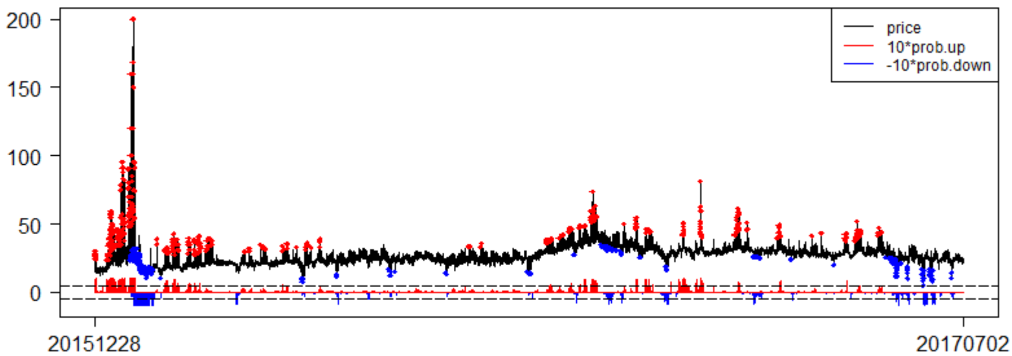

Figure 2. The electricity prices in the out-of-sample period along with upward (red dots) and downward (blue dots) jumps detected (i.e. 'observed') by means of *ADJ* method, forecasted probabilities of upward (red bars) and downward (blue bars) jump occurrences

Note: The probabilities outside the area determined by dashed horizontal lines correspond to the values higher than 0.5.

Based on each of 553 models, we forecast a downward jump if the probability $P(Y = -1) > 0.5$, an upward jump if the probability $P(Y = 1) > 0.5$, and no jump otherwise. The left side of Table 2 reports a percentage of correctly forecasted states, i.e. downward jumps, no jump, upward jumps. The percentage of correctly forecasted no jump states is very high and exceeds 98%, regardless of the method of jump detection used. The percentage of correctly forecasted downward jumps (57.5% – 69.7%) is higher than upward ones (43.5% – 51.3%). These percentages are not too high, however, total accuracy of the generalised ordered logistic regression models i.e. the percentages of correctly forecasted states are satisfactorily high (slightly over 95%). The right side of Table 2 presents the values of pseudo-R^2 measure calculated for *ADJ*, *QUA* and *TUK* methods. The values of this measure range from about 0.5 up to almost 0.76, which is a good outcome. The results are similar regardless which method is used to detect jumps.

Fig. 2 presents the electricity prices in the out-of-sample period along with upward (red dots) and downward (blue dots) jumps detected (i.e. 'observed') by means of *ADJ* method. The forecasted probabilities of downward (blue bars) and upward (red bars) jump occurrences are presented at the bottom of the figure. For the sake of clarity, the probabilities of downward jumps are presented as negative values. The probabilities outside the area bounded by dashed horizontal lines exceed 0.5 – in such a case an upward/downward jump is forecasted. It is easy to see that the red (blue) dots in the time series of prices and red (blue) bars higher than 0.5 appear at similar moments. The plots for the other jump detection methods are similar to the one presented above.

4. Conclusions

The paper analyses and forecasts upward and downward jump occurrences in electricity prices. Using the rolling window scheme, we detect jumps by means of one of the considered methods based on: the adjusted boxplot, the quantiles, and Tukey's criterion. We employ the generalised ordered logistic regression model which allows different impacts of explanatory variables on the probabilities of upward and downward jumps. By contrast, the ordered logistic regression model imposes the same impact of explanatory variables in each equation (parallel regressions assumption). On the other hand, the logistic regression model used in the literature can only handle the upward jump and no jump occurrences.

In our study, regardless of the method used to detect jumps, the values of pseudo-R^2 measure are relatively high. We notice a meaningful impact of the explanatory variables on the probabilities of an upward, downward or no jump occurrence. In particular, we observe that no jump or upward jump probabilities are increased by higher electricity prices appearing 24 hours earlier, higher forecasted consumption, and during peak hours (#8 – #20), and they are decreased by the higher forecasted wind power and on Saturdays and Sundays. We assess the accuracy of the generalised ordered logistic regression model in forecasting downward jump, no jump, or upward jump occurrences by means of the proportion of correctly forecasted states. The total accuracy of the models exceeds 95%.

The results indicate that the generalised ordered logistic regression model is a promising tool for forecasting the probability of upward and downward jump occurrences. The important conclusion is that there is no clear distinction in the goodness-of-fit and the accuracy of the forecasts obtained under the generalised ordered logistic regression model for jump detection methods considered in the study. The effectiveness is high in each case.

Acknowledgements

This work was supported by the National Science Centre, Poland through grant 2016/23/B/HS4/03018.

References

Brys, G., Hubert, M., Struyf, A. (2004). A Robust Measure of Skewness. *Journal of Computational and Graphical Statistics*, *13*(4), 996-1017.

Christensen, T.M., Hurn, A.S., Lindsay, K.A. (2012). Forecasting Spikes in Electricity prices. *International Journal of Forecasting*, *28*(2), 400-411.

Eichler, M., Grothe, O., Manner, H., Türk, D.D.T. (2012). Modeling Spike Occurrences in Electricity Spot Prices for Forecasting, *METEOR*, RM/12/029.

Eichler, M., Grothe, O., Manner, H., Türk, D.D.T. (2014). Models for Short-Term Forecasting of Spike Occurrences in Australian Electricity Markets: a Comparative Study. *Journal of Energy Markets*, *7*, 55-81.

Hellström, J., Lundgren, J., Yu, H. (2012). Why do Electricity Prices Jump? Empirical Evidence from the Nordic Electricity Market. *Energy Economics*, *34*(6), 1774-1781.

Hubert, M., Vandervieren, E. (2008). An Adjusted Boxplot for Skewed Distributions. *Computational Statistics & Data Analysis*, *52*(12), 5186-5201.

Janczura, J., Trück, S., Weron, R., Wolff, R.C. (2013). Identifying Spikes and Seasonal Components in Electricity Spot Price Data: A Guide to Robust Modeling. *Energy Economics*, *38*, 96-110.

Kostrzewska, J., Kostrzewski, M. (2018). The Logistic Regression in Predicting Spike Occurrences in Electricity Prices. In *Proceedings of the 12th Professor Aleksander Zeliaś International Conference on Modelling and Forecasting of Socio-Economic Phenomena. Zakopane: Foundation Cracow Univ Economics*.

Kostrzewska, J., Kostrzewski, M., Pawelek, B., Galuszka, K. (2016). The Classical and Bayesian Logistic regression in the Research on the Financial Standing of Enterprises after Bankruptcy in Poland. In *Proceedings of the 10th professor Aleksander Zeliaś International Conference on Modelling and Forecasting of Socio-Economic Phenomena. Zakopane: Foundation Cracow Univ Economics*.

Kostrzewski, M. (2012). On the Existence of Jumps in Financial Time Series. *Acta Physica Polonica-Series B Elementary Particle Physics, 43*(10), 2001.

Kostrzewski, M. (2019). The Bayesian Methods of Jump Detection: The Example of Gas and EUA Contract Prices. *Central European Journal of Economic Modelling and Econometrics (CEJEME), 11*(2), 107-131.

Pawełek, B., Kostrzewska, J., Lipieta, A. (2015). The Problem of Outliers in the Research on the Financial Standing of Construction Enterprises in Poland. In *Proceedings of the 9th Professor Aleksander Zeliaś International Conference on Modelling and Forecasting of Socio-economic Phenomena. Zakopane: Foundation Cracow Univ Economics*.

Tukey, J.W. (1977). *Exploratory Data Analysis. Addison-Wesley, Massachusetts*.

Weron, R. (2014). Electricity Price Forecasting: A Review of the State-of-the-Art with a Look into the Future. *International Journal of Forecasting, 30*(4), 1030-1081.

Weron, R., Zator, M. (2015). A note on using the Hodrick–Prescott filter in electricity markets. *Energy Economics, 48*, 1-6.

Williams, R. (2006). Generalized Ordered Logit/Partial Proportional Odds Models for Ordinal Dependent Variables. *The Stata Journal, 6*(1), 58-82.