

Evaluation of the Application of a Forecasting Model Using the Classic Least Squares Method in Forecasting Daily 15-Minute Peak Power Demand in the National Power System

Authors

Rafał Czapaj
Jacek Kamiński
Pablo Benalcazar

Keywords

forecasting, electric power demand, national power system load, classic least squares error method

Abstract

The paper discusses the results of historical (*ex post*) forecasting of the 15-minute peak demand for power in the National Power System (NPS). The study employed the classic least squares method based on the autoregressive nature of the analysed quantity (without explanatory variables) only. Simulation tests in the expired mode (with known value of the forecast variable) for the following day included analyses for 2nd and 3rd order polynomials, for the number of values delayed by from 2 to 16 preceding days, and the purpose of the study was to select their most favourable combinations. The analysed time series covered the periods of 13 years and 5 years divided into weekdays. The forecast results (*ex post* and *ex ante*) were compared with naive forecasts (for weekly delays) assuming the assignment of earlier values to the forecast quantities. Effectiveness of the most favourable expired prediction for the 3rd order polynomial and the 15-day delay obtained by the classic least squares method was lower than for the naive forecasts.

DOI: 10.12736/issn.2330-3022.2019104

Received: 22.02.2019

Received in revised form: 13.04.2019

Accepted: 16.05.2019

Available online: 30.08.2019

1. Purpose and scope of study

The subject of power demand forecasting in the National Power System has been known in Poland for many years and is very well recognized [1, 2]. The purpose of this study is to verify the applicability of the classic least squares method (LSM) to the task of forecasting the daily 15-minute peak demand for power in the National Power System (NPS). The study covered the demand time series for a period of 13 years (2002–2014), without division into weekdays and holidays, and with such a division. Subject to additional analyses were a 5-year time series (2010–2014) without division into weekdays and holidays. The time series were subjected to basic statistical analysis, seeking, among other things, observations significantly departing from the usual multimodal distributions, and significant standard deviations of the residual component (i.e. average relative errors). Simulations in the expired mode for the next day, without consideration of explanatory variables, were made for 2nd and

3rd order polynomials. Initial studies for 4th and 5th grade polynomials were abandoned due to the forecasts' poor effectiveness. The number of the delayed values analysed for the polynomial orders qualified for further study covered the range from day $n-3$ to day $n-16$ for the 2nd order polynomial, and from day $n-4$ to day $n-16$ of for the 3rd order polynomial. The analysis of the quality (effectiveness) of the expired forecasts was analysed by the Mean Absolute Percentage Error (MAPE). The forecasts were made in the MS Excel, while the statistical analyses were developed in the STATISTICA 13.3.

2. Classic least squares method

2.1. Introduction

There are many methods for estimating linear econometric models. Often used are interchangeable names: the classic least squares method and the classic linear regression model.

An indispensable postulate regarding the discussed group of methods is the conformity of estimators for the analysed estimation method [3]. Besides the classic LSM presented here, also used are Aitken's generalized LSM [3, 4], D. Cochrane's and G.H. Orcutt's complete differential method [3, 4], maximum likelihood method [3, 5], and many other referenced in the relevant literature [6–9].

2.2. Mathematical form

An advantage of the LSM is its effectiveness resulting from the Gauss-Markov theorem, which provides that this method produces consistent, unbiased and most efficient estimators even without assuming the normality of the considered variables' distribution. The linear econometric model presents [3] a formula (1) which after the estimation takes the form described [3] by formula (2). These formulas, however, were not used in the simulations and are only the starting point for further considerations.

$$Y_n = \alpha_0 + \alpha_1 \cdot X_{1t} + \alpha_2 \cdot X_{2t} + \dots + \alpha_k \cdot X_{kt} + \varepsilon_t \quad (1)$$

where: Y_n – explanatory variable showing the level of the studied phenomenon over time; $X_{1t}, X_{2t}, \dots, X_{kt}$ – explanatory variables; $\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_k$ – structural parameters; ε_t – random element.

$$Y_t = a_0 + a_1 \cdot X_{1t} + a_2 \cdot X_{2t} + \dots + a_k \cdot X_{kt} \quad (2)$$

The least squares method application leads to the determination of estimates $a_0, a_1, a_2, \dots, a_k$ of structural parameters $a_0, a_1, a_2, \dots, a_k$ so that the sum of squares of differences between individual actual values of explained variable Y_t and its theoretical values Y_t^P , obtained from the model, will be the smallest (formula 3), with the difference between the actual and theoretical values being called the model residue [3, 10]:

$$\sum_t^1 (Y_t - Y_t^P)^2 = \min \quad (3)$$

Vector of structural parameter estimates after mathematical operations on the matrices, using the LSM, is presented in formula 4, which was used in the simulations.

$$a = (X'X)^{-1} X'y \quad (4)$$

where: X' – transposed matrix, $(X'X)^{-1}$ – inverse matrix, y – value of n power demand measurements.

The least squares method, always producing the result of the smallest sum of error squares, may yield results distant from the actual trend line in the case of multiple outliers. No such phenomenon was noted in the analysed example; thus, it was assumed that the method does not limit the analysed time series. No observations that could be considered outliers were removed from the analysed time series representing the peak power demand in the National Power System. Since there were no observations, which could be considered as outliers, there

was no need to remove them from the analysed time series. The message of the Gauss-Markov theorem [11] regarding the nature of estimators, assuming the absence of many outliers, indicates that the planned simulations should give results of favourable efficacy in terms of MAPE analysis of errors for expired forecasts. The estimator conformity will lead to smaller estimation errors due to the increase of the sample size. The estimator's non-bias will guarantee that the estimates obtained with its use will be free of systematic error, owing to the fact that the estimator assumes the estimated parameter's value equal to the average value of the sample, while maintaining the sample's randomness. The estimator's highest efficiency will be ensured by obtaining the smallest standard errors. Where the most effective estimator, i.e. the estimator with unstable variance or without random component autocorrelation, has not been obtained, the generalized LSM is used [12]. This study was free of this inconvenience in the analysed examples. Where the model is heteroscedastic and there are no autocorrelations of residues, the weighted LSM is used [13]. There were residue autocorrelations in the analysed cases; therefore, there was no need to apply this method.

3. Statistical analysis of time series

Basic statistics of the analysed time series of peak power demand are presented in Tab. 1. The smallest average values were recorded for days from the 5-year period, i.e. the holiday day (16.089 GW), Sunday and Saturday. The highest average values were recorded for days from the 5-year period, i.e. Tuesday, Wednesday and Thursday (20.981 GW). The same trends for individual days were recorded for the median value for the 5-year period (16.081 GW and 21.021 GW). The smallest standard deviation was characteristic for the holiday for a 5-year period (1.483 GW), while the highest value was recorded for the entire 13-year period (2.582 GW). The smallest minimum occurred for the 13-year series (12.750 GW), then for the 5-year series for Sunday (12.750 GW). The highest maximum was achieved in the period between 2010 and 2014, and amounted to 25.845 GW. The difference between the limit values for the average was 4.892 GW, for the median 4.947 GW, for the standard deviation 1.099 GW, for the minimum 2.630 GW, for the maximum 5.947 GW.

Series	Average	Median	Std. deviation	Minimum	Maximum
	[MW]				
13 years	19,908	20,016	2,582	12,750	25,845
5 years	20,927	20,840	2,324	14,376	25,845
Monday	20,714	20,845	2,237	15,217	25,777
Tuesday	20,870	20,928	2,203	15,298	25,845
Wednesday	20,943	21,028	2,187	15,380	25,649
Thursday	20,981	21,021	2,178	14,661	25,581
Friday	20,681	20,767	2,220	14,314	25,463
Saturday	18,752	18,854	1,970	13,667	23,487
Sunday	17,275	17,432	1,876	12,750	22,117
holidays	16,089	16,081	1,483	13,116	19,898

Tab. 1. Descriptive statistics of time series subjected to the study

13 years
5 years
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
holidays

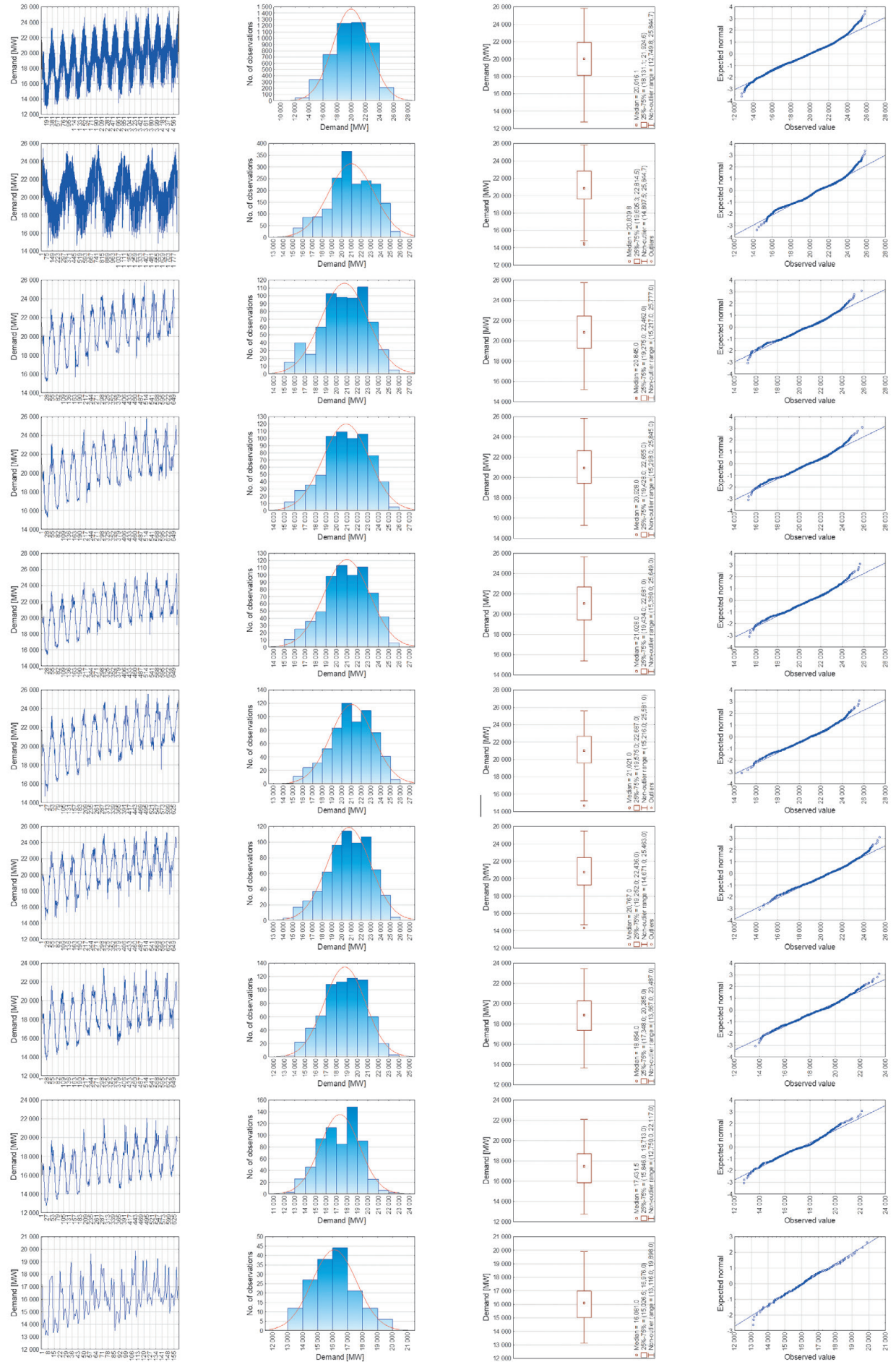


Fig. 1. Time series of peak power demand in the National Power System and their histograms, box-whiskers graphs, expected value graphs

The analysed time series, histograms, frame charts and residue plots of the series in question are presented in Fig. 1. The histograms' analysis indicates that the least effective forecasts can be developed for public holidays due to the histogram's significant asymmetry. The residue plots did not show observations that could be considered outliers, which would cause a malfunction of the classic LSM algorithm. Forecasting for the 13-year and 5-year time series may be low-efficient due to the smaller autocorrelations for the previous day compared to the day with a weekly delay, which were considered for the other time series except for holidays.

4. Ex-ante analysis of models

Forecasting models were analysed in the *ex-ante* anticipation mode. This analysis consisted of model verification, point and interval forecasts, and calculation of error and prediction precision measures. Details are presented in the following sections of this part of the paper.

4.1. Model verification

The verification of the econometric model for a representative measurement day will be presented on the example of the week with the third Wednesday of July, which is a measurement Wednesday (2014). The measures of goodness of fit used in this example were: residual variability coefficient V_e and linear determination coefficient R^2 (Tab. 2). As part of the parameter significance test, a residual graph was plotted. Point forecasts for each day were determined based on 14 previous demand measurements in accordance with formula 5 [3], where $T = n + 1, n + 2, \dots$. The forecasts are set out in Tab. 2 and 3.

$$y_t^p = ca = [1 \ T] \cdot \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} \quad (5)$$

Average prediction error S_{D^T} was used to assess the point forecasts' quality, as well as relative prediction error V_{D^T} . The determined values of *ex ante* evaluation measures are presented in Tab. 2 and 3. Interval forecasts were determined at 95% confidence level. The forecasts are also presented in Tab. 2 and 3. The interval forecasts' quality was assessed using prediction precision D_{IT^p} and the relative prediction precision V_{IT^p} . The determined values of *ex ante* evaluation measures are presented in Tab. 2 and 3.

4.2. Exante model analysis results

Detailed results of the *ex ante* analysis are presented in Tab. 2 and 3. Subject to the analysis was the third week of July 2014, including the third measurement Wednesday, considered as the representative day. The analysed period included all weekdays from Monday 14 July to Sunday 20 July. The forecast highest residual variability coefficient was recorded for Friday (0.06), and the lowest for Monday, Wednesday, Thursday and Sunday (0.02). The forecast highest linear determination coefficient was recorded for Sunday (0.9) and the lowest (0.06) for Friday. The forecast highest average prediction error estimate was recorded on Friday (1.562 GW) and the lowest on Sunday (0.419 GW). The highest relative prediction error estimate of 7.21% may be

Measure	Monday	Tuesday	Wednesday	Thursday
V_e [-]	0.02	0.03	0.02	0.02
R^2 [-]	0.41	0.33	0.46	0.53
Y [MW]	20,901.00	21,312.00	21,324.00	21,511.00
Y_T^p [MW]	22,568.21	22,296.55	22,355.55	23,063.53
S_{D^T} [MW]	681	778	644	460
V_{D^T} [%]	3.02	3.49	2.88	2.00
-95% [MW]	21,069.66	20,585.10	20,937.50	22,050.40
+95% [MW]	24,066.76	24,008.01	23,773.61	24,076.67
D_{IT^p} [MW]	1,498.55	1,711.45	1,418.06	1,013.13
V_{IT^p} [%]	6.64	7.68	6.34	4.39

Tab. 2. Results of model verification, and *ex ante* point and interval forecasts for days from Monday to Thursday

Measure	Friday	Saturday	Sunday	holidays
V_e [-]	0.06	0.03	0.02	0.03
R^2 [-]	0.06	0.18	0.90	0.41
Y [MW]	21,711.00	19,185.00	16,834.00	20,396.86
Y_T^p [MW]	21,648.01	19,226.26	19,238.37	21,485.21
S_{D^T} [MW]	1,562	855	419	771
V_{D^T} [%]	7.21	4.45	2.18	3.60
-95% [MW]	18,210.82	17,344.85	18,315.34	19,787.67
+95% [MW]	25,085.19	21,107.66	20,161.40	23,182.76
D_{IT^p} [MW]	3,437.19	1,881.40	923.03	1,697.54
V_{IT^p} [%]	15.88	9.79	4.80	7.93

Tab. 3. Results of model verification, and *ex ante* point and interval forecasts for days from Friday to Sunday

characteristic for Friday, and the lowest for Thursday and Sunday (2.00% and 2.18%, respectively). Very good forecasts have a level of <3%. The forecast highest prediction precision was recorded on Sunday (0.923 GW), and the lowest on Friday (3.437 GW). The forecast highest relative prediction precision was recorded on Sunday (4.80%), and the lowest on Friday (15.88%).

5. Simulation results

5.1. 2nd order polynomial

Simulation results for the 2nd order polynomial for daily delayed values (daily delays) from 3 to 9 are presented in Tab. 4, while for daily delays from 10 to 16 in Tab. 5. Analysis of the results in the tables indicates that MAPE *ex post* errors decrease for each of the analysed time series with increase in the number of daily delayed values. The biggest MAPE error reduction was recorded for holidays (10.73%), followed by the 5-year period (9.13%), and the 15-year time series (8.71%). The reductions noted for the remaining time series are around 2.9–4.7%. The smallest reductions in the accuracy of forecasts in terms of MAPE error result from the smaller values recorded for the starting point, i.e. the 3-day delay. The lowest MAPE error rates were recorded for Tuesdays, Wednesdays and Thursdays, 2.78%, 2.83%, and 2.89%, respectively. The highest values of these errors were

characteristic for holidays (20.12%), the 5-year period (17.13%), and the 13-year period (15.67%). The smallest arithmetically averaged errors for all daily delays were recorded for Tuesdays (3.26%), followed by Wednesdays (3.33%) and Thursdays (3.40%). The highest averaged values were characteristic for the 5-year period (12.11%), holidays (11.48%), and the 13-year period (11.21%). The daily delay 13 turned out to be the most advantageous for Wednesdays and Thursdays, while the best for Friday and for the 5-year and 13-year periods was the delay 14. The most effective forecasts for Mondays and Saturdays were found for the daily delay 15, while the best for Tuesdays and Thursdays was the daily delay 16. The most effective forecasts for holidays were for the daily delay 9.

Simulation results for the 3rd order polynomial for daily delays from 3 to 9 are presented in Tab. 6, while for daily delays from 10 to 16 in Tab. 7. Analysis of the results in the tables indicates that MAPE *ex post* errors decrease for each of the analysed time series with increase in the number of daily delayed values. The biggest MAPE error reduction was recorded for holidays (29.85%), followed by the 5-year period (21.40%), and the 13-year

time series (20.09%). The reductions noted for the remaining time series are around 11.9–7.8%. The smallest reductions in the accuracy of forecasts in terms of MAPE error result from the smaller values recorded for the starting point, i.e. the 3-day delay. The lowest MAPE error rates were recorded for Tuesdays, Wednesdays and Thursdays, 2.73%, 2.76%, and 2.81%, respectively. The highest values of these errors were characteristic for holidays (36.30%), the 5-year period (29.62%), and the 13-year period (27.19%). The smallest arithmetically averaged errors for all daily delays were recorded for Tuesdays (4.57%), followed by Wednesdays (4.67%) and Thursdays (4.77%). The highest averaged values were characteristic for the 5-year period (17.16%), holidays (15.67%), and the 13-year period (15.05%). The 15-day delay turned out to be the most advantageous for all analysed time series. The 3rd order polynomial application turned out to be more advantageous in most of the analysed time series and concerned all day types analysed separately. The largest increase in the forecast effectiveness between the 2nd and 3rd order polynomials was recorded in holiday days (error reduction by 3.01% in favour of the 3rd order polynomial). The exemplary vector of

Series	MAPE error [%] for previous days						
	(-3) days	(-4) days	(-5) days	(-6) days	(-7) days	(-8) days	(-9) days
13 years	15.11	13.38	11.70	13.69	15.67	11.89	12.65
5 years	17.13	15.60	13.51	10.90	12.43	13.41	14.40
Monday	6.84	4.98	4.25	3.83	3.68	3.50	3.42
Tuesday	5.85	4.32	3.61	3.25	3.06	2.93	2.88
Wednesday	5.73	4.37	3.69	3.44	3.25	3.06	3.00
Thursday	6.01	4.51	3.81	3.48	3.21	3.08	2.98
Friday	8.16	6.21	5.11	4.68	4.29	4.07	3.97
Saturday	7.72	5.66	4.69	4.07	3.73	3.44	3.35
Sunday	6.77	4.61	3.94	3.58	3.35	3.24	3.19
holidays	20.12	15.48	12.82	11.67	11.12	10.24	9.39

Tab. 4. MAPE errors of the *ex post* forecasts for the 2nd order polynomial based on days preceding by from 3 to 9 for individual periods and day types

Series	MAPE error [%] for previous days						
	(-10) days	(-11) days	(-12) days	(-13) days	(-14) days	(-15) days	(-16) days
13 years	11.99	10.74	9.26	7.87	6.96	7.59	8.41
5 years	13.77	12.36	10.63	9.06	8.00	8.70	9.67
Monday	3.35	3.29	3.26	3.22	3.22	3.21	3.22
Tuesday	2.84	2.81	2.84	2.81	2.82	2.81	2.78
Wednesday	2.90	2.84	2.88	2.83	2.86	2.88	2.91
Thursday	2.99	2.94	2.94	2.89	2.91	2.91	2.89
Friday	3.87	3.74	3.65	3.55	3.52	3.52	3.56
Saturday	3.36	3.22	3.19	3.11	3.10	3.06	3.08
Sunday	3.13	3.07	3.07	3.05	3.07	3.07	3.09
holidays	9.72	10.22	10.48	10.12	10.06	9.82	9.46

Tab. 5. MAPE errors of the *ex post* forecasts for the 2nd order polynomial based on days preceding by from 10 to 16 for individual periods and day types

Series	MAPE error [%] for previous days						
	(-3) days	(-4) days	(-5) days	(-6) days	(-7) days	(-8) days	(-9) days
13 years	–	27.19	21.05	19.91	15.30	12.90	14.10
5 years	–	29.62	24.30	23.01	16.75	15.19	16.56
Monday	–	12.79	8.24	6.57	5.60	5.08	4.68
Tuesday	–	10.82	7.20	5.53	4.76	4.24	3.90
Wednesday	–	10.57	7.17	5.53	4.95	4.42	4.06
Thursday	–	11.04	7.50	5.81	5.10	4.46	4.09
Friday	–	15.17	10.46	8.06	6.86	5.96	5.41
Saturday	–	14.37	9.68	7.27	6.11	5.47	4.65
Sunday	–	12.37	7.77	5.98	5.20	4.64	4.24
holidays	–	36.30	24.79	20.91	16.92	15.12	14.24

Tab. 6. MAPE errors of the *ex post* forecasts for the 3rd order polynomial based on days preceding by from 3 to 9 for individual periods and day types

Series	MAPE error [%] for previous days						
	(-10) days	(-11) days	(-12) days	(-13) days	(-14) days	(-15) days	(-16) days
13 years	14.69	14.77	14.24	13.60	12.42	7.10	8.32
5 years	17.18	17.16	16.04	15.40	14.16	8.22	9.51
Monday	4.40	4.23	4.06	3.96	3.84	3.01	3.61
Tuesday	3.66	3.53	3.37	3.32	3.25	2.73	3.15
Wednesday	3.88	3.75	3.61	3.49	3.34	2.76	3.22
Thursday	3.79	3.66	3.60	3.52	3.40	2.81	3.28
Friday	5.12	4.97	4.78	4.59	4.45	3.28	4.12
Saturday	4.22	4.18	4.03	3.94	3.81	3.03	3.57
Sunday	4.04	3.93	3.80	3.71	3.55	3.03	3.43
holidays	12.50	12.11	11.84	10.68	9.99	6.45	11.91

Tab. 7. MAPE errors of the *ex post* forecasts for the 3rd order polynomial based on days preceding by from 10 to 16 for individual periods and day types

structural parameter estimates for Wednesday 16 July 2014, the 3rd order polynomial and 15 daily delays, consisting of parameters a_0 , a_1 , a_2 and a_3 , reached 21 256.98, -145.36, -3.18, and 1.21, respectively; and the MAPE error reached was 0.31%.

5.2. Ex ante vs. ex post evaluation

Ex-ante evaluations from July 14–20, 2014, presented in Tab. 2 and 3, were compared with expired *ex post* forecasts obtained for the same period (Tab. 4–7). The smallest MAPE *ex post* errors for the 3rd order polynomial and 15-day delay were included in 95-percent confidence interval designated *ex ante* for 4 out of 7 days of the week, i.e. for Tuesday, Wednesday, Friday and Saturday. The demands on Monday, Thursday and Sunday were below those determined by the confidence interval.

The demand for peak power on Monday, July 14, 2014 was 278.38 MW below the left end of the confidence interval. On Thursday, 17 July 2014, the actual demand was lower by 1222.85 MW than the left end of the confidence interval, while for Sunday, July 20, 2014, this difference amounted to 2081.16 MW. The detailed values of the forecasts and MAPE errors obtained for the analysed week are as follows: 20,791.28 MW (0.52%), 20,956.61 MW (1.67%), 20,963.20 MW (1.69%), 20,837.55 MW (3.13%), 20,541.70 MW (5.39%), 18,267.16 MW (4.78%), 16,234.18 MW (3.56%). In this period only the expired forecasts for Monday, Tuesday and Wednesday should be considered as satisfactorily precise.

5.3. Naive forecasts

To compare the most favourable results of expired forecasts, evaluated by means of the MAPE measure, obtained for the 2nd order polynomial and the daily delay of 15, expired forecasts with a weekly delay were developed. The naive forecasts for each of the subsequent days of the week were better. The naive method for days from Monday to Sunday was more effective by 0.50%, 0.52%, 0.50%, 0.49%, 0.29%, 0.26%, 0.54%, respectively, compared to the *ex post* forecasts. For the 13-year time series, the naive method was more effective by 3.51%, and for the 5-year period by 4.47%. Only in the case of the public holidays time series the naive method turned out to be less effective than the LSM for the 3rd order polynomial by 0.47%.

6. Conclusions

The simulation analyses showed that for the classic least squares method and the use of only historical information about the shaping of individual time series, the higher efficiency of obtained expired forecasts was obtained using the 3rd order polynomial in comparison with the 2nd order polynomial. The 2nd order polynomial application produced the most effective expired forecasts for daily delays of 9, 13–16. With regard to the 3rd order polynomial, all the most effective forecasts were obtained for daily delay 15. The effectiveness obtained for the 2nd order polynomial was not much higher for the subsequent days of the week, respectively: 0.20%, 0.05%, 0.07%, 0.08%, 0.24%, 0.03%, and 0.02%. A significant improvement in effectiveness was obtained only for holidays and amounted to 3.01%.

The 2nd order polynomial application yielded less effective results compared to the 3rd order polynomial only for the 13-year and 5-year time series, by 0.14% and 0.22%, respectively, using the principle of searching for the lowest MAPE errors for each time series. Direct comparison of MAPE error values between the polynomials for appropriate delay values indicates that the 2nd order polynomial produces lower values than the 3rd degree polynomial. Comparison of both approaches indicates, however, that the application of the 3rd order polynomial and 15 delays in most of the analysed time series produced better results than the 2nd order polynomial and the need to look for the best value of the time delay. The naive method application for 7-day delay and comparing its results with the analogous results for the LSM and the 3rd order polynomial each time yielded better results. The obtained more favourable effectiveness differences, expressed as MAPE errors for the naive method for subsequent days of the week, were: 0.50%, 0.52%, 0.50%, 0.49%, 0.29%, 0.26%, and 0.54%. For the 13-year and 5-year time series, the effectiveness was higher by 3.51% and 4.47%, respectively. The better by 0.47% result obtained for public holidays by the LSM should be considered an insignificant exception compared to the remaining results.

Summarizing the simulation results obtained for the classic LSM, it should be noted that the 3rd order polynomial application had improved the efficiency of expired forecasts in comparison with the 2nd order polynomial. The obtained simulation results against the background of the naive method and against the background of the effectiveness of the forecasts obtained by more sophisticated methods available in the relevant literature should be considered too low for practical applications. Forecasts of this type can be at most an additional model that would be part of a group model.

The low effectiveness of the expired forecasts described by MAPE error may result from the applied approach consisting in the consideration of only the autoregressive information about the historical development of individual time series. Further research may therefore focus on including explanatory variables that will increase the forecasts' effectiveness. These variables may include the variables describing measurements and observations of meteorological parameters as well as information on the day type. Thus, it will be possible to more fully utilize the potential offered by the algorithm of the classic LSM.

REFERENCES

1. Prognozowanie w elektroenergetyce [Forecasting in the Power Sector], edited by I. Dobrzańska, Wydawnictwo Politechniki Częstochowskiej, Częstochowa 2002.
2. Kopecki K. et al., Analiza i prognoza obciążeń elektroenergetycznych [Analysis and Forecast of Power Loads], Wydawnictwa Naukowo-Techniczne, Warsaw 1971.
3. Czyżycki R., Klóska R., Ekonometria i prognozowanie zjawisk ekonomicznych w przykładach i zadaniach [Econometrics and Economic Phenomena Forecasting in Examples and Exercises], Economicus, Szczecin 2011, pp. 30–31.

4. Pawłowski Z., Elementy ekonometrii [Elements of Econometrics], PWN, Warsaw 1981, pp. 186, 209–212, 224–229.
5. Romejko A., Estymacja liniowego modelu wydatków metodą największej wiarygodności [Linear Expenditure Model Estimation by Highest Likelihood Method], *Wiadomości Statystyczne*, No. 75, 1978, pp. 11–14.
6. Hellwig Z., O jakości modelu ekonometrycznego. Część II. Estymacja, [About Econometric Model Quality. Part II. Estimation] *Przegląd Statystyczny*, No. 4, 1985, p. 297.
7. Kolupa M., Metody estymacji modeli ekonometrycznych [Methods for Econometric Models Estimation], PWE, Warsaw 1974.
8. Milo W., Nieliniowe modele ekonometryczne [Non-linear Econometric Models], PWN, Warsaw 1990.
9. Grabiński T., Wydymus S., Szacowanie liniowych funkcji trendu metodą najmniejszych kwadratów przy pomocy wzorów uproszczonych [Estimation of Linear Trend Functions by Least Squares Method with Simplified Formulas], *Wiadomości Statystyczne*, No. 4, 1975, GUS, p. 20.
10. Ekonometria. Metody i analiza problemów ekonomicznych [Econometrics. Methods and Analysis of Economic Problems], ed. K. Jajuga, Akademia Ekonomiczna, Wrocław 2002, p. 59.
11. Gruszczynski M., Kuszewski T., Podgórska M., Ekonometria i badania operacyjne, podręcznik dla studiów licencjackich [Econometrics and Operational Research, a textbook for bachelor studies], PWN, Warsaw 2009.
12. Borkowski B., Dudek H., Szczęśny W., Ekonometria wybrane zagadnienia [Econometrics Selected Issues], PWN, Warsaw 2003.
13. Stanisz A., Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny [Accessible Statistics Course with STATISTICA PL on Examples from Medicine], Vol. 1 i 2, StatSoft Polska, Krakow 2006.

Rafał Czapaj

PSE Innowacje Ltd.

e-mail: rafal.czapaj@pse.pl

A graduate in processing and use of electricity from the Faculty of Electrical Engineering of the Silesian University of Technology (2003). He has been associated with the Capital Group of PSE SA since January 2005. In 2005–2011 he worked at EPC Consulting SA and dealt with the subject of the electricity market and technical and economic analyses. Since 2011 he has been working at PSE Innowacje Ltd. (formerly CATA Ltd.), dealing with issues of short- and long-term forecasting, optimization of selection of input data sets for forecasting models, methods of searching Big Data sets, issues of operation of transmission grid elements, robotization of inspection works, machine learning and artificial intelligence methods in the power sector.

Jacek Kamiński

Mineral and Energy Economy Research Institute of the Polish Academy of Sciences, Krakow

e-mail: kaminski@min-pan.krakow.pl

Associate Professor at the Mineral and Energy Economy Research Institute of the Polish Academy of Sciences in Krakow. Head of the Energy Economics Division and of the Department of Policy and Strategic Research. He received a doctorate in engineering at the AGH University of Science and Technology in Kraków (2006), and a habilitation degree in power engineering at the Silesian University of Technology in Gliwice (2013). In 2007–2009 he was a national expert seconded to the European Research Centre of the European Commission – Institute for Studies of Technological Perspectives (European Commission - DG Joint Research Center - Institute for Prospective Technological Studies – EC / DG JRC / IPTS) in Seville. Author and co-author of over 100 scientific publications, including in journals: *Energy Policy*, *Energy*, *Applied Energy*. An expert in building mathematical models and decision-making supporting systems in energy enterprises, fuel and energy markets and energy policy. Manager of several dozen studies and projects commissioned by energy companies. Since 2016 chairman of the board of the Polish affiliate of the International Association for Energy Economics (IAEE).

Pablo Benalcazar

Mineral and Energy Economy Research Institute of the Polish Academy of Sciences, Krakow

e-mail: benalcazar@min-pan.krakow.pl

Lecturer/researcher at the Mineral and Energy Economy Research Institute of the Polish Academy of Sciences in the Division of Energy Economics of the of Policy and Strategic Research. A PhD candidate at AGH University of Science and Technology in Krakow. He obtained a MSc in chemical technology with the specialization of clean fossil and alternative fuels energy from the AGH University of Science and Technology. In his research he focuses on energy economics, energy policy, mathematical optimization and forecasting.

Ocena zastosowania modelu prognostycznego wykorzystującego klasyczną metodę najmniejszych kwadratów w prognozowaniu 15-minutowego szczytowego zapotrzebowania na moc w KSE

Autorzy

Rafał Czapaj
Jacek Kamiński
Pablo Benalcazar

Słowa kluczowe

prognozowanie, zapotrzebowanie na moc elektryczną, obciążenie KSE, klasyczna metoda najmniejszych kwadratów błędów

Streszczenie

W artykule przedstawiono wyniki prognozowania wygasłego (na podstawie *ex post*) 15-minutowego szczytowego zapotrzebowania na moc elektryczną w Krajowym Systemie Elektroenergetycznym (KSE). Badania przeprowadzono z zastosowaniem klasycznej metody najmniejszych kwadratów (MKNK), bazując jedynie na autokorelacyjnym charakterze analizowanej wielkości (bez udziału zmiennych objaśniających). Testy symulacyjne w trybie wygasłym (przy znanej wartości zmiennej prognozowanej) na następną dobę obejmowały analizy dla wielomianu 2 oraz 3 stopnia, dla liczby wartości opóźnionych od 2 do 16 dób poprzedzających, a celem artykułu było najkorzystniejsze dobranie ich kombinacji. Analizowane szeregi czasowe obejmowały okres 13 lat oraz 5 lat w podziale na dni tygodnia. Otrzymane wyniki prognoz (*ex post* i *ex ante*) porównano z prognozami naiwnymi (dla opóźnień tygodniowych), zakładającymi przypisanie wartości prognozowanym wartości wcześniejszych. Skuteczność najkorzystniejszej wygasłej predykcji dla wielomianu 3 stopnia i opóźnienia 15-dobowego za pomocą klasycznej metody najmniejszych kwadratów była niższa niż dla prognoz naiwnych.

Data wpływu do redakcji: 22.02.2019

Data wystawienia recenzji: 13.04.2019

Data akceptacji artykułu: 16.05.2019

Data publikacji online: 30.08.2019

1. Cel i zakres badań

Tematyka prognozowania zapotrzebowania na moc w KSE jest znana w Polsce od wielu lat i jest bardzo dobrze rozpoznana [1, 2]. Celem przeprowadzonych badań jest weryfikacja możliwości zastosowania klasycznej metody MNK w zadaniu prognozowania 15-minutowego szczytowego dobowego zapotrzebowania na moc elektryczną w Krajowym Systemie Elektroenergetycznym (KSE). Badaniom poddano szereg czasowy zapotrzebowania za okres 13 lat (2002–2014), bez podziału na dni tygodnia i święta oraz z zastosowaniem takiego podziału. Dodatkowo analizom został poddany 5-letni szereg czasowy (2010–2014), bez uwzględnienia podziału na dni tygodnia i dni świąteczne. Szeregi czasowe poddano podstawowej analizie statystycznej, poszukując m.in. obserwacji znacząco odstających od zwyczajowych rozkładów wielomodalnych i znaczących standardowych odchyłek składnika resztowego (czyli przeciętnych błędów względnych). Symulacje w trybie wygasłym dla kolejnej doby, bez uwzględnienia zmiennych objaśniających, wykonano dla wielomianu 2 i 3 stopnia. Wstępne badania dla wielomianów 4 i 5 stopnia zarzucono ze względu na uzyskanie prognoz o niskiej skuteczności. Liczba wartości opóźnionych, dla której przeprowadzono analizę dla zakwalifikowanych do dalszych badań stopni wielomianu, obejmowała zakres od doby $n-3$ do doby $n-16$ dla wielomianu 2 stopnia oraz od doby $n-4$ do doby $n-16$ dla wielomianu 3 stopnia.

Analizę jakości (skuteczności) uzyskanych prognoz wygasłych wykonano z wykorzystaniem średniego bezwzględnego błędu procentowego MAPE (ang. *Mean Absolute Percentage Error*). Prognozy wykonano w programie MS Excel, natomiast analizy statystyczne opracowano w programie STATISTICA 13.3.

2. Klasyczna metoda najmniejszych kwadratów (MKNK)

2.1. Wprowadzenie

Istnieje wiele metod szacowania liniowych modeli ekonometrycznych. Często stosowane są wymiennie nazwy: klasyczna metoda najmniejszych kwadratów oraz klasyczny model regresji liniowej. Niezbędnym postulatem dotyczącym omawianej grupy metod jest zgodność estymatorów dla analizowanej metody estymacji [3]. Oprócz klasycznej metody MNK, zaprezentowanej w niniejszym artykule, stosowane są: uogólniona metoda najmniejszych kwadratów Aitkena [3, 4], metoda różniczki zupełnej D. Cochran'a i G.H. Orcutt'a [3, 4], metoda największej wiarygodności [3, 5] oraz wiele innych przytaczanych w annałach literatury przedmiotu [6–9].

2.2. Postać matematyczna

Zaletą metody MNK jest jej skuteczność wynikająca z twierdzenia Gaussa-Markowa, w którym czytamy, że metoda ta daje estymatory zgodne, nieobciążone i najefektywniejsze nawet bez założenia o normalności rozkładu rozpatrywanych zmiennych.

Liniowy model ekonometryczny przedstawia [3] wzór (1), który po oszacowaniu przyjmuje postać opisaną [3] wzorem (2). Wskazane wzory nie zostały jednak wykorzystane w badaniach symulacyjnych i stanowią wyłącznie punkt wyjściowy do dalszych rozważań.

$$Y_n = \alpha_0 + \alpha_1 \cdot X_{1t} + \alpha_2 \cdot X_{2t} + \dots + \alpha_k \cdot X_{kt} + \varepsilon_t \quad (1)$$

gdzie: Y_n – zmienna objaśniana obrazująca poziom badanego zjawiska w czasie; $X_{1t}, X_{2t}, \dots, X_{kt}$ – zmienne objaśniające; $\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_k$ – parametry strukturalne; ε_t – składnik losowy.

$$Y_t = \alpha_0 + \alpha_1 \cdot X_{1t} + \alpha_2 \cdot X_{2t} + \dots + \alpha_k \cdot X_{kt} \quad (2)$$

Zastosowanie metody MNK prowadzi do wyznaczenia wartości ocen $\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_k$ parametrów strukturalnych $\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_k$ tak że suma kwadratów różnic pomiędzy poszczególnymi wartościami rzeczywistymi zmiennej objaśnianej Y_t a jej wartościami teoretycznymi Y_t^P , otrzymanymi z modelu, będzie najmniejsza (wzór 3), przy czym różnicę wartości rzeczywistych i teoretycznych określa się mianem reszty modelu [3, 10]:

$$\sum_t^1 (Y_t - Y_t^P)^2 = \min \quad (3)$$

PL

Wektor ocen parametrów strukturalnych po przeprowadzeniu operacji matematycznych na macierzach, z wykorzystaniem metody MNK, przedstawia wzór 4, który zastosowano w badaniach symulacyjnych.

$$a = (X'X)^{-1} X'y \quad (4)$$

gdzie: X' – macierz transponowana, $(X'X)^{-1}$ – macierz odwrotna, y – wartość n pomiarów zapotrzebowania na moc.

Metoda MNK, dając zawsze wynik o najmniejszej sumie kwadratów błędów, może przynosić rezultaty odległe od rzeczywistej linii trendu w sytuacji występowania wielu pomiarów odstających. W analizowanym przykładzie nie odnotowano takiego zjawiska, tym samym założono, że metoda nie stanowi ograniczenia dla analizowanych szeregów czasowych. Z analizowanych szeregów czasowych, reprezentujących zapotrzebowanie na moc szczytową w KSE, nie usuwano żadnych obserwacji, które mogłyby być uznane za odstające. Brak obserwacji, które mogłyby zostać uznane za odstające, spowodował brak potrzeby ich usuwania z analizowanych szeregów czasowych. Przesłanie twierdzenia Gaussa-Markowa [11], dotyczące charakteru estymatorów, przy założeniu braku wielu obserwacji odstających, wskazuje, że planowane symulacje powinny dać wyniki o korzystnej skuteczności w sensie analizy błędów MAPE dla prognoz wygasłych. Zgodność estymatora będzie prowadziła do uzyskania mniejszych błędów estymacji dzięki zwiększeniu liczebności próby. Nieobciążoność estymatora będzie gwarantować otrzymanie za jego pomocą ocen pozbawionych błędów systematycznego, dzięki przyjmowaniu przez estymator wartości szacowanego parametru równej średniej wartości próby, z zachowaniem losowości próby. Największa efektywność estymatora będzie zapewniona dzięki uzyskiwaniu najmniejszych wartości błędów standardowego. W sytuacji braku uzyskania estymatora o największej efektywności, czyli estymatora o niestejłej wariancji lub braku autokorelacji składnika losowego, stosuje się uogólnioną metodę najmniejszych kwadratów [12]. W wykonywanych badaniach wskazano niedogodność nie zachodząca w analizowanych przykładach. W sytuacji, gdy model jest heteroskedastyczny, a nie zachodzą autokorelacje reszt, stosuje się ważoną metodę najmniejszych kwadratów [13]. W analizowanych przypadkach występowały autokorelacje reszt, dlatego też nie wystąpiła konieczność zastosowania tej metody.

3. Analiza statystyczna szeregów czasowych

Podstawowe statystyki analizowanych szeregów czasowych zapotrzebowania na moc szczytową zamieszczono w tab. 1. Najmniejsze wartości średnie odnotowano dla dni z okresu 5-letniego, tj. dnia świątecznego (16,089 GW), niedzieli i soboty. Wartości najwyższe średnie odnotowano dla dni z okresu 5-letniego, tj. wtorku, środy i czwartku (20,981 GW). Analogiczne tendencje dla poszczególnych dni odnotowano dla wartości mediany dla okresu 5-letniego (16,081 GW i 21,021 GW). Najmniejsze odchylenie standardowe

Szereg	Średnia	Mediana	Odchyl. stand.	Minimum	Maksimum
	[MW]				
13 lat	19 908	20 016	2 582	12 750	25 845
5 lat	20 927	20 840	2 324	14 376	25 845
poniedziałek	20 714	20 845	2 237	15 217	25 777
wtorek	20 870	20 928	2 203	15 298	25 845
środa	20 943	21 028	2 187	15 380	25 649
czwartek	20 981	21 021	2 178	14 661	25 581
piątek	20 681	20 767	2 220	14 314	25 463
sobota	18 752	18 854	1 970	13 667	23 487
niedziela	17 275	17 432	1 876	12 750	22 117
święta	16 089	16 081	1 483	13 116	19 898

Tab. 1. Statystyki opisowe szeregów czasowych poddanych badaniom

charakteryzowało dzień świąteczny dla okresu 5-letniego (1,483 GW), natomiast najwyższą wartość odnotowano dla całego okresu 13-letniego (2,582 GW). Najmniejsza wartość minimalna wystąpiła dla szeregu 13-letniego (12,750 GW), następnie dla szeregu 5-letniego dla niedzieli (12,750 GW). Wartość najwyższa maksimum osiągnięta została w okresie pomiędzy 2010 a 2014 rokiem, i wyniosła 25,845 GW. Różnica pomiędzy wartościami granicznymi dla średniej wyniosła 4,892 GW, dla mediany 4,947 GW, dla odchylenia standardowego 1,099 GW, dla minimum 2,630 GW, dla wartości maksymalnej 5,947 GW. Przebiegi analizowanych szeregów czasowych, histogramy, wykresy ramkowe i wykresy reszt przedmiotowych szeregów zaprezentowano na rys. 1. Analiza histogramów wskazuje, że najmniej skuteczne prognozy mogą zostać opracowane dla dni świątecznych ze względu na znaczącą asymetrię histogramu. Wykresy reszt nie wykazały obserwacji, które mogłyby być uznane za obserwacje odstające, powodujące wadliwe działanie algorytmu klasycznej metody MNK. Dla szeregu czasowego 13-letniego i 5-letniego może zaistnieć niska skuteczność prognozowania ze względu na mniejsze autokorelacje dla doby poprzedzającej w porównaniu z dobą z opóźnieniem tygodniowym, które brano pod uwagę dla pozostałych szeregów czasowych z wyjątkiem dni świątecznych.

4. Analiza modeli *ex ante*

Wykonano analizę modeli prognostycznych w trybie wyprzedzającym *ex ante*. Na analizę tę składały się: weryfikacja modelu, wyznaczenie prognoz punktowych i przedziałowych oraz obliczenie mierników błędów i precyzji predykcji. Szczegóły zaprezentowano w kolejnych sekcjach niniejszej części artykułu.

4.1. Weryfikacja modelu

Weryfikacja modelu ekonometrycznego dla reprezentatywnego dnia pomiarowego zostanie przedstawiona na przykładzie tygodnia, w którym przypada trzecia środa lipca, będąca środą pomiarową (2014). Miarami dobroci dopasowania zastosowanymi w tym przykładzie były: współczynnik zmienności resztowej i precyzji predykcji. Szczegóły przedstawiono w kolejnych sekcjach niniejszej części artykułu.

na każdą dobę, bazując na 14 wcześniejszych pomiarach wielkości zapotrzebowania zgodnie ze wzorem 5 [3], gdzie $T = n + 1, n + 2, \dots$. Wyznaczone prognozy zamieszczono w tab. 2 i 3.

$$y_t^p = ca = [1 \ T] \cdot \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} \quad (5)$$

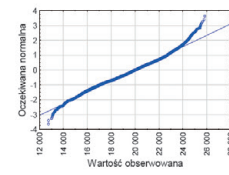
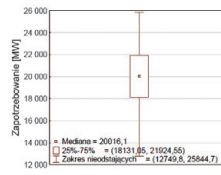
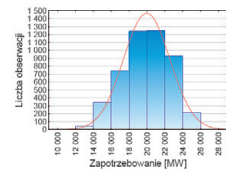
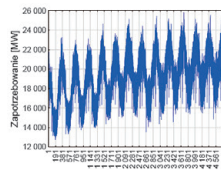
Do oceny jakości prognoz punktowych wykorzystano średni błąd predykcji S_D^T oraz względny błąd predykcji V_D^T . Wyznaczone wartości mierników ocen *ex ante* zaprezentowano w tab. 2 i 3. Wyznaczone prognozy przedziałowe na poziomie wiarygodności 95%. Prognozy zamieszczono również w tab. 2 i 3. Do oceny jakości prognoz przedziałowych wykorzystano precyzję predykcji D_{IT}^P oraz względną precyzję predykcji V_{IT}^P . Wyznaczone wartości mierników ocen *ex ante* zamieszczono w tab. 2 i 3.

4.2. Wyniki analizy modeli *ex ante*

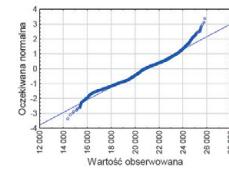
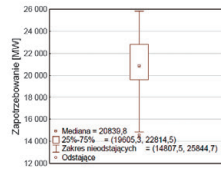
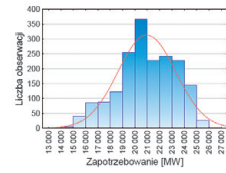
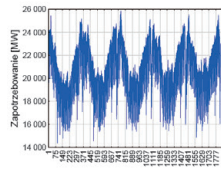
Szczegółowe wyniki analizy *ex ante* zaprezentowano w tab. 2 i 3. Analiza została przeprowadzona dla trzeciego tygodnia lipca 2014 roku, obejmującego trzecią środę pomiarową, określaną jako dzień reprezentatywny. Analizowany okres obejmował wszystkie dni tygodnia z zakresu od poniedziałku 14 lipca do niedzieli 20 lipca. Przewidywana największa wartość współczynnika zmienności resztowej została odnotowana dla piątku (0,06), a najmniejsza dla poniedziałku, środy, czwartku i niedzieli (0,02). Przewidywana największa wartość współczynnika determinacji liniowej została odnotowana dla niedzieli (0,9), a najmniejsza (0,06) dla piątku. Przewidywana największa ocena średniego błędu predykcji dotyczy piątku (1,562 GW), a najmniejsza niedzieli (0,419 GW). Największą oceną względnego błędu predykcji, na poziomie 7,21%, może charakteryzować się piątek, a najmniejszą czwartek i niedziela (odpowiednio 2,00% i 2,18%). Bardzo dobre prognozy charakteryzują się poziomem <3%. Przewidywana najwyższa precyzja predykcji dotyczy niedzieli (0,923 GW), a najmniejsza dotyczy piątku (3,437 GW). Przewidywana największa względna precyzja predykcji dotyczy niedzieli (4,80%), a najmniejsza piątku (15,88%).

PL

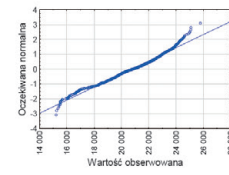
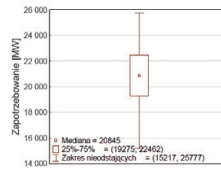
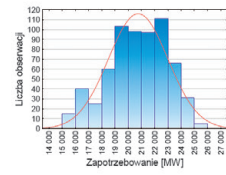
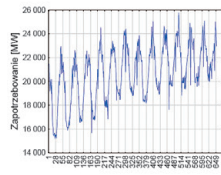
13 lat



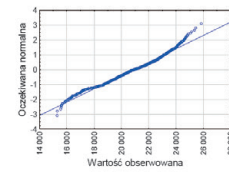
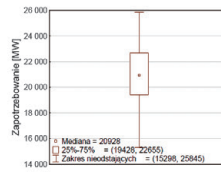
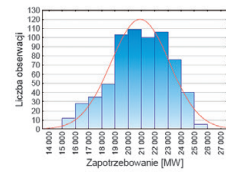
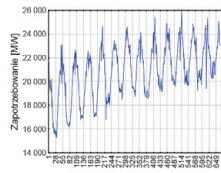
5 lat



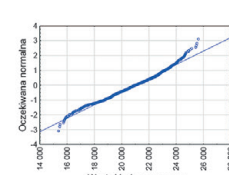
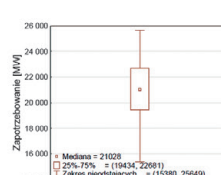
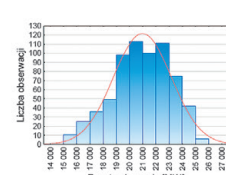
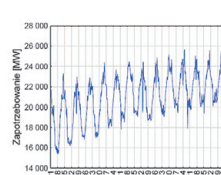
Poniedziałek



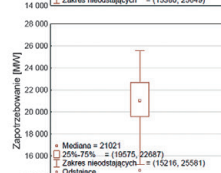
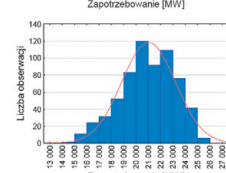
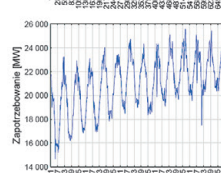
Wtorek



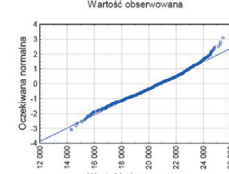
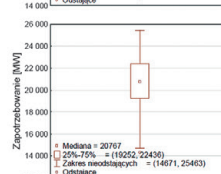
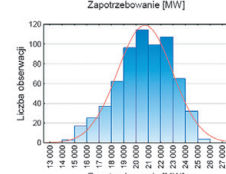
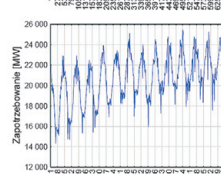
Środa



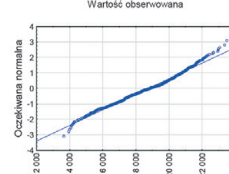
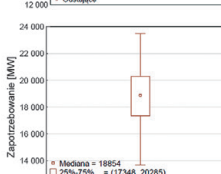
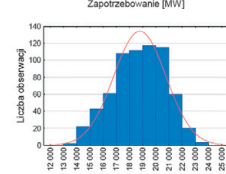
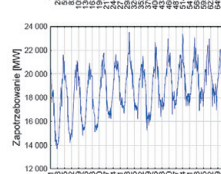
Czwartek



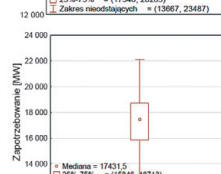
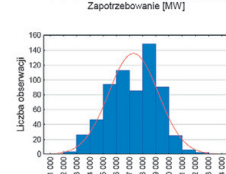
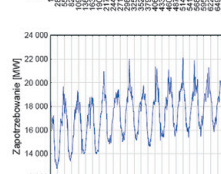
Piątek



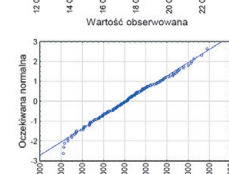
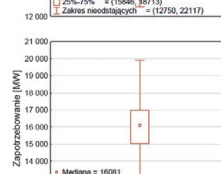
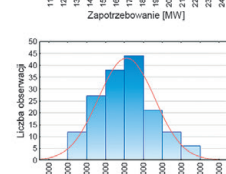
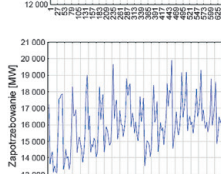
Sobota



Niedziela



Święta



Rys. 1. Przebiegi szeregów czasowych zapotrzebowania na moc szczytową w KSE oraz ich histogramy, wykresy ramka-wąsy i wykresy wartości oczekiwanych

PL

Miernik	poniedziałek	wtorek	środa	czwartek
V_e [-]	0,02	0,03	0,02	0,02
R^2 [-]	0,41	0,33	0,46	0,53
Y [MW]	20 901,00	21 312,00	21 324,00	21 511,00
Y_P [MW]	22 568,21	22 296,55	22 355,55	23 063,53
S_D^T [MW]	681	778	644	460
V_D^T [%]	3,02	3,49	2,88	2,00
-95% [MW]	21 069,66	20 585,10	20 937,50	22 050,40
+95% [MW]	24 066,76	24 008,01	23 773,61	24 076,67
D_{IT}^P [MW]	1 498,55	1 711,45	1 418,06	1 013,13
V_{IT}^P [%]	6,64	7,68	6,34	4,39

Tab. 2. Wyniki weryfikacji modeli, prognoz punktowych i przedziałowych *ex ante* dla dni od poniedziałku do czwartku

Miernik	piątek	sobota	niedziela	święta
V_e [-]	0,06	0,03	0,02	0,03
R^2 [-]	0,06	0,18	0,90	0,41
Y [MW]	21 711,00	19 185,00	16 834,00	20 396,86
Y_P [MW]	21 648,01	19 226,26	19 238,37	21 485,21
S_D^T [MW]	1 562	855	419	771
V_D^T [%]	7,21	4,45	2,18	3,60
-95% [MW]	18 210,82	17 344,85	18 315,34	19 787,67
+95% [MW]	25 085,19	21 107,66	20 161,40	23 182,76
D_{IT}^P [MW]	3 437,19	1 881,40	923,03	1 697,54
V_{IT}^P [%]	15,88	9,79	4,80	7,93

Tab. 3. Wyniki weryfikacji modeli, prognoz punktowych i przedziałowych *ex ante* dla dni od piątku do niedzieli

Szereg	Błąd MAPE [%] dla dób poprzedzających						
	(-3) doby	(-4) doby	(-5) dób	(-6) dób	(-7) dób	(-8) dób	(-9) dób
13 lat	15,11	13,38	11,70	13,69	15,67	11,89	12,65
5 lat	17,13	15,60	13,51	10,90	12,43	13,41	14,40
poniedziałek	6,84	4,98	4,25	3,83	3,68	3,50	3,42
wtorek	5,85	4,32	3,61	3,25	3,06	2,93	2,88
środa	5,73	4,37	3,69	3,44	3,25	3,06	3,00
czwartek	6,01	4,51	3,81	3,48	3,21	3,08	2,98
piątek	8,16	6,21	5,11	4,68	4,29	4,07	3,97
sobota	7,72	5,66	4,69	4,07	3,73	3,44	3,35
niedziela	6,77	4,61	3,94	3,58	3,35	3,24	3,19
święta	20,12	15,48	12,82	11,67	11,12	10,24	9,39

Tab. 4. Błędy MAPE prognoz *ex post* dla wielomianu 2 stopnia na podstawie dób poprzedzających od 3 do 9 dla poszczególnych okresów i typów dni

5. Wyniki symulacji

5.1. Wielomian 2 stopnia

Wyniki symulacji dla wielomianu 2 stopnia dla wartości opóźnionych dobowych (opóźnień dobowych) od 3 do 9 zamieszczono w tab. 4, natomiast dla opóźnień dobowych od 10 do 16 w tab. 5. Wnioski wynikające z obu tablic wskazują, że błędy MAPE *ex post* maleją dla każdego z analizowanych szeregów czasowych wraz ze zwiększeniem liczby wartości opóźnionych dobowych. Największą redukcję błędu MAPE odnotowano dla dni świątecznych (10,73%), następnie dla okresu 5-letniego

(9,13%) i dla 15-letniego szeregu czasowego (8,71%). Odnotowane redukcje dla pozostałych szeregów czasowych kształtują się na poziomie ok. 2,9–4,7%. Najmniejsze redukcje dokładności prognoz w sensie błędu MAPE wynikają z mniejszych wartości odnotowanych dla punktu startowego, czyli wartości opóźnionych 3 dób. Najmniejsze wartości błędów MAPE odnotowano dla wtorków, śród i czwartków, odpowiednio 2,78%, 2,83% oraz 2,89%. Największymi wartościami tych błędów charakteryzowały się dni świąteczne (20,12%), okres 5-letni (17,13%) oraz okres

13-letni (15,67%). Najmniejsze uśrednione arytmetycznie wartości błędów dla wszystkich wartości opóźnionych dobowych odnotowano dla wtorków (3,26%), następnie dla śród (3,33%) oraz dla czwartków (3,40%). Największymi wartościami uśrednionymi charakteryzuje się okres 5-letni (12,11%), dni świąteczne (11,48%) oraz okres 13-letni (11,21%). Opóźnienie dobowe 13 okazało się najkorzystniejsze dla śród i czwartków, natomiast dla piątków oraz dla okresu 5-letniego i 13-letniego opóźnieniem takim było opóźnienie 14. Najskuteczniejsze prognozy dla poniedziałków i sobót okazało się dla opóźnienia dobowego 15, natomiast dla wtorków, czwartków opóźnieniem tym było opóźnienie dobowe 16. Dni świąteczne zyskały najskuteczniejsze prognozy dla opóźnienia dobowego 9.

Wyniki symulacji dla wielomianu 3 stopnia dla opóźnień dobowych od 3 do 9 zamieszczono w tab. 6, natomiast dla opóźnień dobowych od 10 do 16 w tab. 7. Analiza zawartych w tablicach wyników wskazuje, że błędy MAPE *ex post* maleją dla każdego z analizowanych szeregów czasowych wraz ze zwiększeniem liczby wartości opóźnionych dobowych. Największą redukcję błędu MAPE odnotowano dla dni świątecznych (29,85%), następnie dla okresu 5-letniego (21,40%) i dla 13-letniego szeregu czasowego (20,09%). Odnotowane redukcje dla pozostałych szeregów czasowych kształtują się na poziomie ok. 11,9–7,8%. Najmniejsze redukcje dokładności prognoz w sensie błędu MAPE wynikają z mniejszych wartości odnotowanych dla punktu startowego, czyli opóźnienia 3 dób. Najmniejsze wartości błędów MAPE odnotowano dla wtorków, śród i czwartków, odpowiednio 2,73%, 2,76% oraz 2,81%. Największymi wartościami tych błędów charakteryzowały się dni świąteczne (36,30%), okres 5-letni (29,62%) oraz okres 13-letni (27,19%). Najmniejsze uśrednione arytmetycznie wartości błędów dla wszystkich opóźnień dobowych odnotowano dla wtorków (4,57%), następnie dla śród (4,67%) oraz dla czwartków (4,77%). Największymi wartościami uśrednionymi charakteryzuje się okres 5-letni (17,16%), dni świąteczne (15,67%) oraz okres 13-letni (15,05%). Opóźnienie dobowe 15 okazało się najkorzystniejsze dla wszystkich analizowanych szeregów czasowych. Zastosowanie wielomianu 3 stopnia okazało się korzystniejsze w większości analizowanych szeregów czasowych i dotyczyło wszystkich typów dni analizowanych oddzielnie. Największym zwiększeniem skuteczności prognoz pomiędzy wielomianem 2 i 3 stopnia charakteryzowały się dni świąteczne (redukcja błędu o 3,01% na korzyść wielomianu 3 stopnia). Przykładowy wektor ocen parametrów strukturalnych dla środy 16 lipca 2014 roku, wielomianu 2 stopnia oraz 15 wartości opóźnień dobowych, składający się z parametrów a_0 , a_1 , a_2 i a_3 , osiągnął odpowiednio wartości 21 256,98; -145,36; -3,18; 1,21, a osiągnięta wartość błędu MAPE wyniosła 0,31%.

5.2. Ocena *ex ante* vs *ex post*

Oceny *ex ante* z 14–20 lipca 2014 roku, zamieszczone w tab. 2 i 3, porównano z otrzymanymi wygasłymi prognozami

PL

Szereg	Błąd MAPE [%] dla dób poprzedzających						
	(-10) dób	(-11) dób	(-12) dób	(-13) dób	(-14) dób	(-15) dób	(-16) dób
13 lat	11,99	10,74	9,26	7,87	6,96	7,59	8,41
5 lat	13,77	12,36	10,63	9,06	8,00	8,70	9,67
poniedziałek	3,35	3,29	3,26	3,22	3,22	3,21	3,22
wtorek	2,84	2,81	2,84	2,81	2,82	2,81	2,78
środa	2,90	2,84	2,88	2,83	2,86	2,88	2,91
czwartek	2,99	2,94	2,94	2,89	2,91	2,91	2,89
piątek	3,87	3,74	3,65	3,55	3,52	3,52	3,56
sobota	3,36	3,22	3,19	3,11	3,10	3,06	3,08
niedziela	3,13	3,07	3,07	3,05	3,07	3,07	3,09
święta	9,72	10,22	10,48	10,12	10,06	9,82	9,46

Tab. 5. Błędy MAPE prognoz *ex post* dla wielomianu 2 stopnia na podstawie dób poprzedzających od 10 do 16 dla poszczególnych okresów i typów dni

Szereg	Błąd MAPE [%] dla dób poprzedzających						
	(-3) doby	(-4) doby	(-5) dób	(-6) dób	(-7) dób	(-8) dób	(-9) dób
13 lat	–	27,19	21,05	19,91	15,30	12,90	14,10
5 lat	–	29,62	24,30	23,01	16,75	15,19	16,56
poniedziałek	–	12,79	8,24	6,57	5,60	5,08	4,68
wtorek	–	10,82	7,20	5,53	4,76	4,24	3,90
środa	–	10,57	7,17	5,53	4,95	4,42	4,06
czwartek	–	11,04	7,50	5,81	5,10	4,46	4,09
piątek	–	15,17	10,46	8,06	6,86	5,96	5,41
sobota	–	14,37	9,68	7,27	6,11	5,47	4,65
niedziela	–	12,37	7,77	5,98	5,20	4,64	4,24
święta	–	36,30	24,79	20,91	16,92	15,12	14,24

Tab. 6. Błędy MAPE prognoz *ex post* dla wielomianu 3 stopnia na podstawie dób poprzedzających od 3 do 9 dla poszczególnych okresów i typów dni

Szereg	Błąd MAPE [%] dla dób poprzedzających						
	(-10) dób	(-11) dób	(-12) dób	(-13) dób	(-14) dób	(-15) dób	(-16) dób
13 lat	14,69	14,77	14,24	13,60	12,42	7,10	8,32
5 lat	17,18	17,16	16,04	15,40	14,16	8,22	9,51
poniedziałek	4,40	4,23	4,06	3,96	3,84	3,01	3,61
wtorek	3,66	3,53	3,37	3,32	3,25	2,73	3,15
środa	3,88	3,75	3,61	3,49	3,34	2,76	3,22
czwartek	3,79	3,66	3,60	3,52	3,40	2,81	3,28
piątek	5,12	4,97	4,78	4,59	4,45	3,28	4,12
sobota	4,22	4,18	4,03	3,94	3,81	3,03	3,57
niedziela	4,04	3,93	3,80	3,71	3,55	3,03	3,43
święta	12,50	12,11	11,84	10,68	9,99	6,45	11,91

Tab. 7. Błędy MAPE prognoz *ex post* dla wielomianu 3 stopnia na podstawie dób poprzedzających od 10 do 16 dla poszczególnych okresów i typów dni

ex post dla tego samego okresu (tab. 4–7). Najmniejsze wartości błędów MAPE *ex post* dla wielomianu 3 stopnia i opóźnienia 15 zawarte były w 95-proc. przedziale ufności wyznaczonym *ex ante* dla 4 spośród 7 dni tygodnia, tj. dla wtorku, środy, piątku i soboty. W przypadku poniedziałku, czwartku i niedzieli doszło do mniejszego zapotrzebowania niż wartości wyznaczone przez przedział ufności. Zapotrzebowanie na moc szczytową w poniedziałek 14 lipca 2014 roku było o 278,38 MW mniejsze od lewego końca

przedziału ufności. We czwartek 17 lipca 2014 roku rzeczywiste zapotrzebowanie było mniejsze o 1212,85 MW od lewego końca przedziału ufności, natomiast dla niedzieli 20 lipca 2014 roku różnica ta wyniosła aż 2081,16 MW.

Szczegółowe wartości prognoz i uzyskanych błędów MAPE dla analizowanego tygodnia przedstawiają się w sposób następujący: 20 791,28 MW (0,52%), 20 956,61 MW (1,67%), 20 963,20 MW (1,69%), 20 837,55 MW (3,13%), 20 541,70 MW (5,39%), 18 267,16 MW (4,78%),

16 234,18 MW (3,56%). W tym okresie jako zadowalającą precyzję prognoz wygasłych należy uznać jedynie prognozy wyznaczone dla poniedziałku, wtorku i środy.

5.3. Prognozy naiwne

W celu porównania najkorzystniejszych wyników prognoz wygasłych, ocenionych za pomocą miernika MAPE, uzyskanych dla wielomianu 2 stopnia i opóźnienia dobowego 15, opracowano prognozy wygasłe z opóźnieniem tygodniowym. Kolejne dni tygodnia każdorazowo dały wynik korzystniejszy dla prognoz naiwnych. Skuteczność metody naiwnej dla dni od poniedziałku do niedzieli była korzystniejsza odpowiednio o 0,50%, 0,52%, 0,50%, 0,49%, 0,29%, 0,26%, 0,54% w porównaniu z prognozami *ex post*. Dla szeregu czasowego 13-letniego metoda naiwna była skuteczniejsza o 3,51%, a dla 5-letniego o 4,47%. Jedynie w przypadku szeregu czasowego dla dni świątecznych metoda naiwna okazała się mniej skuteczna od metody MNK dla wielomianu 3 stopnia o 0,47%.

6. Wnioski

Przeprowadzone analizy symulacyjne wykazały, że dla klasycznej metody MNK i zastosowania jedynie historycznej informacji o kształtowaniu się poszczególnych szeregów czasowych większą skuteczność uzyskiwanych prognoz wygasłych otrzymuje się przy zastosowaniu wielomianu 3 stopnia w porównaniu z wielomianem 2 stopnia. Zastosowanie wielomianu 2 stopnia dawało najskuteczniejsze prognozy wygasłe dla opóźnień dobowych 9, 13–16. W odniesieniu do wielomianu 3 stopnia wszystkie najskuteczniejsze prognozy uzyskano dla opóźnienia dobowego 15. Uzyskana skuteczność dla wielomianu 2 stopnia była niewiele wyższa dla kolejnych dni tygodnia, odpowiednio: 0,20%, 0,05%, 0,07%, 0,08%, 0,24%, 0,03%, 0,02%. Znaczącą poprawę skuteczności uzyskano jedynie dla dni świątecznych i wyniosła ona 3,01%.

Zastosowanie wielomianu 2 stopnia dało wyniki o mniejszej skuteczności w porównaniu z wielomianem 3 stopnia jedynie dla szeregu czasowego 13-letniego i 5-letniego, odpowiednio o 0,14% i 0,22%, z wykorzystaniem zasady poszukiwania najniższych wartości błędu MAPE dla każdego szeregu czasowego. Bezpośrednie porównanie wartości błędów MAPE pomiędzy wielomianami dla odpowiednich wartości opóźnień wskazuje, że wielomian 2 stopnia daje mniejsze wartości w porównaniu z wielomianem 3 stopnia. Porównanie obu podejść wskazuje jednakże, że zastosowanie wielomianu 3 stopnia oraz 15 wartości opóźnień daje w większości analizowanych szeregów czasowych korzystniejsze wyniki w porównaniu z wykorzystaniem wielomianu 2 stopnia i konieczności poszukiwania najkorzystniejszej wartości opóźnienia czasowego. Zastosowanie metody naiwnej dla opóźnienia 7 dób i porównanie jej wyników z analogicznymi wynikami dla metody MNK i wielomianu 3 stopnia każdorazowo dało wyniki korzystniejsze dla metody naiwnej. Uzyskane korzystniejsze różnice w skuteczności, wyrażonej błędami MAPE dla metody naiwnej dla kolejnych dni

tygodnia, kształtowały się odpowiednio na poziomie: 0,50%, 0,52%, 0,50%, 0,49%, 0,29%, 0,26%, 0,54%. Dla szeregu czasowego 13-letniego i 5-letniego uzyskano skuteczność większą odpowiednio o 3,51% i 4,47%. Uzyskany korzystniejszy wynik dla dni świątecznych o 0,47% na rzecz metody MNK należy uznać za nieznaczący wyjątek na tle pozostałych wyników.

Podsumowując uzyskane wyniki symulacji dla klasycznej metody MNK, należy odnotować, że zastosowanie wielomianu 3 stopnia daje wyższą skuteczność prognoz wygasłych w porównaniu z wielomianem 2 stopnia. Uzyskane wyniki symulacji na tle metody naiwnej oraz na tle skuteczności prognoz uzyskiwanych z wykorzystaniem bardziej wyrafinowanych metod dostępnych w literaturze przedmiotu należy uznać za zbyt niskie do zastosowań praktycznych. Prognozy tego rodzaju mogą stanowić co najwyżej dodatkowy model, który wchodziłby w skład modeli grupowych.

Uzyskana niska skuteczność prognoz wygasłych opisywanych za pomocą błędu MAPE może wynikać z zastosowanego podejścia polegającego na uwzględnieniu jedynie autoregresyjnej informacji o historycznym kształtowaniu się poszczególnych szeregów czasowych. Dalsze badania mogą więc się skupić na uwzględnieniu zmiennych

objaśniających, które pozwolą zwiększyć skuteczność zyskiwanych prognoz. Do zmiennych tych można zaliczyć zmienne opisujące pomiary i obserwacje parametrów meteorologicznych oraz informację o typie dnia. Tym samym możliwe będzie pełniejsze wykorzystanie potencjału oferowanego przez algorytm klasycznej metody MNK.

Bibliografia

1. Prognozowanie w Elektroenergetyce, red. I. Dobrzańska, Wydawnictwo Politechniki Częstochowskiej, Częstochowa 2002.
2. Kopecki K. i in., Analiza i prognoza obciążeń elektroenergetycznych, Wydawnictwa Naukowo-Techniczne, Warszawa 1971.
3. Czyżycki R., Klóska R., Ekonometria i prognozowanie zjawisk ekonomicznych w przykładach i zadaniach, *Economicus*, Szczecin 2011, s. 30–31.
4. Pawłowski Z., Elementy ekonometrii, PWN, Warszawa 1981, s. 186, 209–212, 224–229.
5. Romejko A., Estymacja liniowego modelu wydatków metodą największej wiarygodności, *Wiadomości Statystyczne* 1978, nr 75, s. 11–14.

6. Hellwig Z., O jakości modelu ekonometrycznego. Część II. Estymacja, *Przegląd Statystyczny* 1985, nr 4, Warszawa 1986, s. 297.
7. Kolupa M., Metody estymacji modeli ekonometrycznych, PWE, Warszawa 1974.
8. Milo W., Nieliniowe modele ekonometryczne, PWN, Warszawa 1990.
9. Grabiński T., Wydymus S., Szacowanie liniowych funkcji trendu metodą najmniejszych kwadratów przy pomocy wzorów uproszczonych, *Wiadomości Statystyczne* 1975, nr 4, GUS, Warszawa 1975, s. 20.
10. Ekonometria. Metody i analiza problemów ekonomicznych, red. K. Jajuga, Akademia Ekonomiczna, Wrocław 2002, s. 59.
11. Gruszczyński M., Kuszewski T., Podgórska M., Ekonometria i badania operacyjne, podręcznik dla studiów licencjackich, PWN, Warszawa 2009.
12. Borkowski B., Dudek H., Szczepny W., Ekonometria wybrane zagadnienia, PWN, Warszawa 2003.
13. Stanisław A., Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny, t. 1 i 2, StatSoft Polska, Kraków 2006.

Rafał Czapaj

mgr inż.

PSE Innowacje sp. z o.o.

e-mail: rafal.czapaj@pse.pl

Absolwent Wydziału Elektrycznego Politechniki Śląskiej o specjalizacji: przetwarzanie i użytkowanie energii elektrycznej (2003). Z Grupą Kapitałową PSE SA jest związany od stycznia 2005 roku. W latach 2005–2011 pracował w EPC Consulting SA i zajmował się tematyką rynku energii elektrycznej oraz analiz techniczno-ekonomicznych. Od 2011 roku pracuje w PSE Innowacje sp. z o.o. (wcześniej CATA Ltd.), zajmując się m.in. zagadnieniami prognozowania krótko- i długoterminowego, optymalizacji doboru zbiorów danych wejściowych do modeli prognostycznych, metodami przeszukiwania dużych zbiorów danych, zagadnieniami eksploatacji elementów sieci przesyłowej, robotyzacją prac inspekcyjnych, uczeniem maszynowym i metodami sztucznej inteligencji w energetyce.

Jacek Kamiński

dr hab. inż., prof. IGSMiE PAN

Instytut Gospodarki Surowcami Mineralnymi i Energią Polskiej Akademii Nauk, Kraków

e-mail: kaminski@min-pan.krakow.pl

Profesor nadzwyczajny w Instytucie Gospodarki Surowcami Mineralnymi i Energią Polskiej Akademii Nauk. Kierownik Pracowni Ekonomiki Energetyki oraz Zakładu Polityki i Badań Strategicznych. Uzyskał stopień doktora nauk technicznych na Akademii Górniczo-Hutniczej w Krakowie (2006), a stopień naukowy doktora habilitowanego nauk technicznych w dyscyplinie energetyka na Politechnice Śląskiej w Gliwicach (2013). W latach 2007–2009 oddelegowany ekspert krajowy do Wspólnotowego Centrum Badawczego Komisji Europejskiej – Instytut Studiów Perspektyw Technologicznych (European Commission – DG Joint Research Centre – Institute for Prospective Technological Studies – EC/DG JRC/IPTS) w Sewilli. Autor i współautor ponad 100 publikacji naukowych, w tym m.in. w czasopiśmie: *Energy Policy, Energy, Applied Energy*. Ekspert z zakresu budowy modeli matematycznych oraz systemów wsparcia procesu podejmowania decyzji w przedsiębiorstwach energetycznych, rynków paliw i energii oraz polityki energetycznej. Kierownik kilkudziesięciu prac i projektów zrealizowanych na zlecenie przedsiębiorstw energetycznych. Od 2016 roku jest prezesem zarządu polskiej afiliacji International Association for Energy Economics (IAEE).

Pablo Benalcázar

mgr inż.

Instytut Gospodarki Surowcami Mineralnymi i Energią Polskiej Akademii Nauk, Kraków

e-mail: benalcazar@min-pan.krakow.pl

Asystent w Instytucie Gospodarki Surowcami Mineralnymi i Energią Polskiej Akademii Nauk w Pracowni Ekonomiki Energetyki Zakładu Polityki i Badań Strategicznych. Doktorant Akademii Górniczo-Hutniczej im. Stanisława Staszica w Krakowie. Magister technologii chemicznej ze specjalnością czystych paliw kopalnych i paliw alternatywnych w Akademii Górniczo-Hutniczej. W badaniach naukowych koncentruje się na zagadnieniach ekonomicznych, polityce energetycznej, optymalizacji matematycznej i prognozowaniu.