ECONOMETRIC MODELLING OF COMPOUND CYCLICALITY OF USING TELECOMMUNICATION SERVICES

Paweł Kaczmarczyk

The Mazovian State University in Płock, Płock, Poland
e-mail: p.kaczmarczyk@mazowiecka.edu.pl

ORCID: 0000-0001-5199-6328

Abstract: The aim of this study was to test the effectiveness of an econometric model with dichotomous (binary) explanatory variables in the approximation and prediction of different kinds of cyclicality (annual, monthly, weekly, and daily) of demand for telephone services. The analyses were conducted with the use of data provided by the selected telecommunication network operator. The data included hourly combined demand for specified telephone services (in seconds) of outgoing calls within the framework of the particular subscriber group (business or individual), the given day, the particular month, and the specific category of connection. All kinds of the cyclicality were confirmed in models with 70 explanatory variables (i.e. in models without holidays). The inclusion of the variables set denoting specific holidays improved goodness of the models fit to the data. The econometric modelling of cyclical components and the forecasting of it with the use of dichotomous variables was effective.

Keywords: cyclical components, Decision Support System, Prediction System, short-term forecasting.

1. Introduction

Telecoms operators are functioning in a progressively higher level of competition. It is associated with the development of the "Information Society for All" concept, as well as the liberalisation and harmonisation of the European Union electronic communications market (Begg, 2010; European Commission, 2010; Lisbon European Council, 2000). As a result of achieving the mentioned purposes, telecoms operators strive for the continuous improvement of their decision-making process. In this context, the problem of the reliability of the conducted analytical procedures and the forecast accuracy has an important significance. The forecast accuracy achieved by

telecommunications data analysts has an impact on the quality of the decisions made by management. One of the sources of improvement of the decision-making process is the implementation of an effective Prediction System referred to as PS (Dittmann, 2004). PSs constitute a separate subclass of the Decision Support Systems (DSSs), and enable conducting decision-making processes on the basis of rational procedures and an increase of forecast accuracy.

A PS is very important for the operational planning of an enterprise. In turn, operational management of an enterprise has great significance for the achievement of the objectives defined in the enterprise's strategy (Daft and Marcic, 2011; Griffin, 2015). There is an approach according to which operational management has the greatest impact on enhancing the effectiveness of the growth of the company value (Kasiewicz, 2005). According to another approach, innovation, information and information system are considered as the internal sources of the company value (Marcinkowska, 2000). This is especially linked to the forecasting procedure, because the PS is understood as:

1. a subsystem of an information system,

2. the source of prospective information relating to the external phenomena (the immediate and remote environment of an enterprise) and internal phenomena (sales, costs, receivables, etc.).

The importance of the PS for operational management of a telecommunications operator results from the fact that the PS constitutes, among others, the reliability basis of price calculations, financial planning, and telecommunications network management.

The aim of this research study was to test the effectiveness of the econometric model with dichotomous (binary) explanatory variables in approximation and short--term prediction of different kinds of cyclicality (annual, monthly, weekly, and daily) of demand for telephone services. Therefore, the results of the research study concerned forecasting methods (one of the PS modules). In the subject literature, the forecasting methods are considered as techniques which are used for the approximation of a given variable and the transition from processed data to the forecasting of this variable. At the same time, the formulating of forecasts is executed in accordance with the assumed forecasting principle. The presented analyses were conducted with the use of data provided by the selected telecommunication network operator. The data included hourly combined demand for specified electronic connection services (in seconds) of outgoing calls (from the network of the telecommunications operator) within the framework of a particular subscriber group (business or individual), the given day (e.g. Monday, Saturday, Christmas Day, New Year's Day), the particular month, and the specific category of connection (e.g. mobile networks, calls to the same network). The models were estimated on the basis of data for two years.

The next section of this article is devoted to the literature review of modelling and short-term or mid-term forecasting of hourly demand for connection services. The research methodology applied in this paper, is described in the third section.

The research results obtained by the author are presented in the fourth section. The conclusions and the possibilities of future research are contained in the final section (the fifth section).

2. The literature review of modelling and short-term or mid-term forecasting of hourly demand for telecommunications services

If a forecaster executes short-term or mid-term forecasting, he/she should take into account the quantitative changes which are characterised by an increase or decrease in the value of the forecasting variable in accordance with the previously detected regularity. In this context, the methods of the elimination of deterministic components are a very important step due to the fact that they have a great impact on the forecasts' accuracy (Box, Jenkins, and Reinsel, 1994; Makridakis and Wheelwright, 1989; Makridakis, Wheelwright, and Hyndman, 1998). Factors arising from the calendar have a key importance in this area, however the literature on modelling and short--term or mid-term forecasting of hourly demand for telecommunications services is scarce. There is even a lack regarding such topics as the effectiveness of the forecasting techniques currently used by telecoms operators. This results from the high level of competition observed in the telecommunications services market.

In the literature of modelling and short-term and mid-term forecasting of the hourly demand for telephone services, the models dedicated to such approximation and prediction are called integrated models (Kaczmarczyk, 2016). The idea of the integrated model consists in the fusion of two different techniques of modelling and forecasting in terms of the analysed demand. A linear regression model with binary explanatory variables (which equal 0 or 1) as the first technique was integrated with a feed-forward neural network model, understood as the second technique. The regression model was used for filtering the demand, and the neural network for reflecting the remaining variability of the demand (i.e. the regression residual obtained after the filtration of origin volatility), using the same explanatory variables as in the case of the regression function. The aggregation of the results is executed in the last module. In other words, the values obtained with the use of the regression are corrected by the use of the modelled regression errors. The following segments can be distinguished in the overall understanding of this approach:

1. The first segment: estimation of the linear (multiple) regression model. The regression model enables one to obtain typical demand values for telecommunications services that are generated in many analytical sections (in sections of different categories of connections, diverse groups of subscribers, different types of 24 hours):

$$y_t = a_0 + \sum_{i=1}^{m} a_i x_{it} + \xi_t, \ t = 1, \ 2, \ \ldots, \ n. \tag{1}$$

2. The second segment: calculation of the residual values (i.e. cleaning time series of the response variable):

$$z_t = y_t - \hat{y}_t, \quad t = 1, 2, \ldots, n. \tag{2}$$

3. The third segment: formulation of the demand forecast by using the regression model:

$$y_T^* = a_0 + \sum_{i=1}^{m} a_i x_{iT}^*, \quad T = n+1, n+2, \ldots, n+h. \tag{3}$$

4. The fourth segment: neural modelling and forecasting of residual values of the regression model:

$$z_t = f\left(x_{1t}, x_{2t}, \ldots, x_{mt}\right), \qquad t = 1, 2, \ldots, n. \tag{4}$$

or

$$z_T^* = f\left(x_{1T}^*, x_{2T}^*, \ldots, x_{mT}^*\right), \qquad T = n+1, n+2, \ldots, n+h. \tag{5}$$

5. The fifth segment: correction of values obtained with the use of the regression model by the residuals obtained using the neural model in order to model the original demand/correction of the forecast obtained with the use of the regression model by the prognostic (neural) residuals in order to forecast the demand:

$$d_t = y_t + z_t, \qquad t = 1, 2, \ldots, n. \tag{6}$$

or

$$d_T^* = y_T^* + z_T^*, \qquad T = n+1, n+2, \ldots, n+h. \tag{7}$$

In the empirical part of the quoted concepts, the regression models with 35 binary explanatory variables were constructed as follows:

$$Y_t = \alpha_0 + \sum_{r=1}^{24} \gamma_r H_{rt} + \sum_{i=1}^{3} \beta_i D_{it} + \sum_{j=1}^{6} \delta_j C_{jt} + \sum_{p=1}^{2} \mu_p S_{pt} + \eta_t, \tag{8}$$

$$t = 1, 2, \ldots, 14688,$$

where: Y_t – hourly combined seconds of outgoing calls ($t = 1, 2, \ldots, 14688$); H_{rt} – hours during the day ($r = 1, 2, \ldots, 24$; $t = 1, 2, \ldots, 14688$); D_{it} – the type of day i.e. working 24-hours, Sunday, Saturday ($i = 1, 2, 3$; $t = 1, 2, \ldots, 14688$); C_{jt} – category of connection, namely: mobile networks, local calls to the same network, local calls to other networks, trunk calls, international calls, other connections ($j = 1, 2, \ldots, 6$; $t = 1, 2, \ldots, 14688$); S_{pt} – kind of subscribers group i.e. individual subscribers and business subscribers ($p = 1, 2$; $t = 1, 2, \ldots, 14688$).

The neural network (with 35 inputs and 1 output) is the following:

$$z_t = f\left(H_{1t}, \ldots, H_{24t}, D_{1t}, \ldots, D_{3t}, C_{1t}, \ldots, C_{6t}, S_{1t}, \ldots, S_{2t}\right), \tag{9}$$

$$t = 1, 2, \ldots, 14688,$$

where: z_t – the regression residuals modelled using the neural network model.

In the case of the non-integrated regression, R-square is equal to 0.4748, calculated F statistic is equal to 378.4818 on 35 and 14,652 degrees of freedom, $p < 0.001$. The integration with the neural model caused that the goodness-of-fit of the hybrid model (R square = 0.7149) and the forecast accuracy were significantly improved.

The role of the neural network (i.e. the modelling and forecasting of the regression errors) can also fulfil other methods, for example the SARIMA model (Kaczmarczyk, 2017). Therefore the approximation of the regression errors expressed by the use of equation (8) and prediction of these errors were executed using ARIMA (p, d, q) (P, D, Q)$_s$. The goodness-of-fit of the integrated model (R-square = 0.9480) and the forecast accuracy was also much better than in the case of the non-integrated regression model.

Table 1. The concept of the single-sectional integrated model

Module	Formulas
1. Regression model estimation (module 1):	$y_t = a_0 + \sum_{i=1}^{m} a_i x_{it} + \xi_t, \ t = 1, \ 2, \ \ldots, \ n.$
2. Residuals calculation (module 2):	$z_t = y_t - \hat{y}_t, \ t = 1, \ 2, \ \ldots, \ n.$
3. Prediction of response variable using the regression model (module 3):	$y_T^* = a_0 + \sum_{i=1}^{m} a_i x_{iT}^*, \ T = n+1, \ n+2, \ \ldots, \ n+h.$
4. Modelling and forecasting of the residual with the use of the neural network as autoregressive model (module 4):	$z_t = f\left(z_{t-1}, z_{t-2}, \ldots, z_{t-m}\right), \quad t = 1, \ 2, \ \ldots, \ n.$ $z_T^* = f\left(z_{T-1}, z_{T-2}, \ldots, z_{T-m}\right), \quad T = n+1, \ n+2, \ \ldots, \ n+h.$
5. Results aggregation and forecast accuracy assessment (module 5):	$d_t = y_t + z_t, \quad t = 1, \ 2, \ \ldots, \ n.$ $d_T^* = y_T^* + z_T^*, \quad T = n+1, \ n+2, \ \ldots, \ n+h.$

Source: author's own elaboration on the basis of (Kaczmarczyk, 2018).

Within the framework of the integrated models, two types of models can be distinguished, namely the multi-sectional integrated models described above, and the single-sectional integrated models (Kaczmarczyk, 2018). The idea of the multi-sectional models is to model many analytical sections simultaneously. The distinguished sections are equated with the particular levels of a given classification factor. The classification factors are: the type of 24-hours, the category of connections, and the kind of subscriber groups. In the case of the multi-sectional model, the demand for telecommunications services was modelled in: three types of 24-hours, six categories of connections, and two types of subscriber groups. In turn, the single-sectional model was characterised by the one chosen analytical

section (one level of each classification factor). This model was used to capture the relationship within the framework of a single subscriber group, during a single type of day, and within a single category of connection. This overall approach is shown in Table 1.

In the empirical exemplary of the referred approach, the following formula with 24 binary independent variables was used:

$$Y_t = \alpha_0 + \sum_{r=1}^{24} \gamma_r H_{rt}, \quad t = 1, 2, \ldots, 1248, \tag{10}$$

where: Y_t – hourly counted seconds of outgoing calls to mobile networks generated by business subscribers during working 24 hours – all Wednesdays in the period of a year ($t = 1, 2,\ldots, 1248$); H_{rt} – hours during the day ($r = 1, 2, .., 24; t = 1, 2, \ldots, 1248$).

The neural network is as follows:

$$z_t = f\left(z_{t-1}, z_{t-2}, \ldots, z_{t-24}\right), \qquad t = 25, 26, \ldots, 1248, \tag{11}$$

where: z_t – the regression residuals modelled using the neural network model.

The goodness-of-fit of the single-sectional non-integrated regression model (R-square = 0.9694, calculated F statistics = 1,615.8351 on 24 and 1,223 degree of freedom, $p < 0.001$) is lower than in the case of the single-sectional integrated regression model (R-square = 0.9888).

The analysis of the literature in the explored area encourages one to experiment in terms of other approaches. It is worth mentioning that modelling and forecasting of the demand for telephone services can be carried out using analysis of cyclical components. In the case of hourly data, it is possible to take into account different kinds of cyclicality (annual, monthly, weekly, and daily).

3. Research methodology

In the econometric modelling of periodic seasonality, there are five ways to describe cyclicality using the following types of variables (Kufel, 2010):
1. periodic binary variables (0 or 1) with the intercept,
2. periodic binary variables (0 or 1) without the intercept,
3. centred periodic binary variables (0 or 1),
4. adjusted periodic binary variables (0 or 1),
5. harmonic components.

The first approach was used in this study. If the periodic binary (dichotomous) variable was the variable D_{jt}, it took the value 1 for the period $t - j$ divided by m without rest (m is the number of phases in the cycle). This variable took the value 0 in other periods. If the cycle was equal to 24-hours, then (for hourly data) m was equal to 24. However, if the cycle was equal to e.g. 1 week, then (for the same data) m was equal to 7.

On the basis of hourly data, the cyclical components, i.e. annual, monthly and weekly cycles, were estimated. The evaluation of annual, monthly, weekly and daily cycles based on hourly data was possible using the appropriate sets of binary variables (0 or 1)is presented in Table 2.

Table 2. Possibility of a full and simplified description of individual kinds of cyclicality and the numbers of phases in the cycle

Kind of cycle	Number of phases in cycle (in hours)	
	Full version	Simplified version
Annual	$m = 365 \times 24 = 8760$ $m = 366 \times 24 = 8784$	$m = 12$ (the phase is 1 month)
Monthly	$m = 30 \times 24 = 720$ $m = 31 \times 24 = 744$	$m = 30$ (the phase is 24-hours) $m = 31$ (the phase is 24-hours)
Weekly	$m = 7 \times 24 = 168$	$m = 7$ (the phase is 24-hours)
Daily	$m = 24$	$m = 24$ (the phase is 1 hour)

Source: (Kufel, 2010, p. 113).

The use of the full version of sets of binary variables to describe the cyclical components would require estimating the model with a very large number of such variables. The number of all the explanatory variables would be equal to 9720. The use of variables sets in the simplified version required only 70 binary variables. The simplified approach was used in this study. The model hypothesis can exist in two formulas (Kufel, 2010, p. 113):

$$T_t = \alpha_0 + \sum_{r=1}^{23} \gamma_r D_{rt} + \sum_{i=1}^{6} \beta_i W_{it} + \sum_{j=1}^{30} \delta_j M_{jt} + \sum_{s=1}^{11} \mu_s Y_{st} + \eta_t \tag{12}$$

or

$$T_t = \alpha_0 + \sum_{r=2}^{24} \gamma_r D_{rt} + \sum_{i=2}^{7} \beta_i W_{it} + \sum_{j=2}^{31} \delta_j M_{jt} + \sum_{s=2}^{12} \mu_s Y_{st} + \eta_t, \tag{13}$$

where: T_{rt} – hourly counted time of outgoing calls generated by business subscribers in terms of one chosen category of connection; D_{rt} – daily binary variable which takes value 1 for the hour r during the day and value 0 in other hours during the day; W_{it} – weekly binary variable which takes value 1 for all hours during the day i of the week and value 0 in other hours and days of the week; M_{jt} – monthly binary variable which takes value 1 for all hours during the day j of the month and value 0 in other hours and days of the month; Y_{jt} – annual binary variable which takes value 1 for all hours during the month s of the year and value 0 in other hours and months of the year.

In the research, the second version was applied. In the next stage, the following model with various holidays was tested:

$$T_t = \alpha_0 + \sum_{r=2}^{24} \gamma_r D_{rt} + \sum_{i=2}^{7} \beta_i W_{it} + \sum_{j=2}^{31} \delta_j M_{jt} + \sum_{s=2}^{12} \mu_s Y_{st} + \sum_{l=1}^{11} \eta_l H_{lt} + \eta_t, \tag{14}$$

where: H_{it} – explanatory variable denoting such holidays as: New Year's Day, Easter, Labour Day, Constitution Day, Corpus Christi, Ascension, All Saints', Independence Day, Christmas.

The cycles amplitudes were interpreted as deviations from the average level and fulfilled the following condition:

$$\sum_{j=1}^{m} d_j^* = 0. \tag{15}$$

The cycles amplitudes were calculated according to the formula below:

$$d_j^* = d_j - \overline{d}, \tag{16}$$

where:

$$\overline{d} = \frac{1}{m} \sum_{j=1}^{m} d_j. \tag{17}$$

Verification of the statistical significance of a given type of the cyclicality was carried out by means of the F-test. This was applied to verify the significance of a particular set of variables in the model. The null hypothesis was formulated as follows:

$$H_0: \; d_2 = d_3 = d_4 = ... = d_m = 0. \tag{18}$$

The F-statistics (test statistics), which was applied in the empirical part of the research study (section 3), took the following form (Baltagi, 2008, p. 79; Maddala and Lahiri, 2009, p. 155):

$$F_{emp} = \frac{(RRSS - URSS)/r}{URSS / (n - k - 1)}, \tag{19}$$

where: $RRSS$ – restricted residual sums of squares; $URSS$ – unrestricted residual sums of squares; r – the number of restrictions; $n - k - 1$ – degrees of freedom of the unrestricted model.

For example, when the statistical significance of daily seasonality was tested, then hypothesis (18) for model (13) had the following form:

$$H_0: d_2 = d_3 = d_4 = ... = d_{24} = 0. \tag{20}$$

With this assumption, the number of restrictions was equal to $r = 23$.

4. The research results

Two types of data were used in the empirical research: hourly combined time (in seconds) of outgoing calls generated by business customers to mobile networks, and hourly counted seconds of outgoing calls generated by business subscribers to the same network (i.e. the network of the surveyed operator). The data are presented in the Figure 1.

a) The outgoing connections to mobile networks
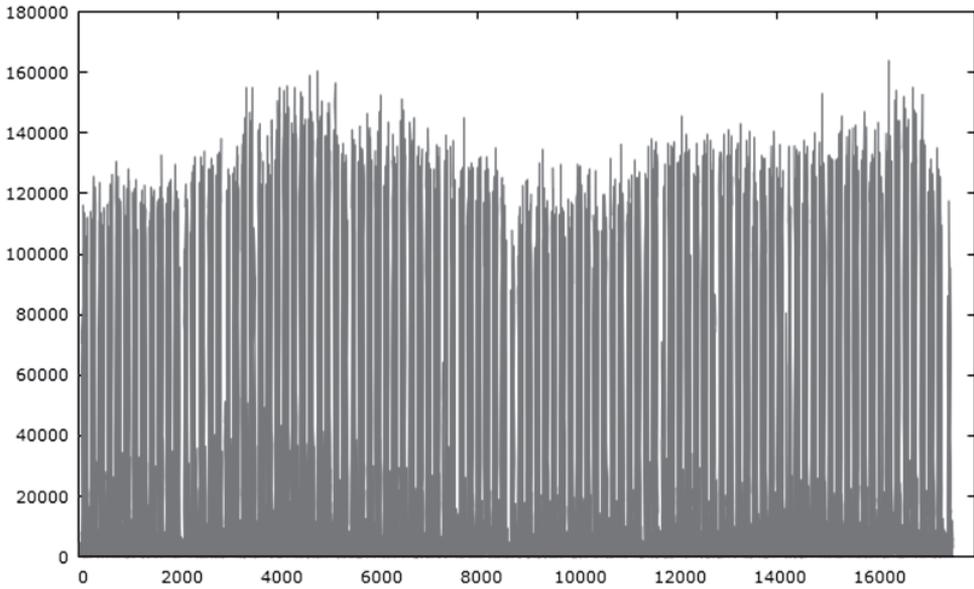
b) The outgoing connections to the same networks
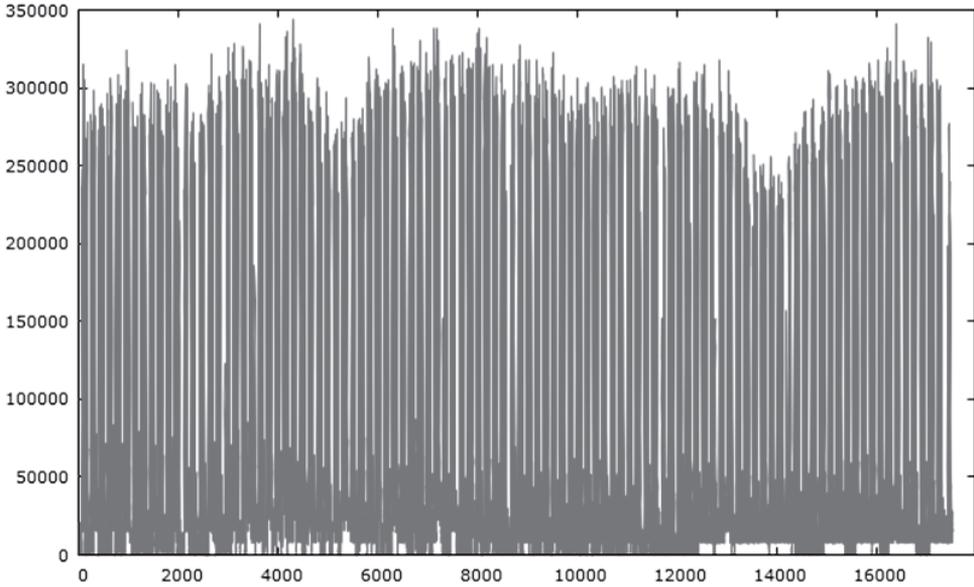
Fig. 1. The course of hourly counted seconds of outgoing calls generated by business subscribers in a period of two years

Source: author's own elaboration.

In the conducted research, formula (13) was used because in the case of this formula, goodness-of-fit was better (in the sense of the number of statistically significant parameters) than in the case of model (12). The number of explanatory variables in model (13) was equal to 70. The model parameters were estimated from the hourly data for a period of two years, i.e. from 1st January of the first year to 31st December of the second year (17,520 cases). The model of demand for connections to mobile networks within the framework of business customers was called model_1(70), and the model of demand for connection services to the same network in the group of business clients was termed as model_2(70). The goodness--of-fit of model (13) is presented in Table 3 and Figure 2.

Table 3. The summary of the regression models with 70 explanatory variables

Coefficient	model_1(70)	model_2(70)
R	0.8638	0.8603
R Square	0.7462	0.7402
Adjusted R Square	0.7451	0.7392
F	732.8933	710.2101
Degrees of freedom	70; 17449	70; 17449
p-value	$p < 0.0001$	$p < 0.0001$
Std. Error of the Estimate	22,019.7080	50,110.9287
ARCH test (p-value)	$p < 0.0001$	$p < 0.0001$

Source: author's own calculations.

a) The autocorrelation of model residuals of connections to mobile network
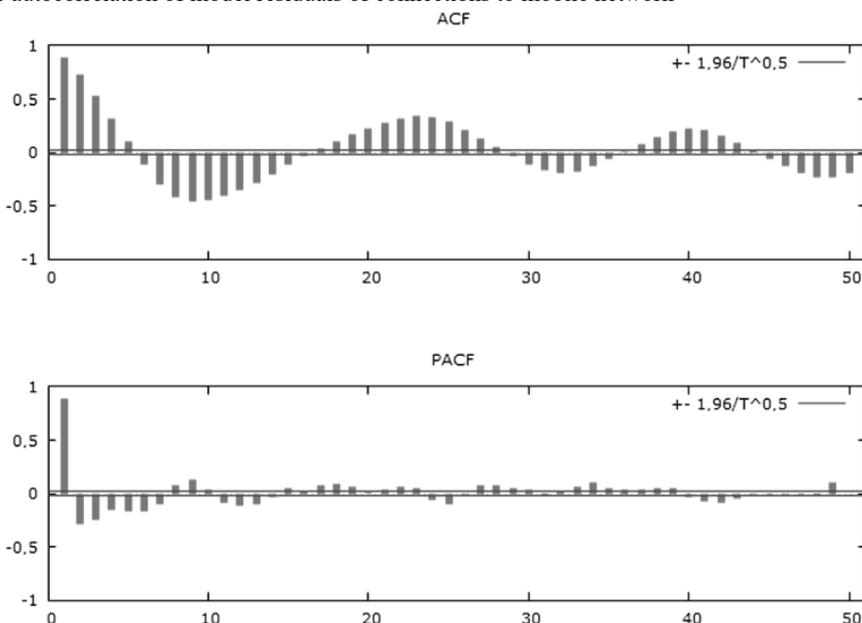
b) The autocorrelation of model residuals of connections to the same network
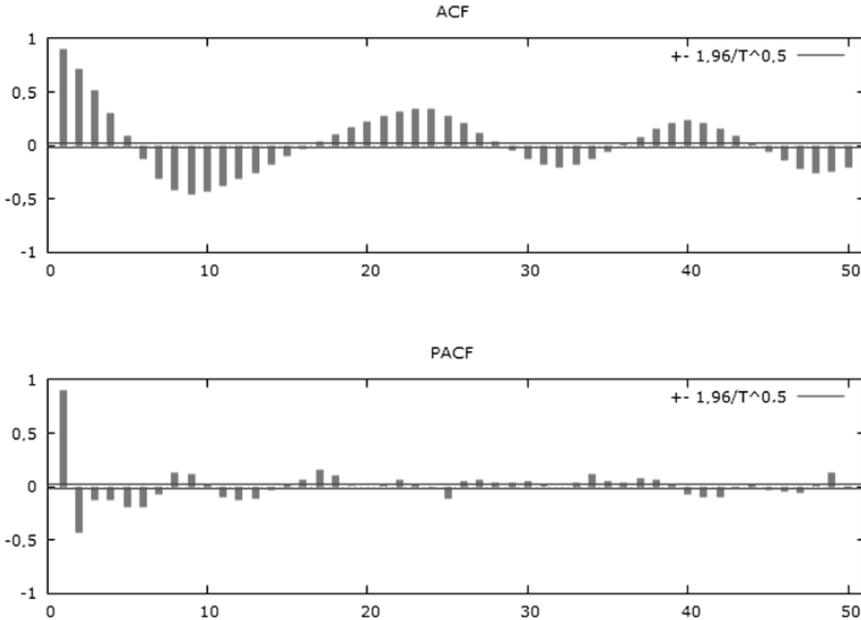
Fig. 2. The ACF and PACF for the model residuals of the demand in the business group

Source: author's own elaboration.

Next, model_1(70) and model_2(70), were used to verify the significance of the particular types of the cyclicality. Formulae (16) to (19) were used. The values of the estimated statistics F_{emp}, together with the critical values of F-statistics at the significance level $\alpha = 0.05$, were as follows:

Table 4. The verification results of statistical significance of all types of the cyclicality

Cycle	F_{emp}	$F(s_1, s_2)$	p-value
		model_1(70)	
Daily	1,815.0260	$F(23; 17,449) = 1.5299$	$p < 0.0001$
Weekly	1,526.2719	$F(6; 17,449) = 2.0991$	$p < 0.0001$
Monthly	4.1251	$F(30; 17,449) = 1.4598$	$p < 0.0001$
Annual	15.0002	$F(11; 17,449) = 1.7892$	$p < 0.0001$
		model_2(70)	
Daily	1,752.8676	$F(23; 17,449) = 1.5299$	$p < 0.0001$
Weekly	1,520.4798	$F(6; 17,449) = 2.0991$	$p < 0.0001$
Monthly	3.9838	$F(30; 17,449) = 1.4598$	$p < 0.0001$
Annual	10.8649	$F(11; 17,449) = 1.7892$	$p < 0.0001$

Source: author's own calculations.

All types of the cyclicality in both models were statistically significant. Therefore, subsequent activities consisted in calculating and interpreting the cycles amplitudes in particular types of the cyclicality. The exemplary results (for connections to mobile networks) are shown in Figure 3.

a) The amplitudes of the daily cycle
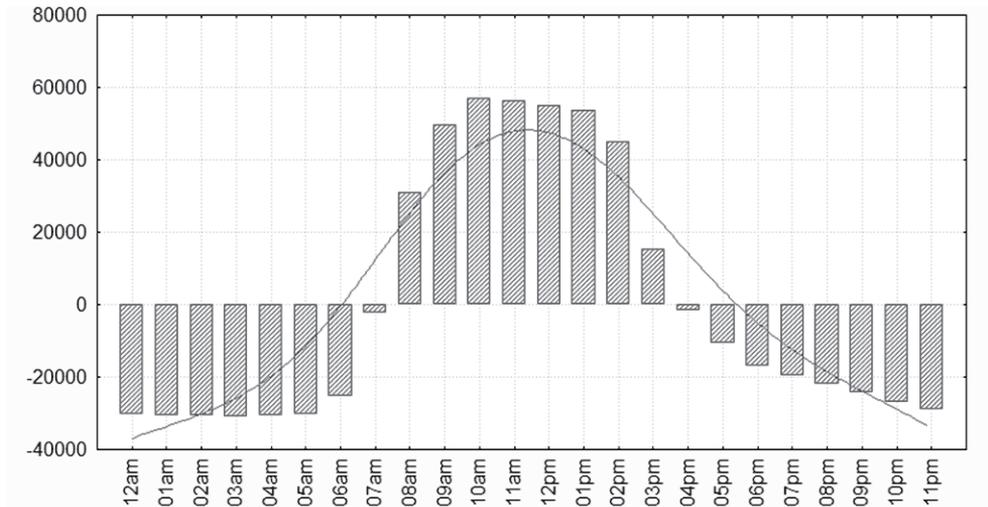
b) The amplitudes of the weekly cycle
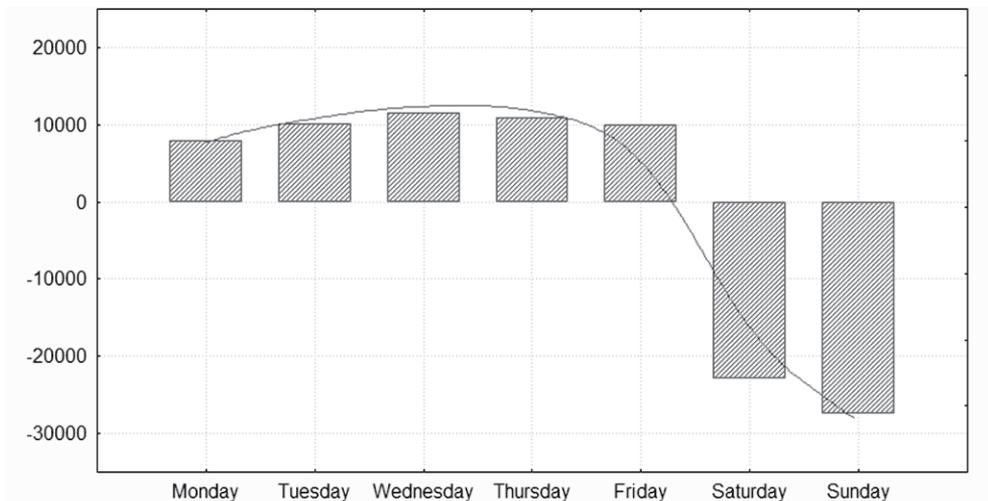
c) The amplitudes of the monthly cycle
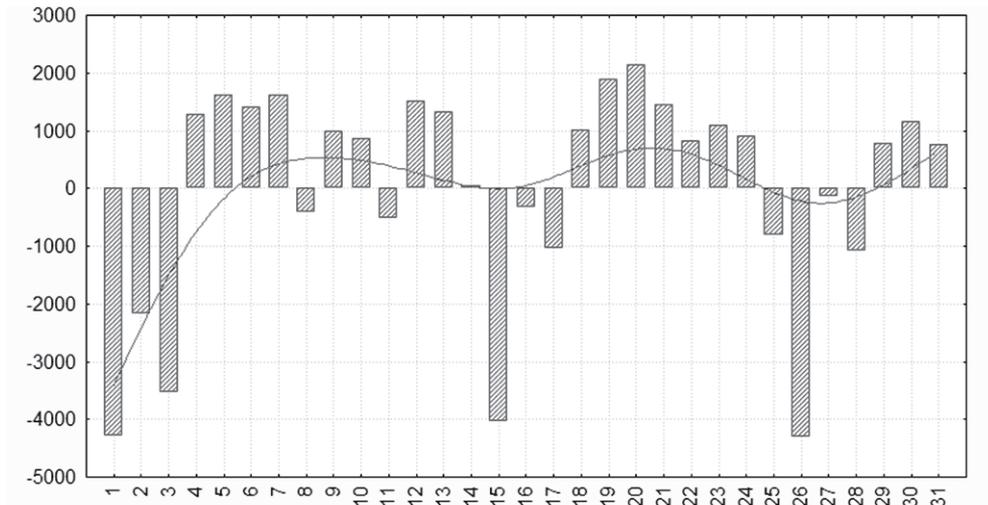
d) The amplitudes of the annual cycle
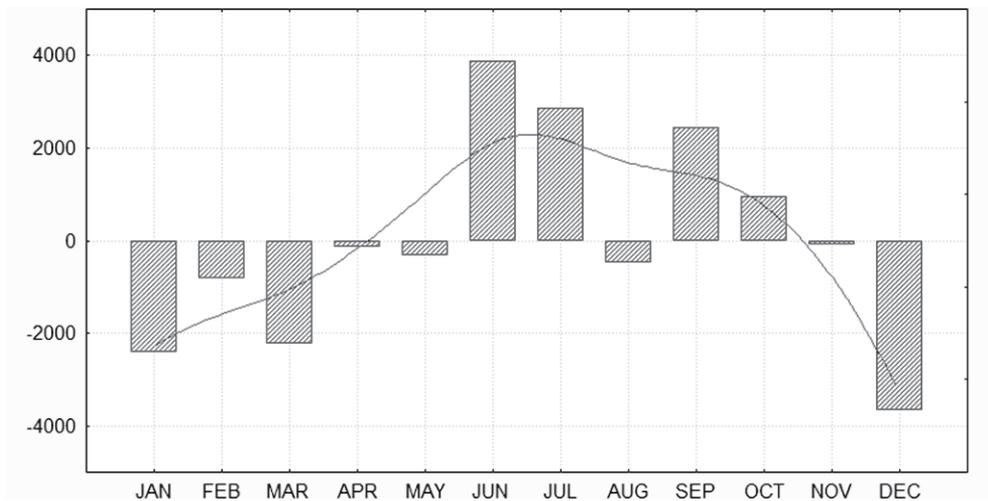
Fig. 3. The amplitudes of all kinds of cycles

Source: author's own calculations and elaboration.

Figure 3 indicates that the highest amplitudes occurred in daily cycles. The night hours were characterised by the lowest demand, and the hours from 8 am to 4 pm characterised by the highest demand. The course of daily demand for telephone services was characterised by the fact that it had only one maximum (first, the values rose systematically and then fell without exception). Therefore, the peak hour was from 10 am to 11 am. The second cyclicality (in terms of the size of the amplitudes)

constituted the weekly cycles. The demand for telephone services during the week was also characterised by having only one maximum. Within the framework of this cyclicality, the days with the highest demand were those in the middle of the week. Wednesday was characterised by the highest demand. On Saturday and Sunday, the demand declined sharply. The lowest demand was on Sunday. Due to the criterion of size of the amplitudes, the next cyclicality was the annual cyclicality. This cycle was dominated by a decline especially in the winter and spring months: December, January, February and March. The highest demand was in June and July. In August there was a clear decrease in demand, which disappeared in September when the level of the demand increased again. This could be in connection with the holiday period. Amplitudes of the monthly cycles were statistically significant, but the smallest from all the kinds of the cycles amplitudes. Within the framework of the monthly cycle, it was noted that positive amplitudes dominated. Until the 3rd day of the month there was a drop in the demand. Then from the 4th day there was an increase. The positive amplitudes persisted (with only a few exceptions) up to the 24th day, after which a drop in the demand occurred over the next four days. A rebound in demand was observed in the last three days.

Next, the effectiveness of the models with various holidays (i.e. with 11 additional explanatory variables) was checked. These models were constructed in accordance with equation (14), so they initially involved 81 explanatory variables. Verification of the statistical significance of particular types of the cyclicality was conducted. The monthly cyclicality proved to be statistically insignificant in both models (Table A1 in Appendix). Therefore, the set of 30 variables describing the monthly cyclicality was eliminated in these models. Finally, the results of the assessment of the goodness-of-fit of models with 51 explanatory variables were obtained, as shown in the table below.

Table 5. The summary of the regression models with 51 explanatory variables

Coefficient	model_1(51)	model_2(51)
R	0.8711	0.8678
R Square	0.7588	0.7531
Adjusted *R* Square	0.7581	0.7524
F	1,077.6153	1,044.6624
Degrees of freedom	51; 17468	51; 17468
p-value	$p < 0.0001$	$p < 0.0001$
Std. Error of the Estimate	21,453.8062	48,825.8154
ARCH test (*p*-value)	$p < 0.0001$	$p < 0.0001$

Source: author's own calculations.

For both model_1 and model_2, the inclusion of additional variables denoting specific holidays, as well as the elimination of the statistically insignificant cyclicality, improved the goodness-of-fit of the model fit. However, residuals of the model were still characterised by autocorrelation.

The forecasting effectiveness of estimated models was then tested. The expired forecasts were verified in the period of the first three months of the next year (2,160 forecasts from 1st January to 31th March). All the forecasting results were justified in Table 6.

Table 6. The comparison of the tested models in terms of forecasting effectiveness

Model	Original forecasts		Adjusted forecasts	
	RMSE	MAE	RMSE	MAE
model_1(70)	19806.0673	15375.5070	17120.7364	11760.6837
model_2(70)	48234.6611	39364.6407	44101.0449	32833.0797
model_1(51)	19426.2272	15436.0735	16393.2977	11555.8726
model_2(51)	47274.7467	39417.1614	42572.3400	32317.0261

Source: author's own calculations.

The visualisation of the original forecasts, formulated with the use of the best fitted model, is presented in Figure 4.

a) The forecasts and empirical values
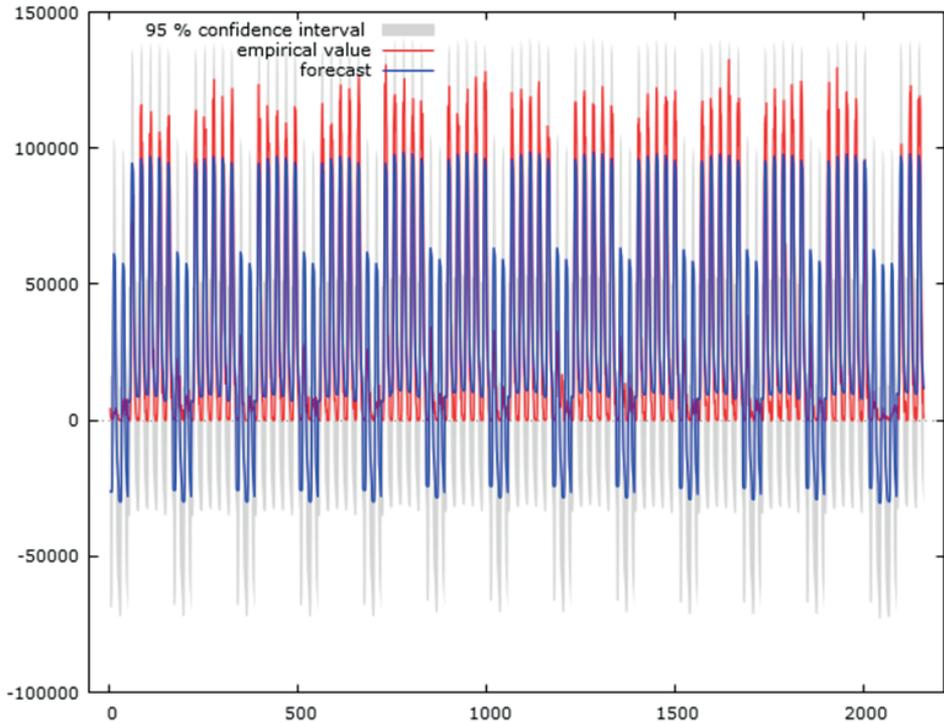
b) The absolute forecasts errors
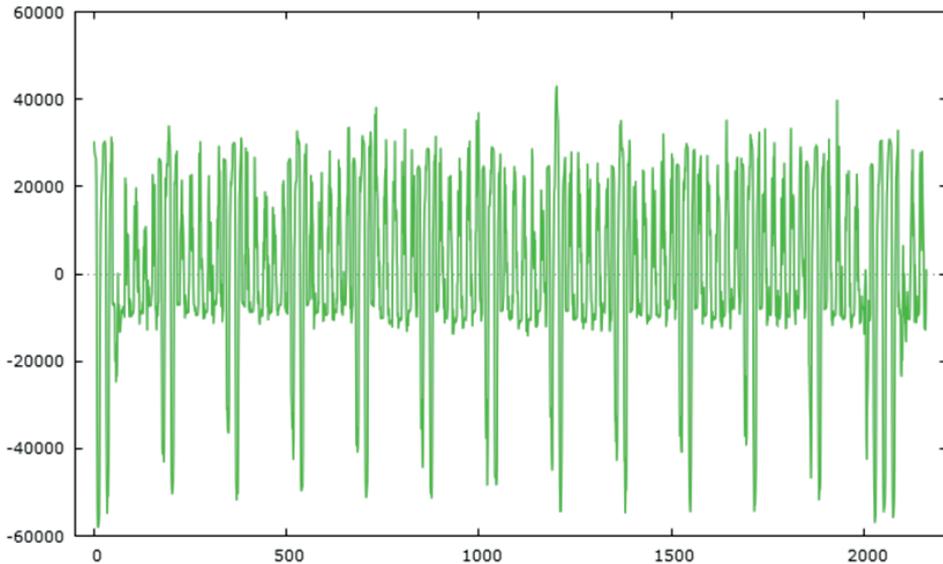
Fig. 4. The forecasting results generated by the use of the model of connection services
to mobile networks which had 51 explanatory variables

Source: author's own calculations and elaboration.

MAE and RMSE were presented for two kinds of forecasts i.e. original and
adjusted. The adjustment of the forecasts consisted in replacing the negative forecasts
with values 0 because the negative demand in selected hours was pointless. Within
the framework of forecasting effectiveness, the best results were achieved by the use
of models without a monthly cycle, i.e. model_1(51) and model_2(51). The forecasting
results for the models of connections to mobile networks were more effective in
comparison to the models of connections to the same network in considering a given
stage of exploration, i.e. with all kinds of the cyclicality or without the monthly cycle
(this was confirmed on the basis of the obtained values of the ratio of RMSE to the
average level of the demand forecasts). The analogical positions of models were
observed due to the comparison of the modelling effectiveness of the demand.

5. Conclusions and possibility of future research

The tested models were difficult to classify to one of the model types of the demand
for telephone services described in the literature (i.e. the multi-sectional or single-
-sectional models). The aim of constructing the models studied in this study was the
approximation of various types of the cyclicality associated with the need to take
into account the demand in different types of days, for example. Such a model

described the demand for telephone services within one connection category, in one subscriber group but in all types of days.

All kinds of the cyclicality were confirmed in the models without variables representing holidays (models with 70 explanatory variables). However, the models with holidays allowed for somewhat better results within the framework of the goodness-of-fit to the data and the forecast accuracy. In these models, one type of the cyclicality was eliminated. The monthly cyclicality proved to be statistically insignificant. Therefore, this kind of models included 51 independent variables.

The obtained goodness-of-fit of the models was definitely higher in comparison to the multi-sectional non-integrated regression model (Kaczmarczyk, 2016). In turn, the results were worse compared to the single-sectional non-integrated regression model (Kaczmarczyk, 2018). However, it is worth mentioning that the approximated data were characterised by higher complexity in this study (the models tested included various types of days), whereas the one-sectional non-integrated regression model was used to model the demand in only one type of day. Hence econometric modelling of cyclicality components with the use of dichotomous variables was effective (R square was higher than 0.70 in all analysed cases).

Due to the fact that the residuals of the tested models were characterised by autocorrelation, attempts to improve the fit of the model to empirical data using ARIMA (p, d, q) $(P, D, Q)_s$ are interesting. The ARIMA model could be used to model the residuals of the regression model. The residuals were also characterised by the occurrence of the ARCH effect, so it is possible and justified to try to reflect them using ARCH (q), which could also cause an increase of the model's fit and the forecast accuracy.

References

Baltagi, B. H. (2008). *Econometrics*. Berlin: Springer.

Begg, I. (2010). Europe 2020 and Employment. *Intereconomics*, *45*(3).

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994). *Time series analysis. Forecasting and control*. Englewood Cliffs: Prentice Hall.

Daft, R. L., and Marcic, D. (2011). *Understanding management*. Mason: South-Western Cengage Learning.

Dittmann, P. (2004). *Prognozowanie w przedsiębiorstwie. Metody i ich zastosowanie*. Kraków: Oficyna Ekonomiczna.

European Commission. (2010). *Europe 2020. A Strategy for Smart, Sustainable and Inclusive Growth*. Communication from the Commission. Brussels. Retrieved July 22, 2017 from http://www.build-up.eu/sites/default/files/content/com2010_2020en01.pdf

Griffin, R. (2015). *Fundamentals of Management*. Boston: Cengage Learning.

Kaczmarczyk, P. (2016). Integrated model of demand for telephone services in terms of microeconometrics. *Folia Oeconomica Stetinensia*, *16*(2). http://dx.doi.org/10.1515/foli-2016-0026

Kaczmarczyk, P. (2017). Microeconometric analysis of the telecommunication services market with the use of SARIMA models. *Dynamic Econometric Models*, *17.* http://dx.doi.org/10.12775/DEM.2017.003

Kaczmarczyk, P. (2018). Neural network application to support regression model in forecasting single-sectional demand for telecommunications services. *Folia Oeconomica Stetinensia*, *18*(2). http://dx.doi.org/10.2478/foli-2018-0025

Kasiewicz, S. (2005). *Budowanie wartości firmy w zarządzaniu operacyjnym.* Warszawa: Szkoła Główna Handlowa w Warszawie.

Kufel, T. (2010). *Ekonometryczna analiza cykliczności procesów gospodarczych o wysokiej częstotliwości obserwowania*. Toruń: Wydawnictwo Naukowe UMK.

Lisbon European Council. (2000). *Presidency Conclusions.* Retrieved June 24, 2017 from http://www.europarl.europa.eu/summits/lis1_en.htm

Maddala, G. S., and Lahiri, K. (2009). *Introduction to econometrics*. Chichester: John Wiley and Sons.

Makridakis, S., and Wheelwright, S. C. (1989). *Forecasting methods for management*. New York: J. Wiley.

Makridakis, S., Wheelwright, S. C., and Hyndman, R. J. (1998). *Forecasting methods and applications*. New York: J. Wiley.

Marcinkowska, M. (2000). *Kształtowanie wartości firmy*. Warszawa: Wydawnictwo Naukowe PWN.

EKONOMETRYCZNE MODELOWANIE ZŁOŻENIA CYKLICZNOŚCI KORZYSTANIA Z USŁUG TELEKOMUNIKACYJNYCH

Streszczenie: Celem artykułu jest zbadanie efektywności modelu ekonometrycznego ze zmiennymi dychotomicznymi (binarnymi) w aproksymacji i prognozowaniu złożonej cykliczności (rocznej, miesięcznej, tygodniowej, dobowej) popytu na usługi telekomunikacyjne. Analizy przeprowadzono w oparciu o dane udostępnione przez wybranego operatora sieci telekomunikacyjnej. Dane obejmują cogodzinnie sumowane zapotrzebowanie na usługi połączeń wychodzących z sieci operatora (w sekundach) w przekroju: wybranej grupy abonentów, poszczególnych dni, kolejnych miesięcy oraz danej kategorii połączenia. Wszystkie rodzaje cykliczności zostały potwierdzone w modelach zawierających 70 zmiennych objaśniających (tzn. w modelach bez świąt). Włączenie do modelu zbioru zmiennych reprezentujących święta poprawiło dopasowanie modelu do danych. Ekonometryczne modelowanie złożenia cykliczności i jego prognozowanie z wykorzystaniem zmiennych dychotomicznych okazało się efektywne.

Słowa kluczowe: komponenty cykliczne, System Wspomagania Decyzji, System Predykcyjny, prognozowanie krótkookresowe.

Appendix

Table A1. The verification results of statistical significance of all types of the cyclicality (81 explanatory variables)

Cycle	F_{emp}	$F(s_1, s_2)$	p-value
model_1(81)			
daily	1,904.4045	F(23; 17,438) = 1.5299	$p < 0.0001$
weekly	1,984.8483	F(6; 17,438) = 2.0991	$p < 0.0001$
monthly	1.4486	F(30; 17,438) = 1.4598	$p = 0.0536$
annual	13.8193	F(11; 17,438) = 1.7892	$p < 0.0001$
model_2(81)			
daily	1,848.0489	F(23; 17,438) = 1.5299	$p < 0.0001$
weekly	1,936.6240	F(6; 17,438) = 2.0991	$p < 0.0001$
monthly	1.3990	F(30; 17,438) = 1.4598	$p = 0.0722$
annual	7.6215	F(11; 17,438) = 1.7892	$p < 0.0001$

$F(s_1, s_2)$ – critical value of F-statistics at the significance level 0.05.

Source: author's own calculations.