

THE MULTIPLE REGRESSION MODEL WITH DICHOTOMOUS VARIABLES IN ANALYSIS OF MULTI-SECTIONAL DEMAND FOR CONNECTIVITY SERVICES – APPROACH BASED ON PER SECOND BILLING

Paweł Kaczmarczyk✉

The Mazovian State University in Płock, Poland

ABSTRACT

The article presents the results of comparative research of the effectiveness of two types of models in terms of approximation and short-term forecasting of the multi-sectional demand for connectivity services. The presented results of the analyses are related to the selection of an appropriate forecasting method as an element of the Prediction System dedicated to telecommunications operators. The first tested model was a multiple regression model with dichotomous explanatory variables. The second model was a multiple regression model with dichotomous explanatory variables and autoregression. In both models, the dependent variable was the hourly counted seconds of outgoing calls within the network of the selected operator. Telephone calls were analysed in terms of such classification factors as: type of day, category of call, group of subscribers. Taking into account all levels of classification factors of the explanatory variable, 35 dichotomous explanatory variables were specified. The defined set of dichotomous explanatory variables was used in the estimation process of both compared regression models. However, in the second model, first-order autoregression was additionally applied. The second model (multiple regression model with dichotomous explanatory variables with first-order autoregression) was found to have higher approximation and predictive capabilities than the first model (multiple regression model with dichotomous explanatory variables without autoregression).

Key words: Prediction System, connectivity services, dichotomous variable, autoregression, forecasting

JEL codes: C31, C53, L96

INTRODUCTION

Statistical research and prediction of the number of hourly counted seconds of outgoing calls generated in individual, separately analysed analytical sections may form a substantive platform for the construction of call tariffs. Such a platform plays the role of a detailed picture of the demand for connectivity services at a selected time in: a specific type of day (working day, Saturday, Sunday, specific holidays), a specific sub-

scriber group (business and individual subscribers), a call category (Internet, mobile networks, local external, local internal, long-distance, international).

The multi-sectional analyses of volumes of connectivity services and the models based on these analyses for the determination of the demand forecasts for these services (Prediction Systems) support the sales planning process. Plans for the volume of services provided in combination with connection tariffs as well as network maintenance costs allow financial

Paweł Kaczmarczyk <https://orcid.org/0000-0001-5199-6328>

✉ p.kaczmarczyk@mazowiecka.edu.pl

and marketing planning [Daft and Marcic 2011, Griffin 2015].

Thus, the aim of this article is to present the results of econometric research in the hourly approach on the demand for telecommunications operator services in sections of hours, days, subscribers, and calls categories. The empirical material was provided by a selected telecommunications operator. The data used for the analyses consisted of hourly counted seconds of outgoing calls generated by business or individual customers during a working day, Saturday, Sunday, or holiday over a period of one year. Various analytical sections in terms of calls categories were also taken into account in the empirical material.

THE ISSUES OF MODELLING THE TELECOMMUNICATIONS MARKET

Modelling and forecasting of the demand for the services of a telecommunications operator requires higher and higher interest. Market mechanisms are being activated in the telecommunications sector and operators are making their own decisions. As a result, commercial decision support systems are placing increasingly higher demands on the forecasting techniques implemented in them.

The configuration of the forecasting model for describing and forecasting telecommunications traffic depends on the forecasting horizon. If the purpose of building a model is a long-term prediction, then in addition to obvious quantitative changes, qualitative changes should also be taken into account. Quantitative changes consist of changes in the value of the explained variable in accordance with the detected regularity, for example the regression function. On the other hand, qualitative changes are the transformations of essential features of the phenomenon, for example, the transformation of the existing regularity, which is expressed by the change of parameters or the functional form of the model. Qualitative changes may also be manifested by the disappearance of previous relationships and the emergence of new ones, that is a change in the set of explanatory variables [Dittmann et al. 2011, Nadolny 2011]. Therefore, the group of phenomena which must be taken into account in the models describing telecommunications traffic in the long (multiannual) period should include:

- the number and structure of subscribers,
- the tendency of customers to increase the demand for services or to make savings,
- changes in the nature of service reception (changes in the daily cycle of telecommunications traffic or changes in the strength of the dependence of demand on other phenomena).

If the purpose of model building is short- to medium-term forecasting, the qualitative changes described above do not occur or occur only in trace amounts. Therefore, it is not necessary to take them into account in the forecasting process. In the case of constructing models for determining short- and medium-term forecasts, the following should be taken into account [Kaczmarczyk 2016, 2017]:

- the type of day (working day, Saturday, Sunday, New Year's Day, etc.),
- the time (06:00:00 to 07:00:00, 07:00:00 to 08:00:00, 08:00:00 to 09:00:00, etc.),
- the category of a call (e.g. to mobile network, long-distance, international),
- type of subscribers (business and individual),
- promotional techniques (e.g. cheaper calls during fixed time periods, free installations).

The scope of the experiments described (later in article) was limited to short- and medium-term forecasting of the demand for the services of a telecommunications operator.

Characteristics of linear regression models with dichotomous variables

Linear regression models with dichotomous variables are used to describe the dependence of qualitative phenomena. Such phenomena are therefore represented in the model by qualitative variables (verbal descriptions of the event state). These variables are introduced into the model as zero-one variables. A zero-one variable is a dichotomous variable because it takes only two values:

- 1 if the event occurs (the object has the given feature),
- 0 if the event does not occur (the object does not have the given feature).

The general form of the tested linear regression model with dichotomous variables can be expressed by a multiple regression formula (causal model) for variables observed in successive time periods [Jain

and Malehorn 2005, Hendry and Nielsen 2007, Verbeek 2012, Wooldridge 2016, Greene 2019]:

$$y_t = \alpha_0 + \alpha_1 x_{1t} + \alpha_2 x_{2t} + \dots + \alpha_k x_{kt} + \varepsilon_t,$$

$$t = 1, 2, \dots, n,$$

where:

y_t – response (dependent) variable,

$x_{1t}, x_{2t}, \dots, x_{kt}$ – explanatory (independent) variables,

$\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_k$ – unknown structural model parameters,

ε_t – model error.

Dichotomous variables can be either explanatory variables or explained ones in the above model. The model verified in the empirical part of the paper uses dichotomous explanatory variables.

The dichotomous variables introduced into the model are treated as quantitative variables. The estimation of parameters of such a model is performed using the classical method of least squares. The assessment value a_i of parameter a_i informs how much the value of the dependent variable (quantitative variable) will change if the explanatory variable x_{it} takes the value of 1, with this change being independent of the development of the other explanatory variables.

The determination of forecasts on the basis of the described model proceeds analogically to the causal model, where independent variables and the dependent variable are quantitative. It is possible to determine the forecast after the prior determination of states of all explanatory variables.

The regression model with dichotomous variables can also be used to filtration process. The filtration process is conducted when there is a need to eliminate deterministic components [Makridakis and Wheelwright 1989, Box et al. 1994, Makridakis et al. 1998].

ESTIMATION AND ASSESSMENT OF THE EFFECTIVENESS OF ZERO-ONE MODELS IN THE CONTEXT OF THE DEMAND FOR TELECOMMUNICATIONS OPERATOR SERVICES

The computational studies described in this section were carried out on the basis of hourly measurements of outgoing seconds over a period of one year, made available by the selected telecommunications operator.

The group of information affecting the demand for telecommunications services included:

- the time of day – classification factor X_1 , (level of classification factor X_{1r} ; 24 levels were defined; $r = 1, 2, \dots, 24$),
- the type of day – classification factor X_2 (level of classification factor X_{2i} ; three variants were distinguished: working day, Saturday, Sunday; $i = 1, 2, 3$),
- the category of a call – classification factor X_3 (level of classification factor X_{3j} ; six levels of this information were identified: mobile networks, local outside the operator's network, internal local, long-distance, international, other; $j = 1, 2, \dots, 6$),
- the type of a customer – classification factor X_4 (level of classification factor X_{4p} ; business or individual subscribers were included; $p = 1, 2$).

In the first of the models tested (in Model 1; Table 1), due to the zero-one representation of the input variables, 35 explanatory variables were distinguished. The dependent variable was the hourly measurements of outgoing seconds (variable Y). Therefore, the following dichotomous explanatory variables: $X_{1,1}, \dots, X_{1,24}, X_{2,1}, X_{2,2}, X_{2,3}, X_{3,1}, \dots, X_{3,6}, X_{4,1}, X_{4,2}$ were included in the regression model as explanatory variables:

$$Y = \alpha_0 + \sum_{r=1}^{24} \gamma_r X_{1r} + \sum_{i=1}^3 \beta_i X_{2i} + \sum_{j=1}^6 \delta_j X_{3j} + \sum_{p=1}^2 \mu_p X_{4p} + Z,$$

$$y_t = \alpha_0 + \sum_{r=1}^{24} \gamma_r x_{1rt} + \sum_{i=1}^3 \beta_i x_{2it} + \sum_{j=1}^6 \delta_j x_{3jt} + \sum_{p=1}^2 \mu_p x_{4pt} + z_t, t = 1, 2, \dots, n,$$

The estimation procedure was carried out using the standard method. The interval on the basis of which the parameters were estimated was the period from January, 1st to February, 20th (14,688 cases), and the interval for verification of forecasts was the period from February, 21st to February, 28th (2,304 cases). The use of such a large learning set was intended to illustrate the complexity of telecommunications traffic. The goodness of Model 1 fit (which was received after the estimation process) is presented in Table 1.

Table 1. Summary of Model 1

Coefficient	Value
Sum of squared errors	4.96e + 13
Standard error of estimation	58177.46
R^2	0.498175
Adjusted R^2	0.497114
$F(31, 14656)$	469.3362
p -value	< 0.000001
Logarithm of the likelihood function	-181971.1
Akaike information criterion	364006.2
Schwarz-Bayesian information criterion	364249.3
Hannan-Quinn information criterion	364087.0
Residuals autocorrelation – rho1	0.906369
Durbin-Watson statistic	0.187254

Source: author’s own calculations.

The number of cases which was used in the estimation process was equal to 14,688. Due to the strong collinearity, the following variables were omitted: $X_{1,24}$, $X_{2,3}$, $X_{3,6}$, $X_{4,2}$ (i.e. 23:00:00 to 00:00:00, Sunday, other connections, a group of individual subscribers). Not all parameters proved to be statistically significant (Table 2).

To fully assess the quality of the model, the autocorrelation of the residuals was also verified and it is shown in Figure 1.

The values of residuals of the verified model are characterised by distinct cyclicality in the daily rhythm. This is certainly determined by the fact that the daily demand patterns are differentiated in various analytical dimensions (a type of day, a subscriber group, a call category). The regression model identifies these runs with accuracy to a constant.

The selected plots of estimation results are presented in Figures 2 and 3.

The presented figures enable assessment of residual values of the regression model (i.e. comparison of calculated values of the response variable with empirical values of the variable) and juxtaposition of the predictor values with the empirical values. So, the regression model fit was not high, moreover speaking the regression model fit was even low (for example high regression errors and high autocorrelation in terms of regression errors).

Table 2. The results of Model 1 estimation

Variable	Structural parameter	Standard error	t -student	p
const	-49222.6	2826.37	-17.42	2.99e-067***
$X_{1,1}$	-2178.0	3325.78	-0.66	0.5126
$X_{1,2}$	-4678.7	3325.78	-1.41	0.1595
$X_{1,3}$	-5221.2	3325.78	-1.57	0.1165
$X_{1,4}$	-5382.4	3325.78	-1.62	0.1056
$X_{1,5}$	-5491.3	3325.78	-1.65	0.0987*
$X_{1,6}$	-5075.8	3325.78	-1.53	0.1270
$X_{1,7}$	-1050.0	3325.78	-0.32	0.7522
$X_{1,8}$	16740.1	3325.78	5.03	4.87e-07***
$X_{1,9}$	54512.5	3325.78	16.39	7.57e-060***
$X_{1,10}$	83468.5	3325.78	25.10	3.91e-136***
$X_{1,11}$	91604.4	3325.78	27.54	7.15e-163***
$X_{1,12}$	88110.9	3325.78	26.49	4.09e-151***
$X_{1,13}$	84610.0	3325.78	25.44	9.48e-140***
$X_{1,14}$	83925.6	3325.78	25.23	1.41e-137***
$X_{1,15}$	76363.5	3325.78	22.96	1.19e-114***
$X_{1,16}$	60080.0	3325.78	18.06	3.65e-072***
$X_{1,17}$	52430.8	3325.78	15.76	1.55e-055***
$X_{1,18}$	52342.3	3325.78	15.74	2.34e-055***
$X_{1,19}$	60507.7	3325.78	18.19	3.70e-073***
$X_{1,20}$	66284.5	3325.78	19.93	3.16e-087***
$X_{1,21}$	53640.4	3325.78	16.13	5.07e-058***
$X_{1,22}$	38474.5	3325.78	11.57	8.10e-031***
$X_{1,23}$	13138.0	3325.78	3.95	7.84e-05***
$X_{2,1}$	29575.9	1281.24	23.08	7.75e-116***
$X_{2,2}$	4091.2	1727.62	2.37	0.0179**
$X_{3,1}$	18129.4	1662.89	10.90	1.43e-027***
$X_{3,2}$	117666.0	1662.89	70.76	0.0000***
$X_{3,3}$	76293.8	1662.89	45.88	0.0000***
$X_{3,4}$	32863.5	1662.89	19.76	8.12e-086***
$X_{3,5}$	-1330.0	1662.89	-0.80	0.4238
$X_{4,1}$	-11936.6	960.07	-12.43	2.61e-035***

***denote significance at 1% level; **denote significance at 5% level; *denote significance at 10% level

Source: Author’s own calculations.

Then, the forecasting process was carried out. The visualisation of forecasts, which were calculated with the use of Model 1, is presented in Figure 4.

The predictive effectiveness of the analysed model was tested using the following measures:

- mean error of *ex post* forecasts, ME (mean error),

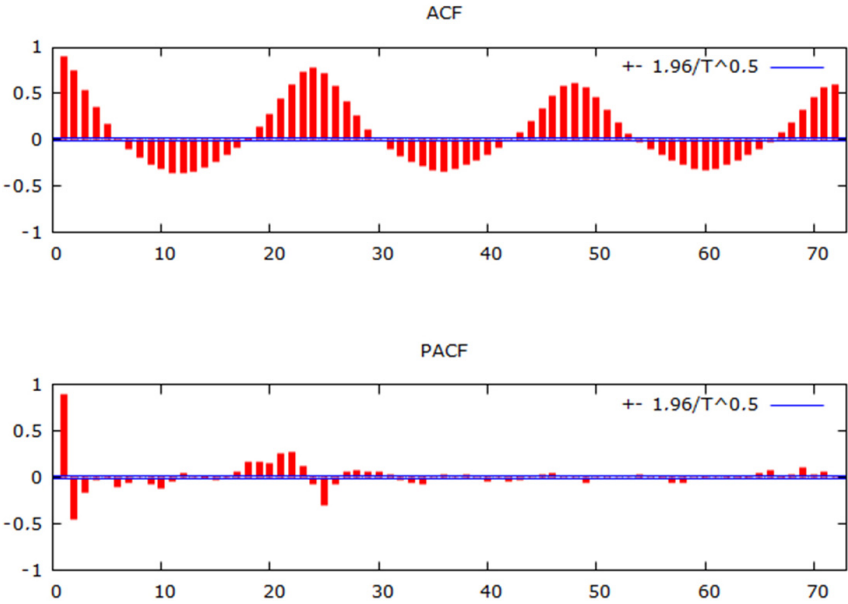


Fig. 1. The ACF and PACF for the residuals of Model 1

Source: Author’s own elaboration.

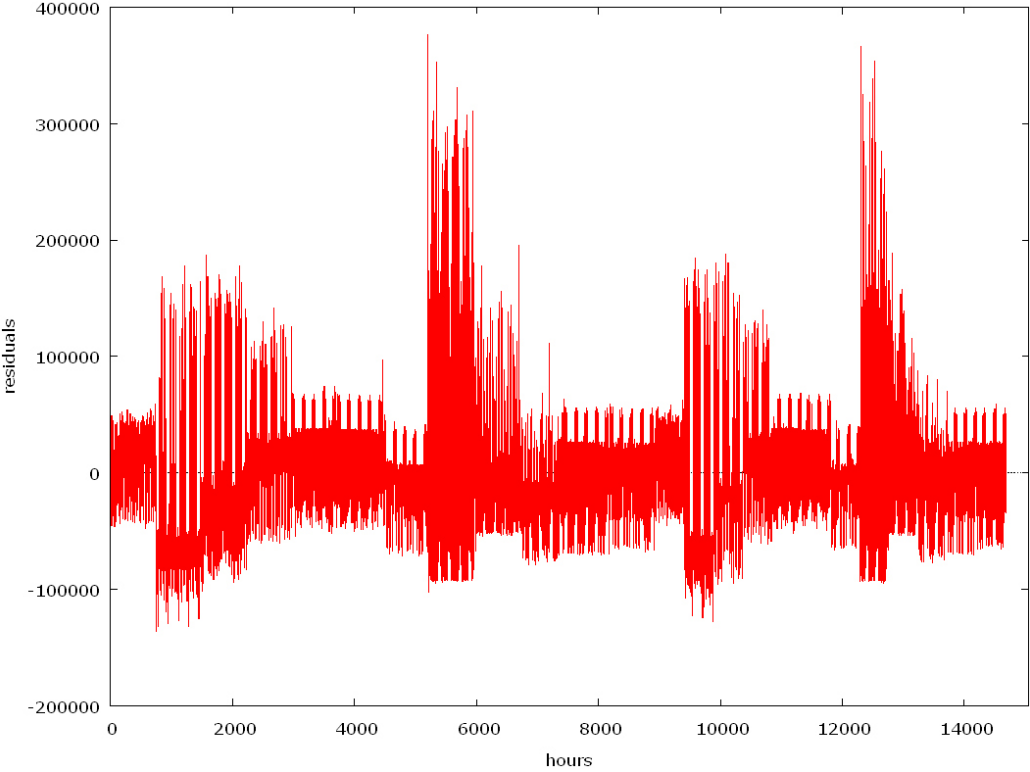


Fig. 2. Residuals of Model 1

Source: Author’s own elaboration.

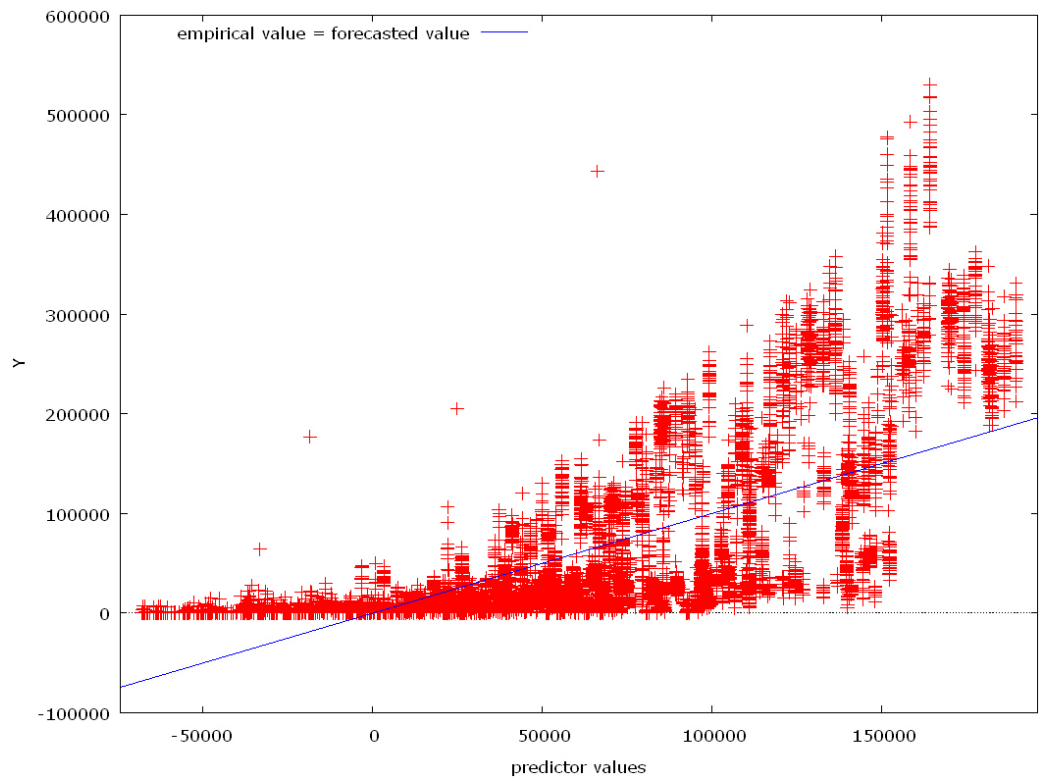


Fig. 3. Empirical and predictor values of Model 1
 Source: Author's own elaboration.

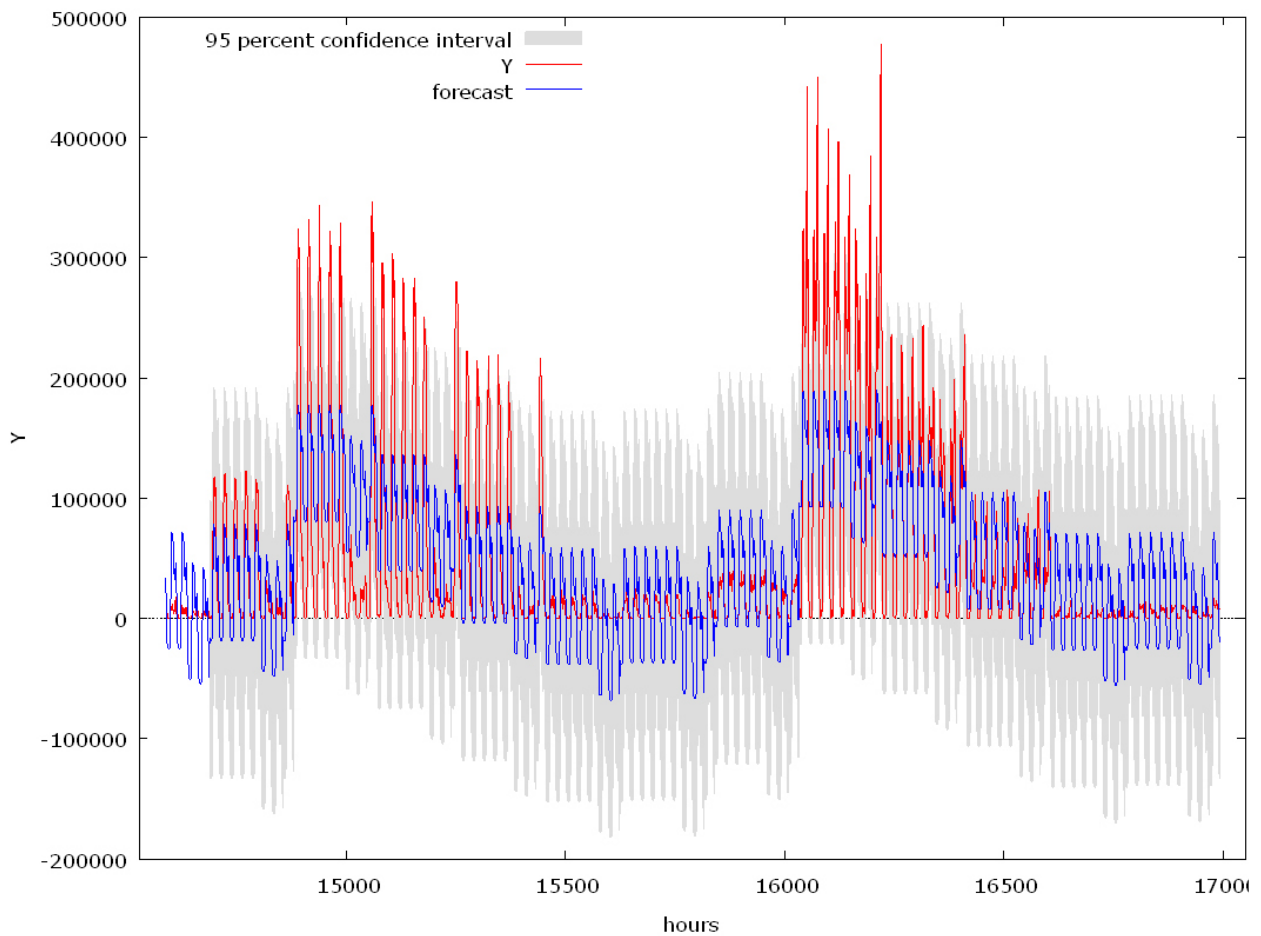


Fig. 4. The forecasting results generated by the use of Model 1
 Source: Author's own elaboration.

- mean absolute error of ex post forecasts, MAE (mean absolute error),
- root mean square of the ex post forecast error, RMSE (root mean square error).

In connection with the adopted measures, the forecast's accuracy was verified. Therefore, ME, MAE, and RMSE were: -391.53, 43,677, and 57,250 respectively.

In order to improve the approximation efficiency of the model and the accuracy of forecasts, the research analysis was based on inclusion of additional explanatory variable y_{t-q} in the estimated equation for y_t . This results from the fact that the correlogram analysis suggests that inclusion of AR(q) model could be appropriate:

$$y_t = \alpha_0 + \varphi y_{t-q} + \sum_{r=1}^{24} \gamma_r x_{1rt} + \sum_{i=1}^3 \beta_i x_{2it} + \sum_{j=1}^6 \delta_j x_{3jt} + \sum_{p=1}^2 \mu_p x_{4pt} + z_t,$$

The estimation process of several models with 35 dichotomous variables and with autoregression was realised. Dichotomous variables were the same as in the case of Model 1. The period, which was used in the estimation, was the same as in the case of Model 1. The best results were achieved with an additional element – first-order autoregression, i.e. AR(1). So, Model 2 was as follows:

$$y_t = \alpha_0 + \varphi y_{t-1} + \sum_{r=1}^{24} \gamma_r x_{1rt} + \sum_{i=1}^3 \beta_i x_{2it} + \sum_{j=1}^6 \delta_j x_{3jt} + \sum_{p=1}^2 \mu_p x_{4pt} + z_t,$$

The selected coefficients used to assess the model fit are juxtaposed in Table 3.

Due to the application of module AR(1) in Model 2, the number of cases in estimation of Model 2 was 14,687. The constructed model is characterised by significantly higher fit compared to Model 1. Four variables were omitted because of strong collinearity. The omitted variables were the following: $X_{1,24}$, $X_{2,3}$, $X_{3,6}$, $X_{4,2}$. Therefore, these variables were the same as

Table 3. Summary of Model 2

Coefficient	Value
Sum of squared errors	8.68e + 12
Standard error of estimation	24333.66
R^2	0.912218
Adjusted R^2	0.912026
$F(32, 14654)$	4758.822
p -value	<0.000001
Logarithm of the likelihood function	-169156.5
Akaike information criterion	338379.0
Schwarz-Bayesian information criterion	338629.6
Hannan-Quinn information criterion	338462.2
Residuals autocorrelation – rho1	0.415358
Durbin-Watson statistic	55.42949

Source: Author's own calculations.

in the case of Model 1. Not all structural parameters were statistically significant in the regression model (Table 4). The calculated significance levels for individual model parameters was higher than 0.1 in the case of variable $X_{1,22}$ (21:00:00–22:00:00), $X_{2,2}$ (Saturday), $X_{3,5}$ (international calls).

Table 4. The results of Model 2 estimation

Variable	Structural parameter	Standard error	t -student	p
const	-16470.2	1188.73	-13.86	2,22e-043***
$X_{1,1}$	9739.2	1392.38	7.00	2,77e-012***
$X_{1,2}$	9232.2	1392.07	6.63	3,43e-011***
$X_{1,3}$	10961.1	1392.42	7.87	3,74e-015***
$X_{1,4}$	11292.5	1392.51	8.11	5,48e-016***
$X_{1,5}$	11330.1	1392.53	8.14	4,40e-016***
$X_{1,6}$	11844.4	1392.55	8.51	1,98e-017***
$X_{1,7}$	15492.9	1392.49	11.13	1,22e-028***
$X_{1,8}$	29626.5	1391.93	21.28	4,95e-099***
$X_{1,9}$	51240.9	1391.12	36.83	3,87e-284***
$X_{1,10}$	45889.7	1398.39	32.82	5,60e-228***
$X_{1,11}$	27726.1	1412.12	19.63	9,64e-085***
$X_{1,12}$	16842.9	1417.23	11.88	2,01e-032***
$X_{1,13}$	16515.1	1414.97	11.67	2,45e-031***
$X_{1,14}$	19010.5	1412.81	13.46	4,99e-041***
$X_{1,15}$	12070.0	1412.40	8.55	1,40e-017***
$X_{1,16}$	2654.8	1408.11	1.89	0,0594*
$X_{1,17}$	9795.3	1400.48	6.99	2,78e-012***

cont. table 4

Variable	Structural parameter	Standard error	t-student	p
$X_{1,18}$	16654.3	1397.67	11.92	1.39e-032***
$X_{1,19}$	24900.0	1397.64	17.82	2.93e-070***
$X_{1,20}$	23260.5	1400.66	16.61	2.26e-061***
$X_{1,21}$	5369.6	1403.13	3.83	0.0001***
$X_{1,22}$	1687.8	1398.08	1.21	0.2274
$X_{1,23}$	-9874.1	1393.81	-7.08	1.46e-012***
$X_{2,1}$	2745.5	545.62	5.03	4.92e-07***
$X_{2,2}$	384.2	722.81	0.53	0.5951
$X_{3,1}$	1657.2	698.41	2.37	0.0177**
$X_{3,2}$	10800.4	805.60	13.41	9.61e-041***
$X_{3,3}$	6993.1	743.81	9.40	6.15e-021***
$X_{3,4}$	3012.0	704.74	4.27	1.93e-05***
$X_{3,5}$	-126.1	695.55	-0.18	0.8561
$X_{4,1}$	-1093.8	403.69	-2.71	0.0067***
Y_{t-1}	0.9	0.00	262.93	0.0000 ***

*** denote significance at 1% level; **denote significance at 5% level; *denote significance at 10% level

Source: Author's own calculations.

The autocorrelation function and the partial autocorrelation function of the regression model residuals are presented in Figure 5.

The analysis indicates a significantly lower autocorrelation of the residuals compared to Model 1. However, the observed values of correlation coefficients and Q Box and Ljung statistics are still quite significant. This creates opportunities for further attempts to improve the quality of the tested tool.

The plots (Figures 6 and 7) showing the extent to which the model describes the data are presented below.

The residuals of Model 2 were not as high as previously. Therefore, the estimated values and the empirical values were not so different as in the case of Model 1. The predictor values were also closer to the estimated line than in the case of Figure 3. These properties confirm the conclusion about a better fit of Model 2.

Then the forecasts with the use of Model 2 were calculated and verified. The forecast accuracy of Model 2 was tested in the same period as in the case of Model 1. The visualisation of forecasts, which were formulated using Model 2, is shown in Figure 8.

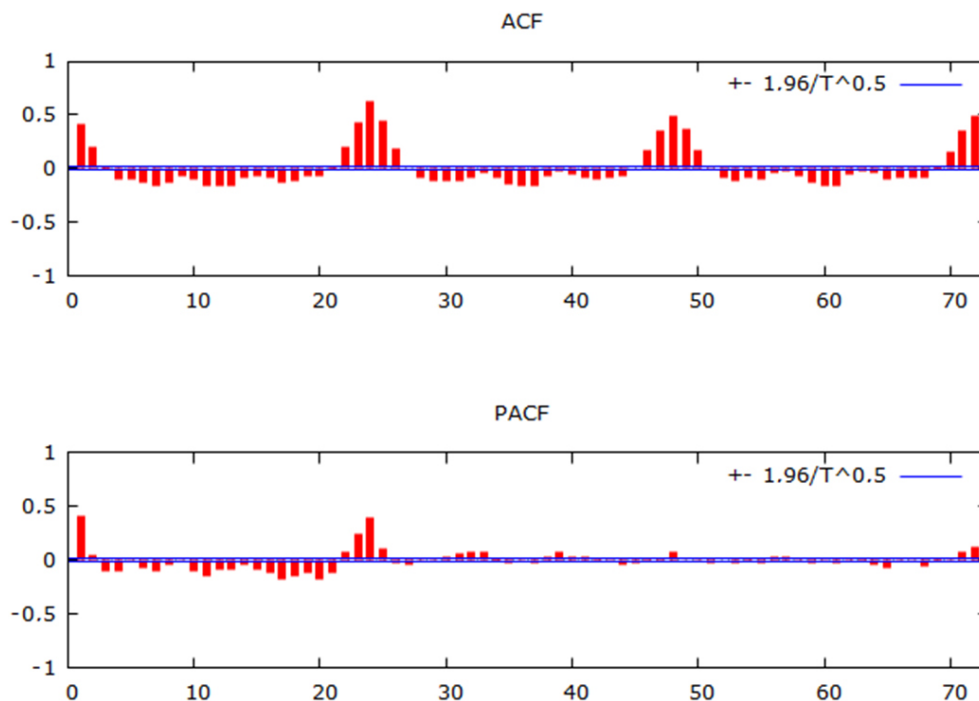


Fig. 5. The ACF and PACF for the residuals of Model 2

Source: Author's own elaboration.

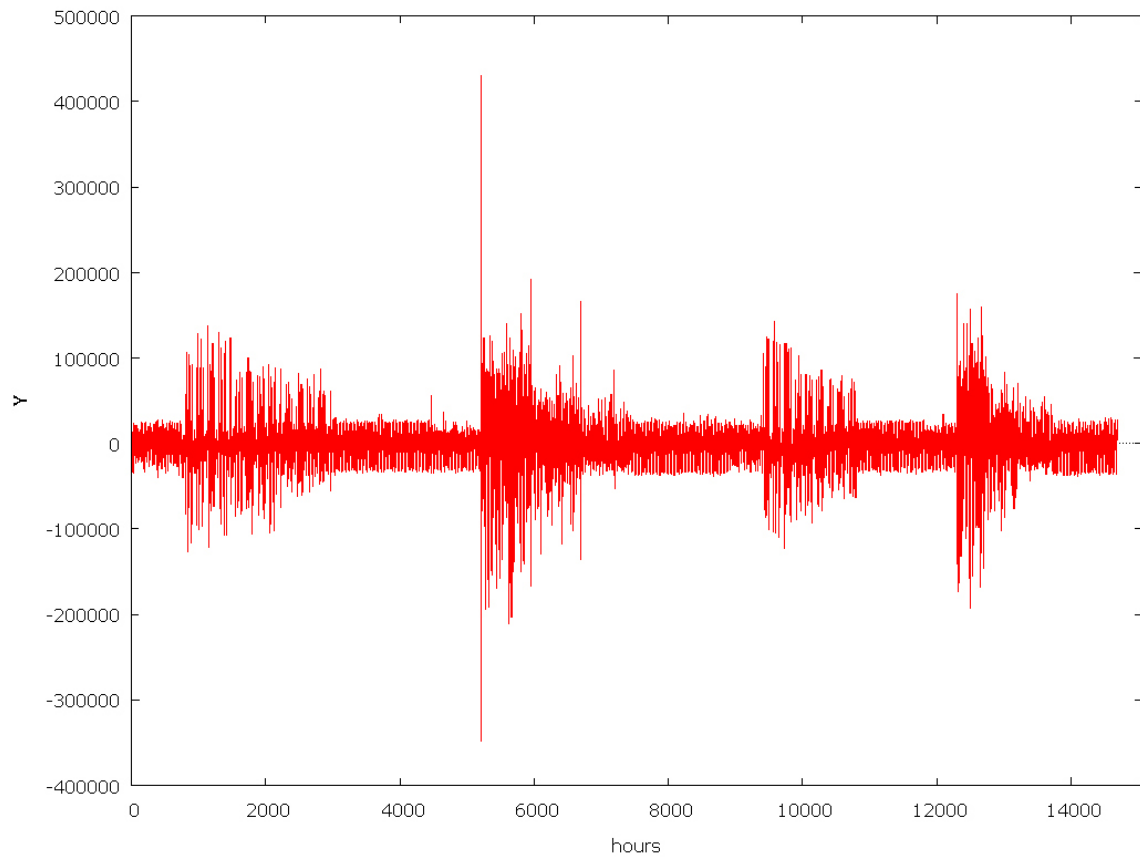


Fig. 6. Residuals of Model 2

Source: Author's own elaboration.

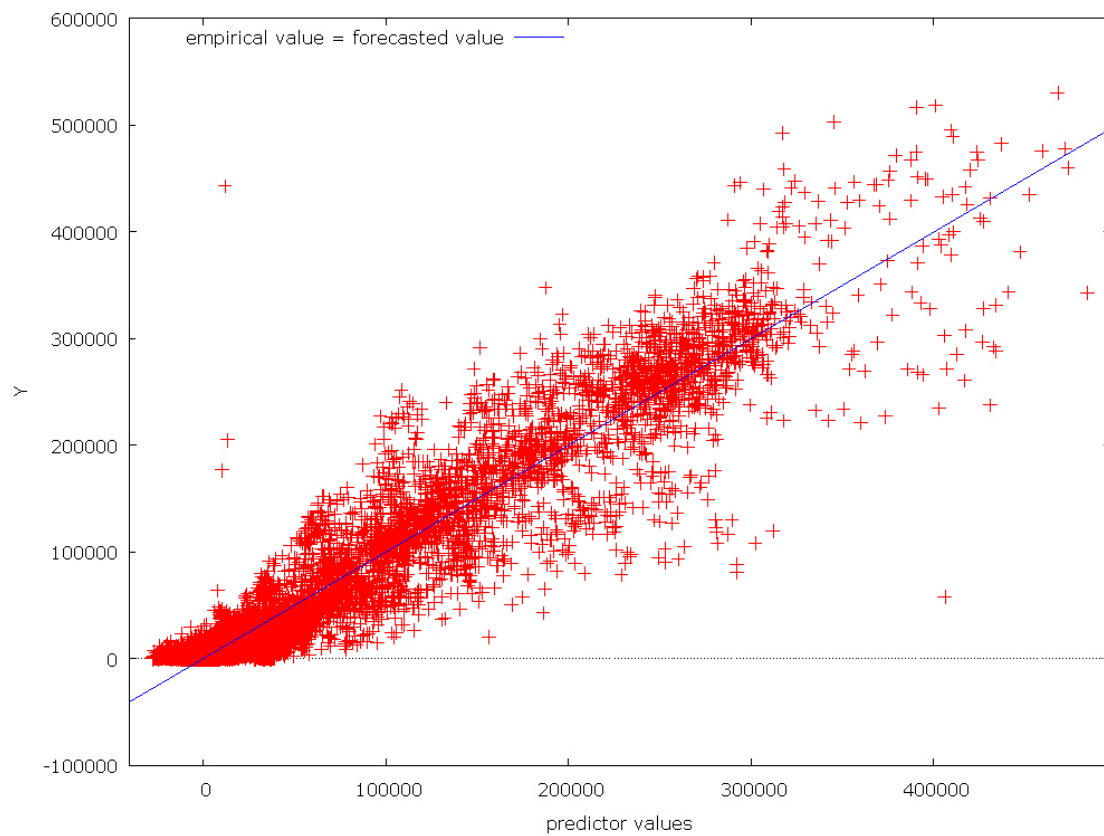


Fig.7. Empirical and predictor values of Model 2

Source: Author's own elaboration.

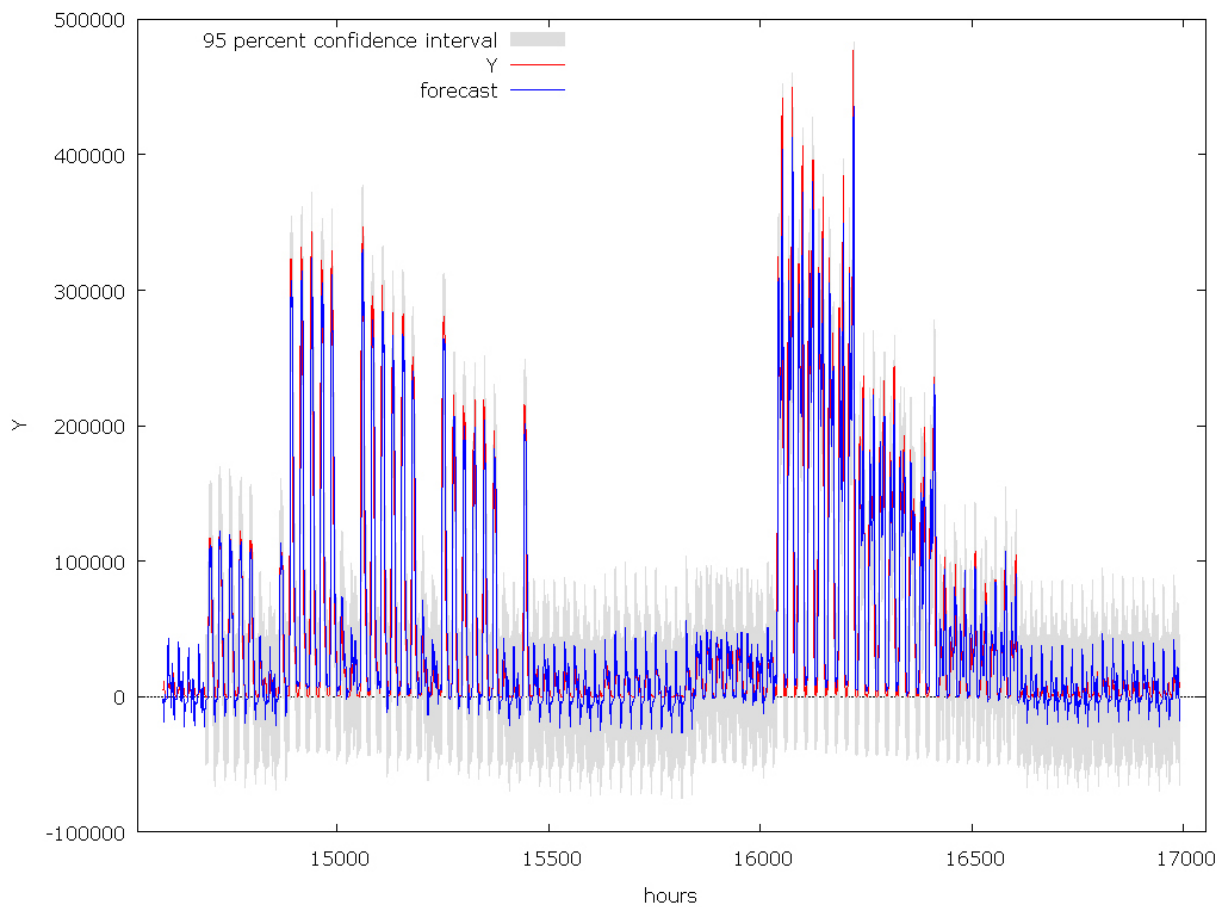


Fig. 8. The forecasting results generated by the use of Model 2

Source: author's own elaboration.

ME, MAE, and RMSE values determined on the basis of forecasts made with the use of Model 2 amounted to -34.736 , $14,583$, and $24,127$ respectively. This argues much higher accuracy of forecasts with respect to Model 1.

FINAL CONCLUSIONS

The linear regression model describing outgoing telecommunications traffic with the use of 35 dichotomous explanatory variables (concerning the type of day, the call category, the subscriber group) – Model 1, is characterised by lower predictive efficiency in comparison with linear regression models with the same explanatory variables and an additional element – first-order autoregression (Model 2). The goodness

of Model 1 fit to the empirical data is absolutely not satisfactory, which is indicated by the coefficients of its quality as well as a high autocorrelation of the residuals. This results from an attempt to fit a single model to all analytical dimensions of telecommunications traffic. Much better results (in the learning and testing interval) were obtained by using Model 2 to represent the demand course in the same analytical sections as in the case of Model 1. The values of average errors of expired forecasts made with this tool are definitely lower than the values of these errors obtained for Model 1.

It is therefore reasonable to further attempt to improve the predictive performance of Model 2 by separating demand information into different day types, subscriber groups and call categories. It wo-

uld then be possible to use independent models to separately represent the isolated dependencies. This would make it possible to focus the capabilities of individual models on describing only specific parts of the reality explored.

REFERENCES

- Box, G.E.P., Jenkins, G.M., Reinsel, G.C. (1994). *Time Series Analysis. Forecasting and Control*. Prentice Hall, Englewood Cliffs.
- Daft, R.L., Marcic, D. (2011). *Understanding Management*. South-Western Cengage Learning, Mason.
- Dittmann, P., Szabela-Pasierbińska, E., Dittmann, I., Szpula, A. (2011). *Prognozowanie w zarządzaniu sprzedażą i finansami przedsiębiorstwa*. Wolters Kluwer Polska, Warszawa.
- Greene, W.H. (2019). *Econometric Analysis*. Pearson Education Limited, London
- Griffin, R. (2015). *Fundamentals of Management*. Cengage Learning, Boston.
- Hendry, D. F., Nielsen, B. (2007). *Econometric Modeling: A Likelihood Approach*. Princeton University Press, Princeton.
- Jain, Ch.L., Malehorn, J. (2005). *Practical Guide to Business Forecasting*. Graceway Publishing Company, Inc., New York.
- Kaczmarczyk, P. (2016). Integrated Model of Demand for Telephone Services in Terms of Microeconometrics. *Folia Oeconomica Stetinensia* 16, 72–83, <https://doi.org/10.1515/fofi-2016-0026>
- Kaczmarczyk, P. (2017). Microeconomic Analysis of Telecommunication Services Market with the Use of SARIMA Models. *Dynamic Econometric Models* 17, 41–57, <https://doi.org/10.12775/DEM.2017.003>
- Makridakis, S., Wheelwright, S.C. (1989). *Forecasting Methods for Management*. J. Wiley, New York.
- Makridakis, S., Wheelwright, S.C., Hyndman, R.J. (1998). *Forecasting Methods and Applications*. J. Wiley, New York.
- Nadolny, M. (2011). Podstawowe modele dyfuzji dóbr telekomunikacyjnych. [In:] J. Łyka (Ed.), *Wybrane modele matematyczne w ekonomii. Globalizacja i rozwój*. Wydawnictwo Uniwersytetu Ekonomicznego, Wrocław, 70–108
- Verbeek, M. (2012). *A Guide to Modern Econometrics*. John Wiley & Sons, Ltd, Chichester.
- Wooldridge, J.M. (2016). *Introductory Econometrics. A Modern Approach*. Cengage Learning, Boston.

MODEL REGRESJI WIELOKROTNEJ ZE ZMIENNYMI DYCHOTOMICZNYMI W ANALIZIE WIELOPRZEKROJOWEGO POPYTU NA USŁUGI POŁĄCZENIOWE – PODEJŚCIE OPARTE NA SEKUNDOWYM NALICZANIU

STRESZCZENIE

W artykule przedstawiono wyniki porównawczych badań efektywności dwóch rodzajów modeli w zakresie aproksymacji oraz krótkookresowego prognozowania wieloprzekrojowego popytu na usługi połączeniowe. Zaprezentowane wyniki analiz są więc związane z doбором odpowiedniej metody prognozowania jako elementu systemu prognostycznego kierowanego do operatorów telekomunikacyjnych. Pierwszym badanym modelem był model regresji wielokrotnej z dychotomicznymi zmiennymi objaśniającymi, a drugim modelem był model regresji wielokrotnej z dychotomicznymi zmiennymi objaśniającymi z autoregresją. W obu modelach zmienną objaśnianą były sumowane co godzinę liczby sekund rozmów wychodzących z sieci wybranego operatora. Połączenia telefoniczne były analizowane pod względem takich czynników klasyfikacyjnych jak: typ doby, kategoria połączenia, grupa abonentów. Biorąc pod uwagę wszystkie poziomy czynniki klasyfikacyjnych zmiennej objaśnianej, wyszczególniono 35 dychotomicznych zmiennych objaśniających. Zdefiniowany zestaw dychotomicznych zmiennych objaśniających wykorzystano w procesie estymacji obu porównywanych modeli regresji. Jednakże w drugim modelu zastosowano dodatkowo autoregresję rzędu pierwszego. Stwierdzono, że drugi model (model regresji wielokrotnej z dychotomicznymi zmiennymi objaśniającymi z autoregresją rzędu pierwszego) charakteryzuje się wyższymi możliwościami aproksymacyjnymi i predykcyjnymi niż model pierwszy (model regresji wielokrotnej z dychotomicznymi zmiennymi objaśniającymi bez autoregresji).

Słowa kluczowe: system prognostyczny, usługi połączeniowe, zmienne dychotomiczne, autoregresja, prognozowanie