MATHEMATICAL MODEL OF TRAVEL TIMES RELATED TO A TRANSPORT CONGESTION: AN EXAMPLE OF THE CAPITAL CITY OF POLAND – WARSAW

Anna Borucka
Military University of Technology, Warsaw, Poland
E-mail: anna.borucka@wat.edu.pl

Szymon Mitkow
Military University of Technology, Warsaw, Poland
E-mail: mitkow@wat.edu.pl

Abstract

The contemporary explosion of urbanization processes, dynamic urban development, as well as intensive growth of economic activity and constantly growing demand result in an increased demand for transport services. The direct consequence of these phenomena is emergence of transport congestion. This problem affects many cities around the world. It is an undesirable phenomenon and has an adverse impact on the flow of traffic, causing problems not only for inhabitants, but also for the development of economy, as well as disruptions in the supply chain. Moreover, it is a cause of increasing pollution levels in cities and excessive energy consumption.

The article proposes a detailed analysis of the phenomenon of transport congestion based on empirical studies carried out on a selected section in the capital city of Poland – Warsaw. Inspired by sustainable transport paradigm, the real transport congestion level and chosen reasons for its occurrence along the studied route were identified. Then, on the basis of selected mathematical methods and tools, this phenomenon was analyzed and mathematical models were proposed. At first, a multiple regression (MR) and influence of such factors as day of the week or holidays was used, and then more advanced econometric model of ARIMA was used. The adequacy of both models was finally compared. We also propose selected solutions to increase capacity, which can be adopted in most cities.

An aim of this article was not only to present mathematical analysis of phenomenon and to identify factors that affect it, but also to present negative social, economic and environmental effects.

The author's intention was also to stress out that achieving satisfactory results in limiting transport congestion requires constant and effective shaping of social attitudes and permanent changes in the way of thinking, to which the presented article also contributes.

Key words: congestion, multiple regression (MR) model, ARIMA model

1. INTRODUCTION - PHENOMENON OF TRANSPORT CONGESTION

The progress of civilization is shown in many aspects, including the development of widely understood transport. However, not only the possibilities, but also the communication requirements are growing. An increasing number of traffic users and motor vehicles lead to congestion in the road network and disturbances to free movement, resulting in transport congestion. There are many definitions of this phenomenon. A. Altshuler defines congestion as a condition in which demand for a given infrastructure object prevents free movement, at the maximum permitted speed of traffic (Altshuler, 1979). J. Rothenberg, on the other hand, refers to a situation where more than one buyer applies for certain goods, which cannot be provided in the form of separate units (Rothenbarg, 1970). Similarly, M. Ciesielski says about congestion occurring when demand for objects of transport infrastructure or transport services exceeds the possibilities of its efficient service (Ciesielski, 1986).

The congestion concerns practically every progressive city in the world, which shows its scale and difficulty in finding a solution to this problem. The latest reports from INRIX (INRIX, 2018), which sum up the past year, show that in the most crowded city in the world, Los Angeles, the driver spent on average 102 hours in traffic. The survey covered 1360 cities on five continents (except Australia), and the top ten are presented in Table 12.

Table 1. The most crowded cities in the world according to INRIX

No.	Country	City	Hours spent on crowded streets (% of changes in the ratio to 2016)	% of the total driving time in congestion conditions
1	USA	Los Angeles	102 (-2%)	12%
2	Russia	Moscow	91 (+2%)	13%
3	USA	New York City	79 (-5%)	12%
4	Brazil	Sao Paulo	70 (-1%)	10%
5	USA	San Francisco	64 (-2%)	9%
6	Colombia	Bogota	63 (+3%)	11%
7	Britain	London	60 (+3%)	14%
8	Russia	Magnitogorsk	57 (0%)	10%
9	Russia	Yurga	55 (0%)	12%
10	USA	Atlanta	54 (-8%)	6%

Source: INRIX, 2018.

In the survey, Poland was ranked 122nd in the world ranking with the city of Cracow, and 131st in the ranking with the capital city of Warsaw. Among the most crowded cities in Europe, Cracow has 66 positions, and Warsaw 71. Moreover, there were such Polish cities as Poznań, Szczecin, Wrocław, Gdańsk, Łódź, Katowice (INRIX, 2018). Solving the congestion problem requires involvement of all elements of the transport system, but even the best solutions will not achieve their full potential without the right approach of traffic participants, which is why it is so important to

build their awareness. It is not only necessary to highlight the negative impact of congestion on economic and environmental issues, but also to point out individual benefits (lower costs, time savings, reduction of driving stress, possibility of using driving time for other activities, e.g. reading a book) that encourage people to switch from car to alternative forms of transport.

Congestion in cities is a subject of many scientific dissertations, especially in the era of such intensive development of urban logistics. This is mainly due to the scale of problem, as well as its constant increase (Miklińska, 2017). The explosion of urbanization processes drastically reduces capacity, forcing the company to seek solutions to counteract the phenomenon. A large number of publications are dedicated to this issue. They indicate the possibilities of expansion and modernization of the existing infrastructure (Jamroz, 2016), both linear (construction of new roads, extension of the existing ones) and point infrastructure (parking lots, stations and stops) (Krajewska & Łukasik, 2017), as well as its more effective use through the control of real-time flows. Such methods are favored by the intensive development of intelligent traffic control systems (Kamiński et al., 2016; Selwon & Roman, 2017; Kornaszewski & Gwiazda, 2017), such as lanes with variable traffic direction (Igliński, 2009; Szołtysek, 2011; Kulińska, et al., 2014) or dynamic control of traffic lights (Ruchaj, 2012; Lejda & Siedlecka, 2016). Another measure to counteract urban congestion is to encourage travelers to use collective forms of transport (Nosal & Starowicz, 2010) or more environmentally friendly and congested roads, such as bicycles (Radzimski, 2012; Dębowska-Mróz et al., 2017). An important element in the analysis of congestion is also the costs it incurs, which lead to social and economic losses (Dyr & Kozłowska, 2017, 2018). In addition, mathematical models are presented in the literature for the purpose of mapping transport systems and processes (Jacyna, 2001, 2009; Świderski, 2011; Kamińska & Chalfen, 2017). They not only allow for assessment of transport flows, but also have an important role in the planning processes, better use of the existing infrastructure, as well as shaping social attitudes and even transport policy instruments (Koźlak, 2015).The issues addressed in this article are a part of the above objectives. The presented study describes methods of mathematical description and analysis of transport congestion with multiple regression method and ARIMA model in order to predict the travel time over the studied section.

2. RESEARCH METHODOLOGY

Analysis of above-mentioned phenomena and processes is possible for modelling using the theory of stochastic processes. We are dealing with a family of random variables defined in a certain probabilistic space with values in a certain measurable space (Taylor & Karlin, 1998). If the domain of stochastic process realization is time, as in the case of our analysis, then we address a time series in which individual measurements constitute a set of observations representing the realization of studied phenomenon and are characterized by its changes. The correct identification of series consists in identification of its elements, which may be

systematic components, such as trends, periodical or cyclical fluctuations, and accidental-random components (Fig. 13) (Dittmann et al., 2011).

Figure 1. Components of time series
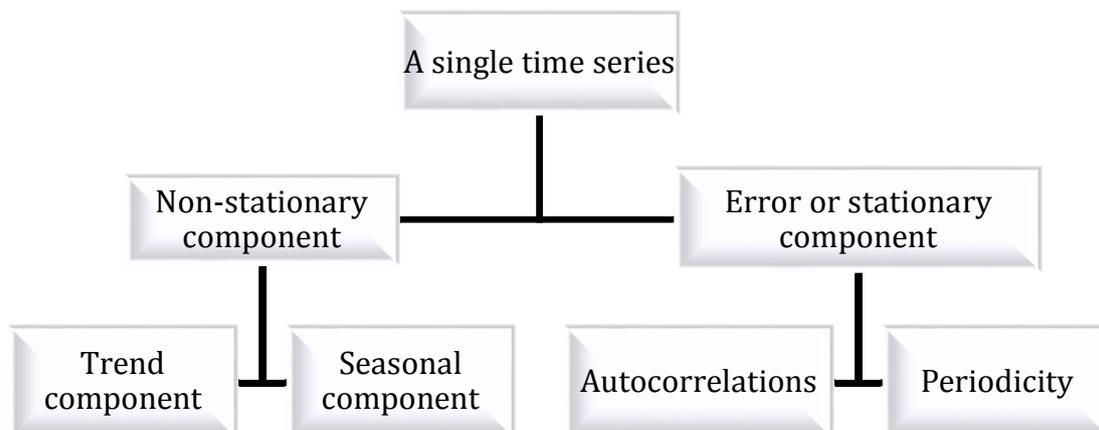
Source: own elaboration

Decomposition of time series enables to define possible to use forecasting models, among the single equation we can distinguish: classic trend models, adaptive trend models, causal-descriptive models and autoregressive models (Cieślak, 2005). In this paper, the multiple regression (MR) method is used as the most common form of linear regression analysis. As a predictive analysis, the multiple linear regression is used to explain the relationship between one continuous dependent variable and two or more independent variables (Dittmann et al., 2011; Rabiej, 2012). Conversely, the ARIMA (Autoregressive Integrated Moving Average) process can be modelled via the group of autoregressive models. Both types of models, MR model and ARIMA model, were developed in this research and compared among each other.

Regression consists in finding a function that reflects implementation of the process and enables to determine how the phenomenon will evolve in the future. It is based on finding correlations between variables, which makes it possible to extend the analysis also to function arguments out-of-sample data. The basis for estimation is to find a proper trend function and then to describe and isolate (if any) seasonal and cyclical fluctuations. For this purpose, the classical method of least squares or method of maximum likelihood are usually used.

The simplest type of regression is a simple linear regression, which describes the relations between variables by means of straight line (1).

$$y = \beta_0 + \beta_1 x + \varepsilon \tag{1}$$

where betas are the parameters, while ε is the noise disturbance assumed to be a white noise (Dittmann et al., 2011; Osińska et al, 2007). If there are more explanatory variables, then we have a multiple regression (MR) and its linear model takes the form (2):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon \tag{2}$$

Regression coefficients describe how much the average value of dependent variable y will change if the value of independent (explanatory) variable x_i to which it relates changes by an entity assuming a fixed level of other independent variables.

Independent variables in the analysis of economic phenomena are sometimes of a qualitative rather than quantitative nature. In this case, it is necessary to translate them into binary variables. Since the number of their values is reduced, they cannot be treated in the same way as continuous variables in regression, which is because of the fact that they make no economic sense. Then such qualitative, or discrete variables have to be re-coded into binary ones and estimation must be carried out with subtraction of one of them. The excluded variable is then the reference level for the others.

Simple autoregressive (AR) models (3) are a group for which it is assumed that there is a relation between the values of time series in the given time and those of the same series in the earlier time, separated by a certain period of time:

$$y_t = f(y_{t-1}, y_{t-2} \dots y_{t-p}, \varepsilon_t) = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + \varepsilon_t \quad (3)$$

where alfas are the parameters, while ε_t is the noise disturbance assumed to be a white noise (Osińska et al, 2007). Therefore, there are no explanatory variables here, and the values of forecast variable are estimated on the basis of their own components, which are distant in time. Their use is limited by the stationary nature of the observed process, which can be ensured by a number of d, differentiating operations: $\Delta y_t^d = (y_t - y_{t-1})^d$. The ARIMA process presented in this article is modelled using the ARIMA model, which is based on the assumption that a value of forecasted variable in time t depends both on its past values and on differences between past values of a true ARIMA process and values obtained from the ARIMA model – i.e. on forecast model errors. The form of classical ARMA model (without integrated I part) is as follows (4):

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + \varepsilon_t \beta_0 + \varepsilon_t - \beta_1 \varepsilon_{t-1} -$$
$$-\beta_2 y_{t-2} - \dots - \beta_q \varepsilon_{t-q} \quad (4)$$

where alfas and betas are the regression parameters, while ε_{t-i} represent the noise disturbances at time t and past time points $t - I$, where the assumption of being the white noise is adopted again. The ARIMA model is obtained from the ARMA model (4) by applying the afore-mentioned differential treatments $\Delta y_t^d = (y_t - y_{t-1})^d$ of the analyzed series to remove time-variant non-stationarity. This is evidenced by the use of the letter I in the name of ARIMA model.
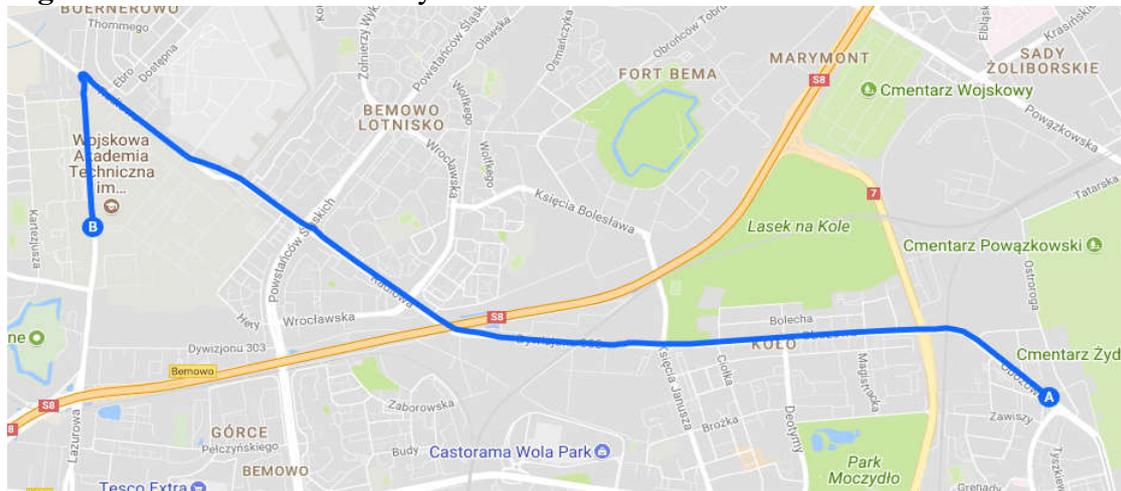
The estimation process is to a large extent formalized (Bielińska, 2007), which results from the procedure created in the 1960s and is proposed by Box and Jenkins. It consists of four main stages: identification, estimation, verification and forecasting. First of all, model structure and parameters should be determined, i.e. such values of *p, d* and *q* model candidates' orders should be found that the best possible model is detected via the estimation procedure in a model selection process. Then it is necessary to analyze the significance of estimated regression parameters of the model (parsimony) and check whether the residual component possesses the properties of the white noise. Moreover, due to stationarity, invertibility and stability reasons, the position of zeros and poles of the ARIMA model's transfer function should be investigated regarding the unit circle (Box et al, 1994). At the end, forecasts are conducted, i.e. future values of the time sequence are determined. Their quality can

be assessed by comparing them with the recorded empirical observations. In addition, the quality of the developed model should be verified via the performance error-based criteria indicators (e.g. Mean Absolute error – MAE, etc.) (Bielińska, 2007).

3. RESEARCH SUBJECT

In order to illustrate how the phenomenon of congestion can be presented experimentally, the journeys over a selected road section were analyzed. The analysis was conducted for the city of Warsaw, which is the capital of Poland. The length of surveyed route was 6 km. Travel time *TT* and fuel consumption (*TT*) were measured for 235 days. They took place on working days, always at the same time. The analyzed route (Obozowa Street 14 – Urbanowicza Street 21) is shown in Figure 2.

Figure 2. Visualization of analyzed section

Source: https://www.google.com/maps/d/edit?mid=15e9NyOU3K0Ec-771KK2T5-GROzj1OlRw&ll=52.25178912954547%2C20.93190000000004&z=15

The surveyed route is an asphalt road of medium quality, on which the local surface defects also occur. The whole length is subject to a speed limit of 50 km/h. There were several reasons for choosing this route. First of all, it should be emphasized that this is a representative section for Warsaw, especially as the alternative route is closed due to ongoing repairs associated with building of a subway system. In addition, data acquisition was related to commuting, which enabled the daily measurement and facilitated to carry out analyses.

4. RESEARCH RESULTS

4.1. Preliminary Analysis of Factors Influencing the Studied Process

Research was carried out in the period from 16 January 2017 to 8 December 2017 every day – on working days – conducting measurements at the same hour. Travel time *TT* and fuel consumption (*TT*) were measured. At first, the relation

between journey time and fuel consumption (*TT*) was studied in order to see if it exists and how strong the relation between these measurements is. The diagram of analyzed values is shown in Figure 3, where the observed variables fuel consumption (*TT*) and a travel time *TT* are intentionally shown to more clearly indicate the relation between them. It is clear that the increase in travel times causes an increase in the amount of consumed fuel. In a diagram, the current consumption measured over analyzed distance was converted into incineration in liters per 100 km, in order to better visualize the phenomenon.

**Figure 3.** Measurements of time dependent travel time *TT*, *t* in days, and fuel consumptions (*TT*) in the analyzed period
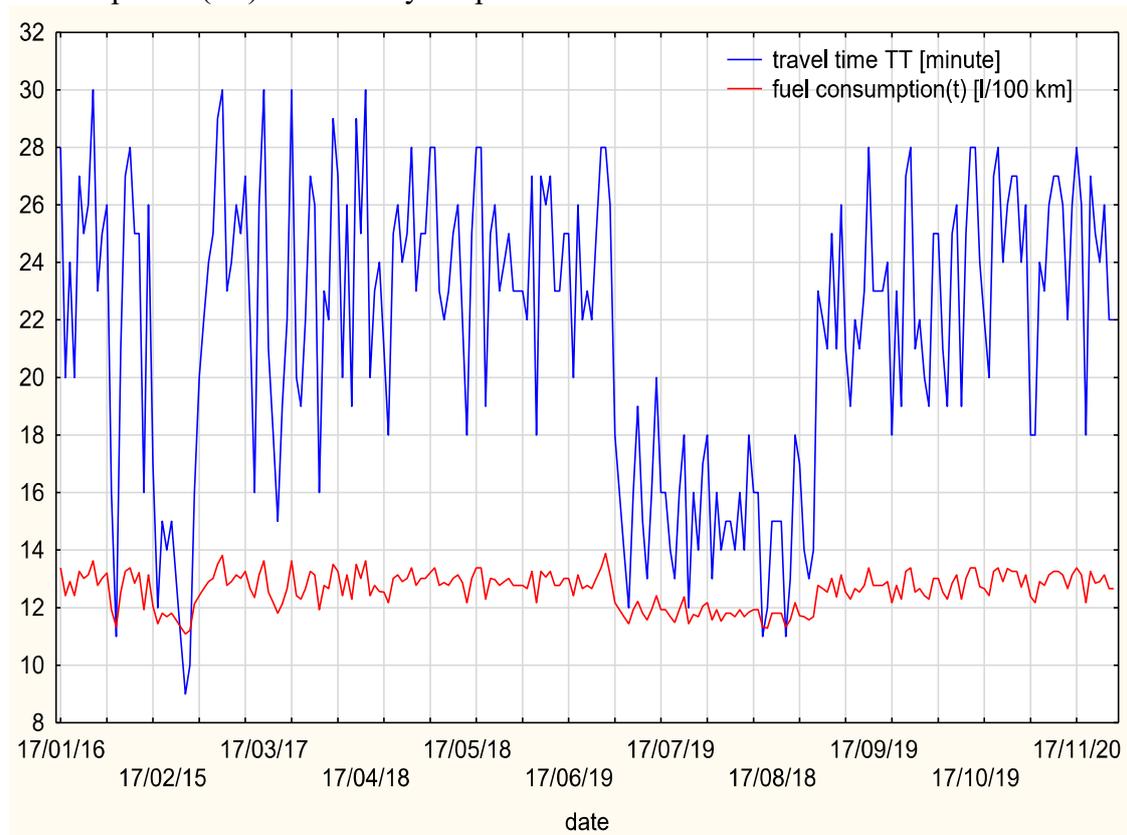
Source: own study

Determined linear regression model, as shown in Figure 4, takes the form of (5):
Fuel (*TT*) = 0.5988 + 0.0073 *TT*   (5)
A strong correlation is confirmed by the scatter diagram on Figure 4 and calculated correlation coefficient of 98%. It is clear that the increase in driving time (in minutes) results in an increase in the amount of consumed fuel. Therefore, every additional minute spent in congestion causes a higher amount of combusted diesel oil (and thus higher costs). The variation of results from uneven traffic intensity on studied section in individual days, as well as from a certain inaccuracy of the measurement device (measurements were conducted on the basis of readings of the on-board computer of vehicle).

Figure 4. Diagram of the linear regression model determined on the basis of actual travel time TT measurements and fuel consumption (TT)
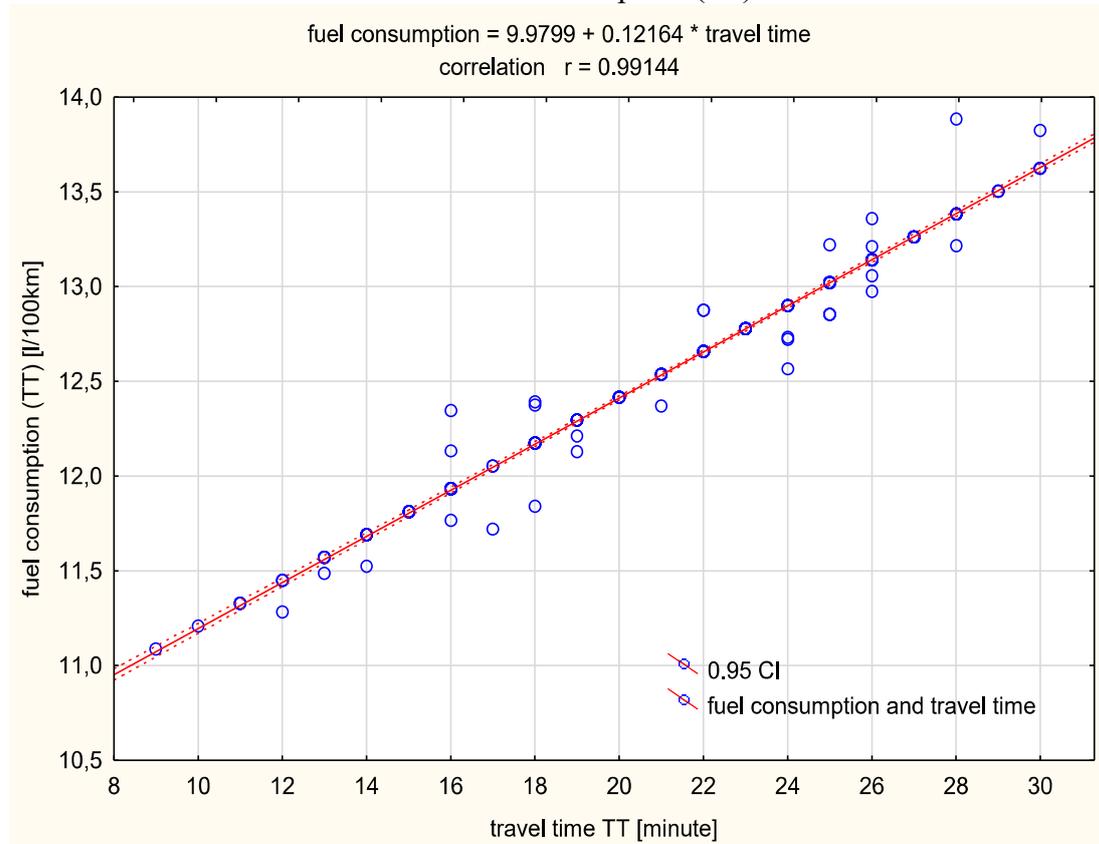
Source: own study

The study showed a strong correlation between the travel time TT and fuel consumption (TT) variable. This is valuable information, which indicates that every additional minute spent in congestion causes an increase in fuel consumption, and as a result, undesirable consequences, such as an increase in travel costs or an increase in emissions to the environment. Since the main objective of this study was to determine the impact of selected factors on the phenomenon of congestion in cities, as a further step only the travel time TT variable was analyzed.

Diagram 3 clearly shows two periods in which the surveyed variables take significantly different values. The analysis of basic descriptive statistics of the travel time TT indicates a high fluctuation of this variable – from 9 to 30 minutes. The coefficient of variation is about 22% (Table 2), so it is quite high. Therefore, the cause of this phenomenon was studied.

Table 2. Descriptive statistics for variable journey time

Variable: travel time TT	Descriptive statistics					
	N	Mean	Min.	Max.	Standard deviation	Coefficient of variation
time	229	21.61	9	30	4.99	23.10

Source: own study

Detailed study of the travel times showed that during the school year, the average travel time *TT* is much longer than during both winter and summer holidays. This is due to the fact that during this period a significant part of the population takes holiday leaves and stays outside the city, which means that there are fewer traffic participants. This relation is shown by the following box plot (Fig. 5).

Figure 5. Box plot of variable travel time *TT* for school year, as well as summer and winter holiday
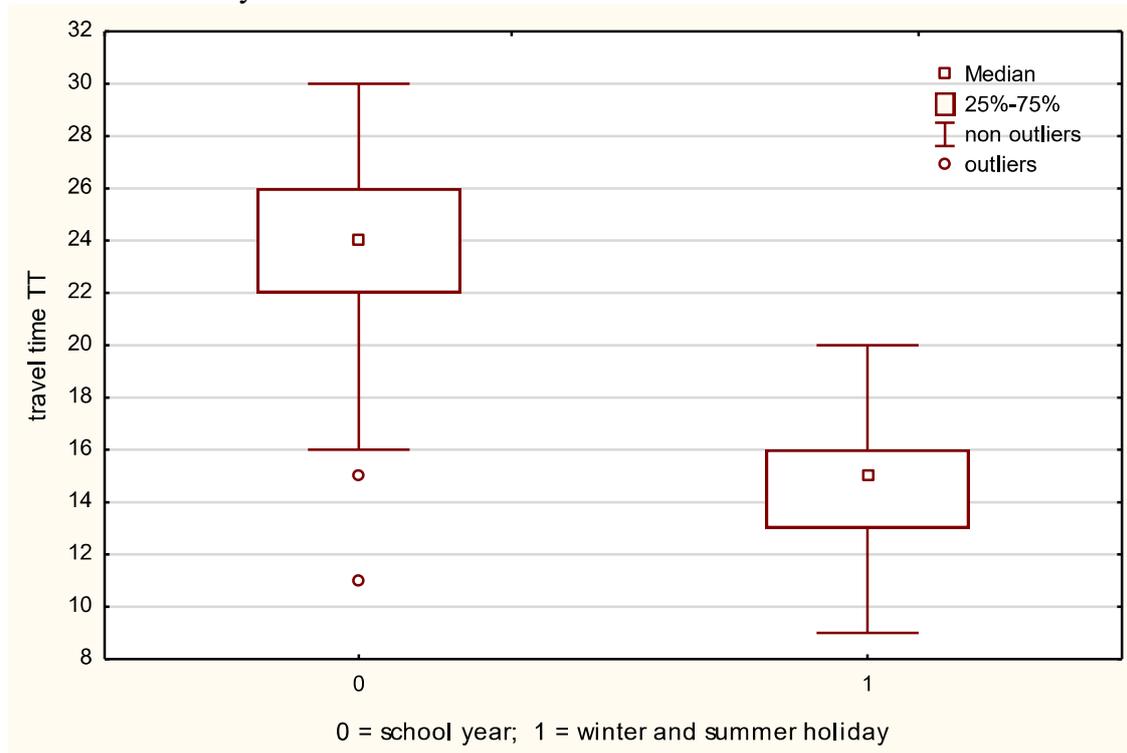
Source: own study

The chart (in figure 5) shows atypical observations. These are not measurement errors, but observations, which indeed took place. Their further analysis leads to conclusion that shorter journey times are achievable not only in school-free period, but also on Tuesdays, Wednesdays and Thursdays when – as the study shows – they are lower than those recorded on Mondays and Fridays. The reason for such phenomenon is a significant migration of people from outside Warsaw to the city, who arrive to work or school from nearby towns. Quite often, the route between home and capital city is covered by car, hence the increased traffic on Mondays and Fridays, while on other working days it is covered by public transport. In relation to overlapping seasonality resulting from holidays and from weekdays, both these factors had to be taken into account. This was shown in Figure 6, where the average journey times recorded on individual weekdays for two periods were compared: school year period and for both: winter and summer holidays.

Figure 6. Box plot considering the seasonality related to holidays and weeks
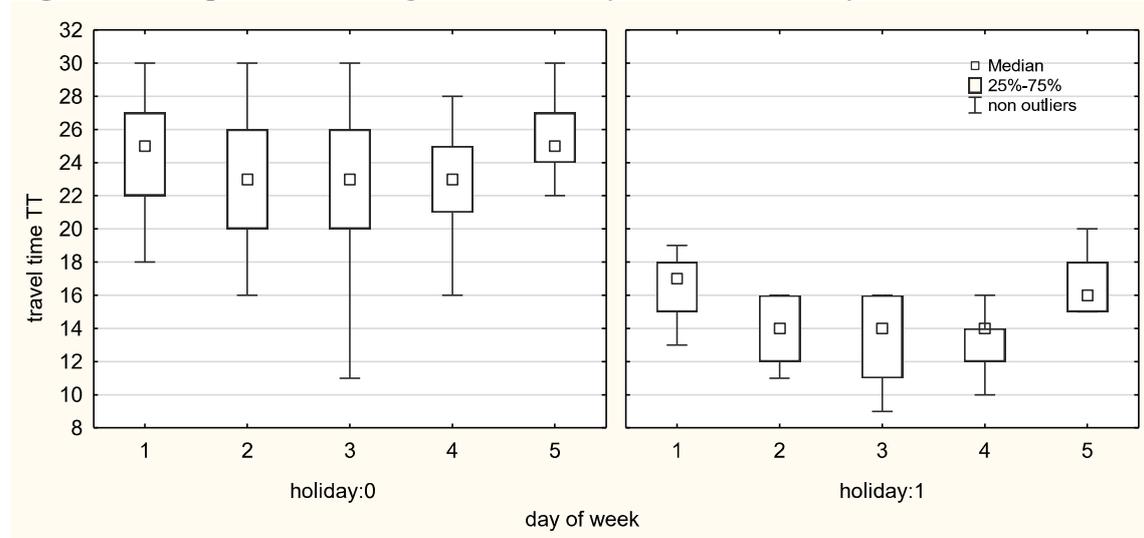
Source: own study

Taking into account the seasonality on the box causes plot that there are no atypical observations. The descriptive statistics presented in Table 3 and 4 are also different.

Table 3. Descriptive statistics of variable travel time *TT* for the school year

Variable: travel time *TT* for the school year	Descriptive statistics					
	N	Mean	Minimum	Maximum	Standard deviation	Coefficient of variation
time	175	23.73	11	30	3.44	14.51

Source: own study

**Table 4.** Descriptive statistics of variable travel time *TT* for holidays

Variable: travel time *TT* for holidays	Descriptive statistics					
	N	Mean	Minimum	Maximum	Standard deviation	Coefficient of variation
time	54	14.74	9	20	2.32	15.71

Source: own study

As can be seen, the average travel time *TT* during the school year is almost 23.73 minutes, which is much longer than the average travel time *TT* on holidays, which is almost 9 minutes less. The maximum and minimum journey times also vary, i.e. appropriately 9 and 20 minutes for holidays and 11 and 30 minutes for the rest of period. The coefficient of variation for both processes is similar at around 15%, which allows for continuing the study.

Before the estimation, a preliminary trend analysis was carried out, calculating the correlation coefficient between the variable travel time *TT* and time *t* (observation day). A negligible result was obtained, r=-0.006471 that informs about the lack of dependence. The scatter diagram was also drawn (Fig. 7), which clearly shows that most observations are beyond the confidence limits (indicated by red lines on diagram). Therefore, there is no travel time dependence from the next day of observation (*t*), and there is no basis to estimate the trend in the model, which describes the travel time.

Figure 7. Scatter diagram of variable journey time in relation to variable *t* (observation day)
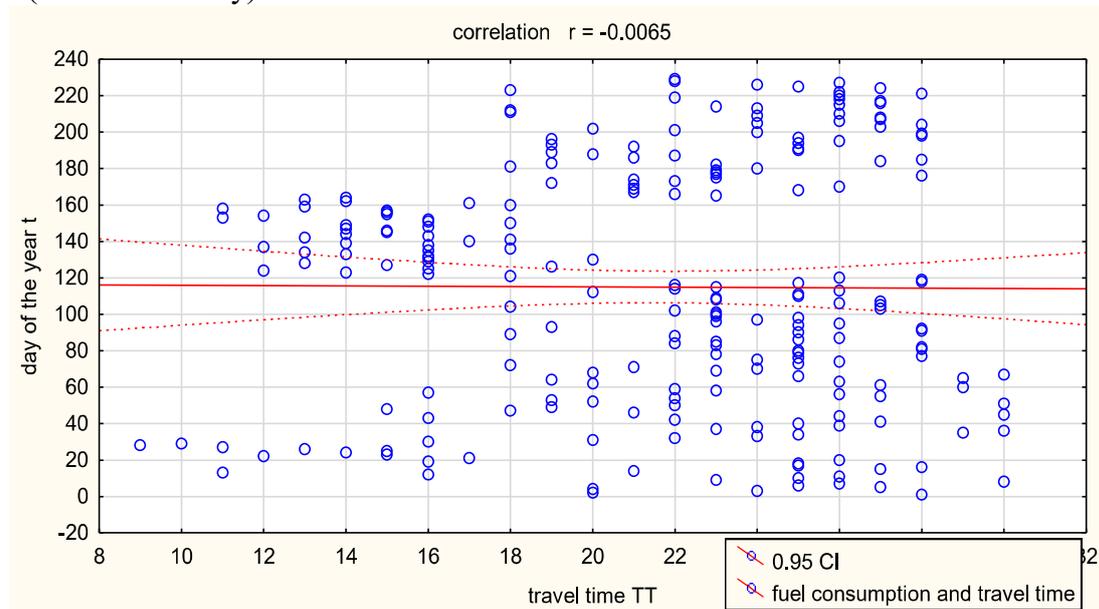
Source: own study

4.2. Multiple Regression (MR) Model With Binary Variables

4.2.1 Model 1 – Only Three Daily-Based Dummy Variables

According to quoted analyses, the stochastic model of daily trips should not take into account the long-term trend, but should describe short-term seasonality for particular days of the week, or sets of groups: working days (Monday and Friday, Tuesday, Wednesday and Thursday) and weekend days. As mentioned earlier, the qualitative nature of explanatory variables causes the need to estimate a model in which the number of variables is reduced by one.

First of all, model 1 was proposed, which takes into account weekdays, i.e. Monday, Friday and a group of days from Tuesday to Thursday, due to the fact that travel times for these days were similar. Multiple regressions (MR) were used for this purpose, estimating the model with binary variables *D1* (Monday) and *D3* (Friday). The estimated regression parameters were obtained using the Statistica software. They are shown in Table 5, while the model has the following structure (6):

$$\widehat{TT} = \hat{a}_0 + \hat{b}_1 \cdot D_1 + \hat{b}_{2-4} \cdot D_2 + \hat{b}_5 \cdot D_3 \qquad (6)$$

where the $\widehat{TT}$ is the estimate of the real $TT(t)$, while t represents the days of the year. During the estimation in Statistica program, variables corresponding to individual days or groups of days were coded with 0 and 1, where 0 means that variable is excluded from the model, and vice-versa.

Table 5. Estimation results – model 1

	Constant	Monday	Tuesday - Thursday	Friday
Parameter	$\hat{a}_0$	$\hat{b}_1$	$\hat{b}_{2-4}$	$\hat{b}_5$
Value	20.65	1.96	-4.84	-2.89
Standard deviation SD	0.41	0.83	1.18	0.83
t - values	49.89	2.36	-3.48	3.45
Relative forecast error δ [%]	2	42.31	-24.30	28.97

Source: own study

All estimated parameters turned out to be statistically significant. However, relative forecast error δ [%] while doing the estimation of parameters are high, and the corrected determination coefficient is only 5%, which means that random component has a definite advantage over the deterministic. In addition, the residue distribution deviates significantly from normal distribution shown in the diagram by red line (Fig. 8).

Figure 8. Histogram of residuals of the model 1


SW-W = 0.9577; p value = 0.00000
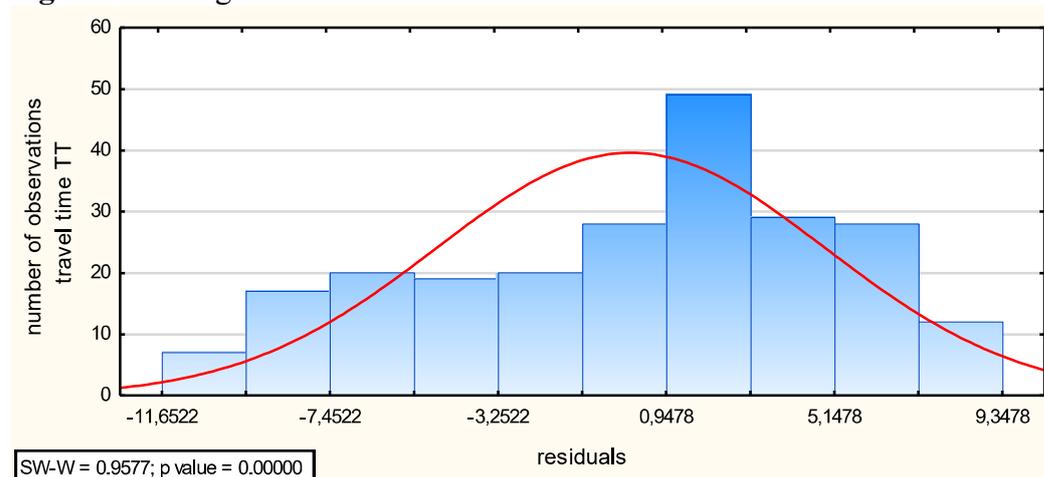
Source: own study

Forecasting quality of the model is negligible ($R^2 = 5$ %). Its causes are shown in the diagram of time series and forecast function (Fig. 9), where it can be seen that the model, generally speaking, does not take into account the seasonality resulting from winter holidays and summer holidays, which significantly increases

the variance, deteriorates determination coefficient and desymmetricizes the distribution of residuals.

Figure 9. Diagram of empirical variable TT (t) and forecasts $\hat{TT}$ for the model 1
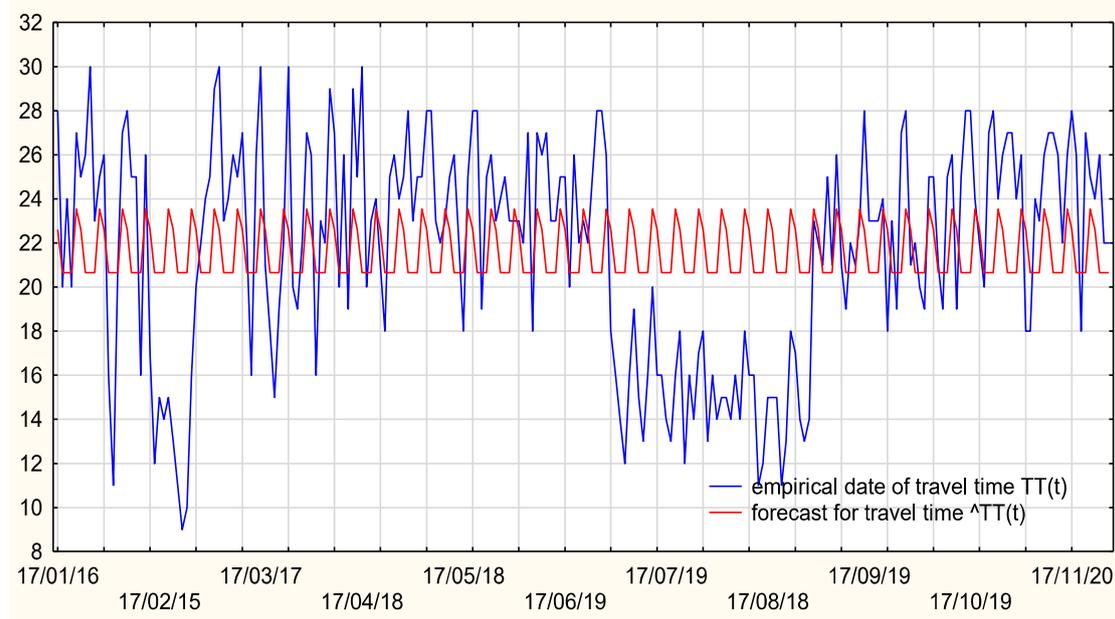
Source: own study

Due to the insignificant forecasting quality of the model, it was not subjected to further statistical tests to verify its properties.

4.2.2 Model 2 – Applying an Additional Holidays' Dummy Variable

Therefore, analyzed model 1 should be adjusted to take into account not only weekly, but also seasonality due to winter holidays. A new binary variable *Dh* has been created for the school holidays and the parameter estimates have been calculated again. The obtained results are presented in Table 6, while the model has now the following structure (7):

$$\hat{TT} = \hat{a}_0 + \hat{b}_1 \cdot D_1 + \hat{b}_{2-4} \cdot D_2 + \hat{b}_5 \cdot D_3 + \hat{b}_h \cdot D_h \qquad (7)$$

Table 6. Estimation results – model 2

	Constant	Monday	Tuesday -Thursday	Friday	Holidays
Parameter	$\hat{a}_0$	$\hat{b}_1$	$\hat{b}_{2-4}$	$\hat{b}_5$	$\hat{b}_h$
Value	22,79	1,96	-4,27	2,73	-8,96
Standard deviation SD	0,28	0,51	13,28	0,52	0,47
t - values	81.61	3.82	-0.32	5.29	-19.13
Relative forecast error δ [%]	1,23	26,16	20,24	18,91	-5,22

Source: own study

Errors in estimation of parameters are smaller, but still quite high, especially for Mondays. However, the determination ratio has increased from 5% to 64%, indicating a significant improvement in fitting model 2.

Figure 10. Histogram of residual model 2

SW-W = 0.9825; p = 0.0064
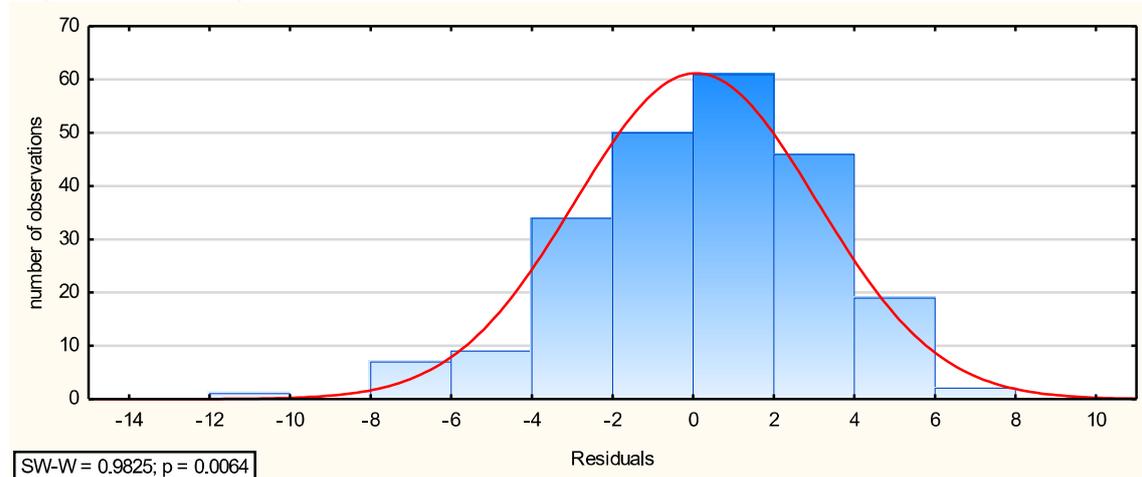
Source: own study

Residuals distribution is more similar to normal (Fig. 10) and the model significantly better reflects variability resulting from school holidays (Fig. 11). Nevertheless, it was decided to make one more attempt to estimate the regression model.

Figure 11. Diagram of empirical variable TT(t) and forecasts $\hat{TT}$ for the model 2

Source: own study

514
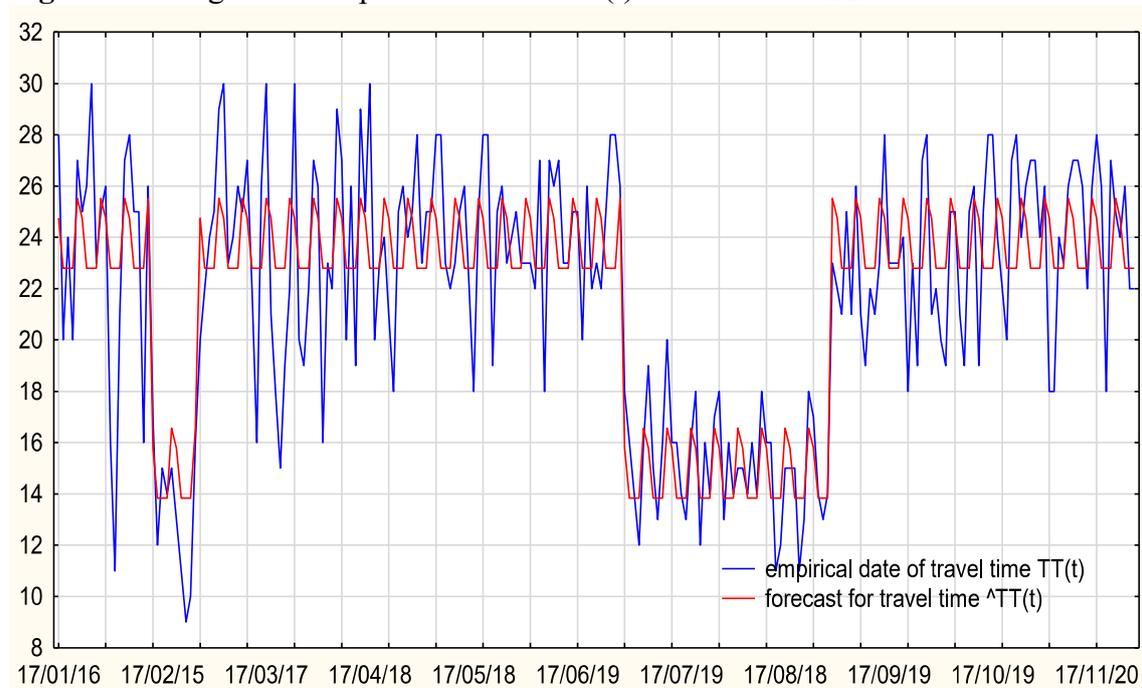
4.2.3 Model 3 – Applying All Daily Dummy Variables And Additional Holidays

Finally, a third model of multiple regression (MR) was proposed, in which the explanatory variables concern all days of the week and days related to winter holidays, and the model 3 was estimated for Mondays with the following results (Table 7), while the model has now the following structure (8):

$$\widehat{TT} = \hat{a}_0 + \hat{b}_1 \cdot D_1 + \hat{b}_2 \cdot D_2 + \hat{b}_3 \cdot D_3 + \hat{b}_4 \cdot D_4 + \hat{b}_5 \cdot D_5 + \hat{b}_h \cdot D_h \qquad (8)$$

Table 7. Estimation results – model 3

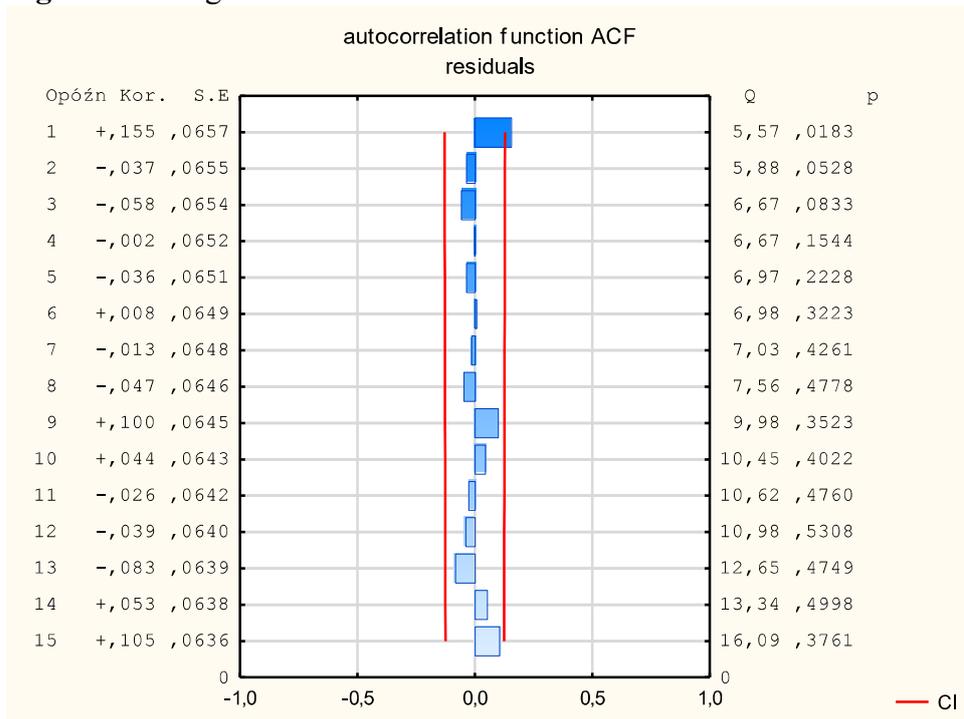|	Constant	Monday	Tuesday	Wednesday	Thursday	Friday	Holiday
Parameter	$\hat{a}_0$	$\hat{b}_1$	$\hat{b}_2$	$\hat{b}_3$	$\hat{b}_4$	$\hat{b}_5$	$\hat{b}_h$
Value	24.75	14.05	-1.83	-2.04	-2.00	0.77	-8.96
Standard deviation SD	0.46	1.81	0.63	0.63	0.63	0.63	0.47
t - values	53.91	7.76	-2.90	-3.25	-3.178	1.22	-19.06
Relative forecast error δ [%]	1.85	9.57	-34.47	-30.80	-31.47	81.86	-5.25

Source: own study

The corrected determination factor R^2 is 63%, but variable for Friday is insignificant, large errors in estimation of parameters do not allow for considering model 3 as better than the previous one. This way constructed model is indeed parsimonious, but in regression analysis with a use of dummy variable, it is also practiced to leave insignificant variables in the model, because thanks to that simple and clear interpretation of seasonal parameters is preserved (Sokołowski, 2016).
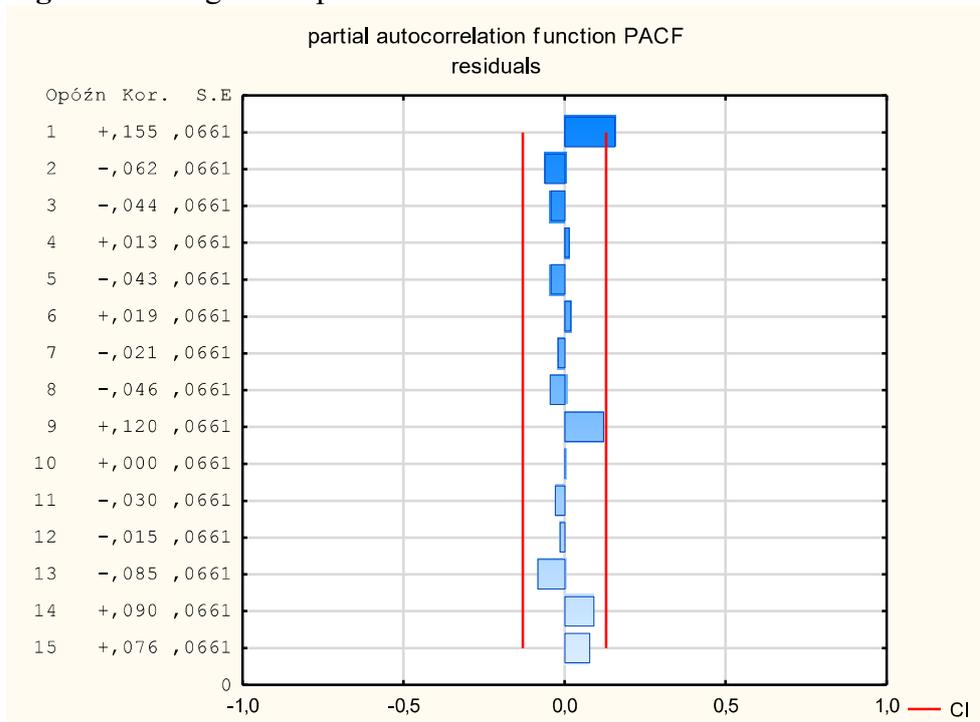
4.2.4 Validation Of The Best Model 2

Therefore, it was finally decided that model 2 is the most reliable for analyzed empirical data and it was subjected to further research. In order to check whether all relations existing in time series have been explained, a diagram of autocorrelation and partial autocorrelation functions of the residual model has been prepared (Fig. 12, Fig. 13).

Figure 12. Diagram of autocorrelation function of the residual model

Opóźn	Kor.	S.E		Q	p
1	+,155	,0657		5,57	,0183
2	-,037	,0655		5,88	,0528
3	-,058	,0654		6,67	,0833
4	-,002	,0652		6,67	,1544
5	-,036	,0651		6,97	,2228
6	+,008	,0649		6,98	,3223
7	-,013	,0648		7,03	,4261
8	-,047	,0646		7,56	,4778
9	+,100	,0645		9,98	,3523
10	+,044	,0643		10,45	,4022
11	-,026	,0642		10,62	,4760
12	-,039	,0640		10,98	,5308
13	-,083	,0639		12,65	,4749
14	+,053	,0638		13,34	,4998
15	+,105	,0636		16,09	,3761

Source: own study

Figure 13. Diagram of partial autocorrelation function of the residual model

Opóźn	Kor.	S.E
1	+,155	,0661
2	-,062	,0661
3	-,044	,0661
4	+,013	,0661
5	-,043	,0661
6	+,019	,0661
7	-,021	,0661
8	-,046	,0661
9	+,120	,0661
10	+,000	,0661
11	-,030	,0661
12	-,015	,0661
13	-,085	,0661
14	+,090	,0661
15	+,076	,0661

Source: own study

The above diagrams in figures 12 and 13 show that apart from the first delay, there are no significant dependencies, which may suggest that most of dependencies

that occur in the analyzed phenomenon have been explained by the model. This is confirmed by the Ljung-Box test, which assumes the following in the hypothesis H0: The data are independently distributed. The calculated value of test statistics up to the order of 10 was Q = 10.447, with a p-value = P (Chi-square(10) > 10.447) = 0.402, which means that the investigated series is a process of white noise.

However, White's test for heteroscedicity of residuals showed that the variance of random element is not constant. For the zero hypothesis, which assumes that the residual heteroscedicity is not present, the value of test statistics LM=22.0237 with the p-value = P(Chi-square(5) > 22.0237) = 0.000518201 was obtained. In addition, the study of normality of the residues distribution at a significance level of α=0.05 indicated the need to reject H0 hypothesis, which assumes that random element has a normal distribution (obtained Chi-square test statistic=6.49899 with p-value=0.0387938). All this indicates an imperfection of the proposed model. Nevertheless, it was decided to verify it on the basis of preserved test observations. Estimated forecasts along with their relative forecast errors are presented in Table 8.

Table 8. Comparison of empirical and forecast data for the test interval

| Date	Regression model (estimated values $\widehat{TT}$)	Empirical data (observations TT)	Forecasting relative error Ψ [%]
17-12-01	25,52	26	1,8%
17-12-04	24,75	25	1,0%
17-12-05	18,52	22	15,8%
17-12-06	18,52	24	22,8%
17-12-07	18,52	22	15,8%
17-12-08	25,52	26	1,8%

Source: own study

In accordance with the above table, the forecasts obtained for days when the traffic intensity is high are satisfactory and amount to less than 2%. On the other hand, on the remaining days of the week, the forecast was much more optimistic than empirical values, and therefore the underestimation of errors is between 15 and 20%.

4.3. ARIMA Model

In the next part of study, the ARIMA model was proposed. As mentioned above, it can only be used for stationary series, and therefore the first step was to analyse series under stationary performance. For this purpose, the diagram of autocorrelation function, shown in Figure 14, was calculated at first.

Figure 14. Diagram of autocorrelation function (ACF) of the variable travel time *TT*.
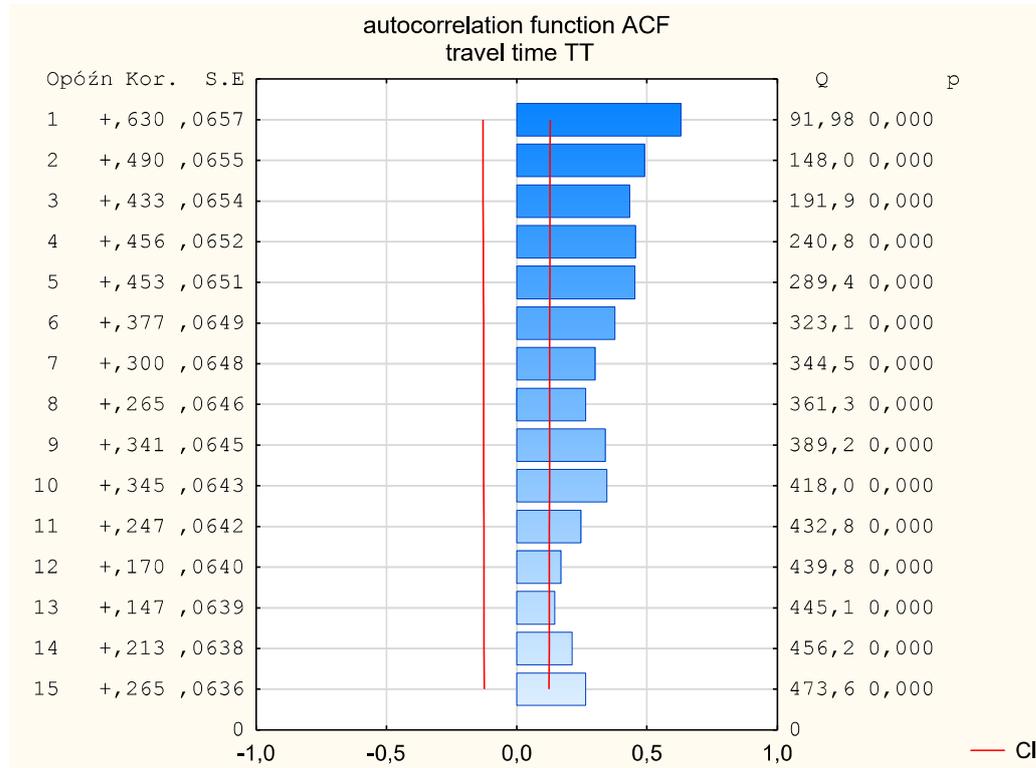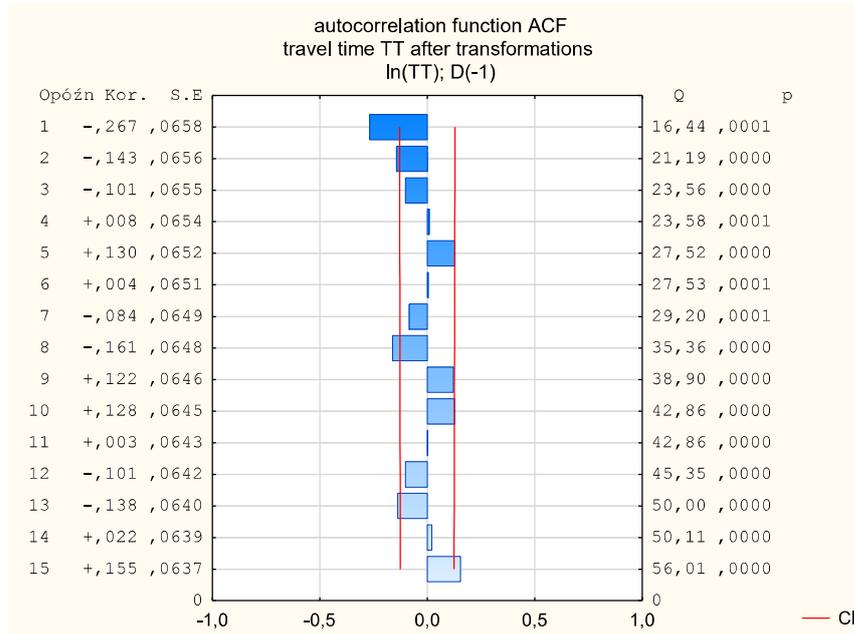

Source: own study

The ACF function does not expire, its values are statistically significantly different from zero. Presence of a unit root was confirmed by the Dickey-Fuller test, for which the H0 zero hypotheses assumes that a unit root exists. The calculated empirical significance level is high and amounts to approximately 50% (p-value= 0.4996), which does not allow to reject H0 and indicates that the tested series is stationary due to the presence of a unit root. Therefore, it must be reduced to a stationary form. For this purpose, two transformations were conducted. Firstly, the logarithmic transformation has been carried out (9):

$$\tilde{y}(t) = \log(TT(t)) \tag{9}$$
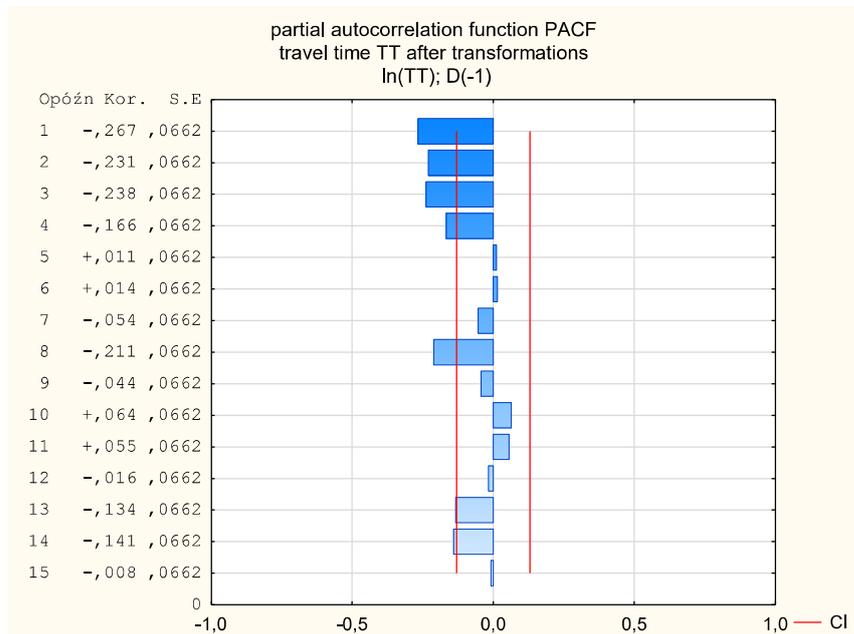
Then first order differentiation (10):

$$\Delta\tilde{y}_t^1 = (\tilde{y}_t - \tilde{y}_{t-1})^1 \tag{10}$$

Figure 15. Diagram of autocorrelation function (ACF) of the of the modified variable travel time TT after transformations

Source: own study

Figure 16. Diagram of partial autocorrelation function (PACF) of the modified variable travel time *TT* after transformations

Source: own study

The ACF function decreases rapidly (Fig. 15), while the PACF function (Fig. 16) breaks off for a delay value k>4, which may initially suggest that this is the AR(4) autoregressive process. Since the course of correlograms does not always provide clear results, several models were estimated, including parameters *p* and *q* in the range from 0 to 4, and also those taking into account the parameter of moving average

process MA(q). In the first place, they were selected due to statistical significance of estimated parameters, and in this way the models for which all parameters are statistically significant were selected, as presented in Table 9.

The best results were obtained for models (formula11 and 12):

ARIMA (2,1,0):

$$\tilde{y}(t) = \hat{a}_0 + \hat{a}_1 \cdot \tilde{y}(t-1) + \hat{a}_2 \cdot \tilde{y}(t-2) = 0 - 0.33 \cdot \tilde{y}(t-1) -$$
$$-0.23 \cdot \tilde{y}(t-2) \tag{11}$$
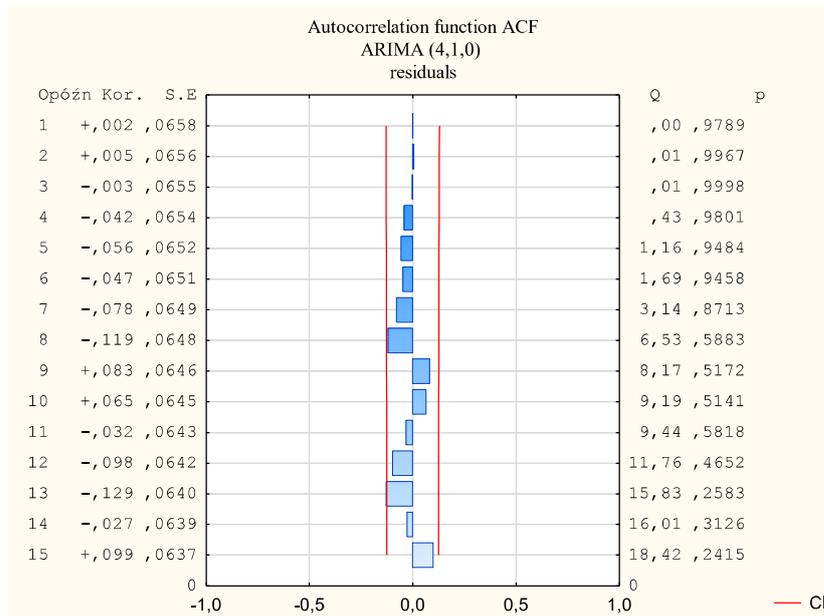
and ARIMA (4,1,0):

$$\tilde{y}(t) = \hat{a}_0 + \hat{a}_1 \cdot \tilde{y}(t-1) + \hat{a}_2 \cdot \tilde{y}(t-2) + \hat{a}_3 \cdot \tilde{y}(t-3)$$
$$+\hat{a}_4 \cdot \tilde{y}(t-4) = 0 - 0,42 \cdot \tilde{y}(t-1) - 0.36 \cdot \tilde{y}(t-2) -$$
$$-0.3 \cdot \tilde{y}(t-3) - 0.17 \cdot \tilde{y}(t-4) \tag{12}$$

**Table 9.** Summary of estimation results

model	Model ARIMA (4,1,0)	Model ARIMA (3,1,0)	Model ARIMA (2,1,0)	Model ARIMA (1,1,1)
Transformations:	ln(TT) D(1)	ln(TT) D(1)	ln(TT) D(1)	ln(TT) D(1)
a_1 p(1)	-0.42	-0.38	-0.33	0.38376
a_2 p(2)	-0.36	-0.31	-0.23	0.82903
a_3 p(3)	-0.30	-0.24		
a_4 p(4)	-0.17			
MSE Mean Squared Error	0.035	0.036	0.038	0.035
AIC Akaike information criterion	-110.12	-105.77	-94.21	-110.68
BIC Bayesian information criterion	-92.98	-92.05	-83.93	-96.97
HQC Hannan–Quinn information criterion	-103.21	-100.23	-90.06	-105.15
ACF and PACF functions of the residuals	not significantly different from zero	significantly different from zero	significantly different from zero	significantly different from zero

Source: own study

Other factors determining the quality of model were then analyzed. Due to the lowest value of all calculated information criteria, the best turned out to be model (2,1,0). However, the analysis of diagram of the autocorrelation function and partial autocorrelation of its residuals showed that there exist for the model values of these functions significantly different from zero, which indicates that in the model remained relations, which were not fully explained by it. Only for the ARIMA (4,1,0) model, all values of ACF and PACF functions of the residuals distribution (Fig. 17 and Fig. 18) were statistically significantly different from zero, which suggests that this model will better explain the relations existing in series. Consequently, both ARIMA (2,1,0) and ARIMA (4,1,0) were further verified.

Figure 17. Autocorrelation function (ACF) of the ARIMA (4,1,0) residual model distribution

Source: own study

Figure 18. Partial autocorrelation function of the ARIMA (4,1,0) residual model distribution
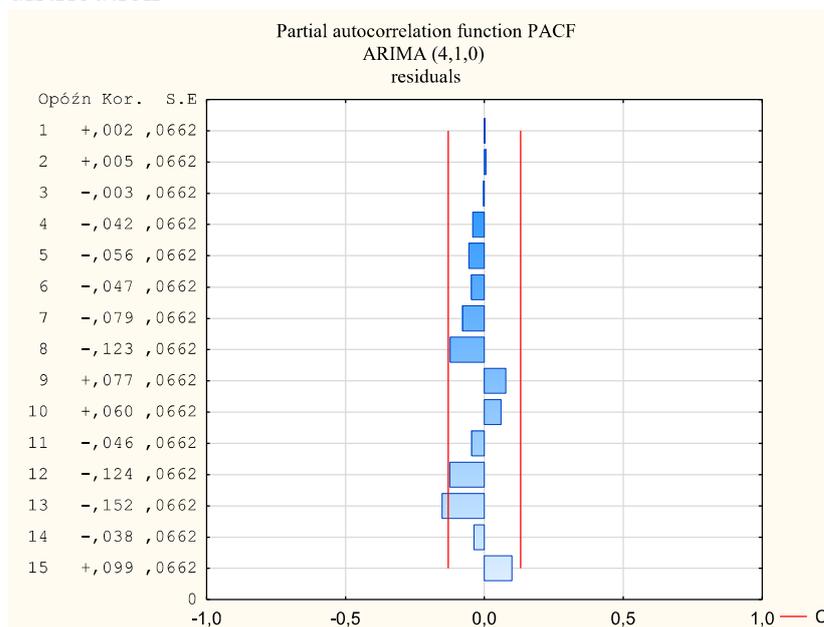
Source: own study

The last stage of model verification is to check its reliability in forecasting. Preserved test observations were reused and three proposed models were compared: the best multiple regression (MR) model from section 5.2, ARIMA (4,1,0) and ARIMA (2,1,0) model. Relative forecast errors were calculated for each observation. The results are presented in Table 10.

Table 10. Forecasting errors for regression model, ARIMA (4,1,0) and ARIMA (2,1,0) models

Date	Day of week	Empirical data	Forecasts			Relative forecast error Ψ [%]		
			Regression model	ARIMA (4,1,0)	ARIMA (2,1,0)	Regression model	ARIMA (4,1,0)	ARIMA (2,1,0)
17.12.01	5	26	25,52	22,97	22,87	1,83	11,67	12,05
17.12.04	1	25	24,75	23,41	22,58	1,00	6,37	9,69
17.12.05	2	22	18,52	23,51	22,47	15,80	-6,86	-2,14
17.12.06	3	24	18,52	23,00	22,57	22,82	4,15	5,95
17.12.07	4	22	18,52	22,88	22,56	15,80	-4,01	-2,56
17.12.08	5	26	25,52	23,01	22,54	1,83	11,50	13,30

Source: own study

Conducted research shows that none of models managed perfectly with empirical data, but the average forecast errors for all the models are satisfactory and they do not exceed 10%. The best result was obtained for the ARIMA (4,1,0) model, which diagram is shown in Figure 19.

Figure 19. Diagram of empirical variable TT(t) and forecasts $\hat{TT}$ for the ARIMA (4,1,0) model
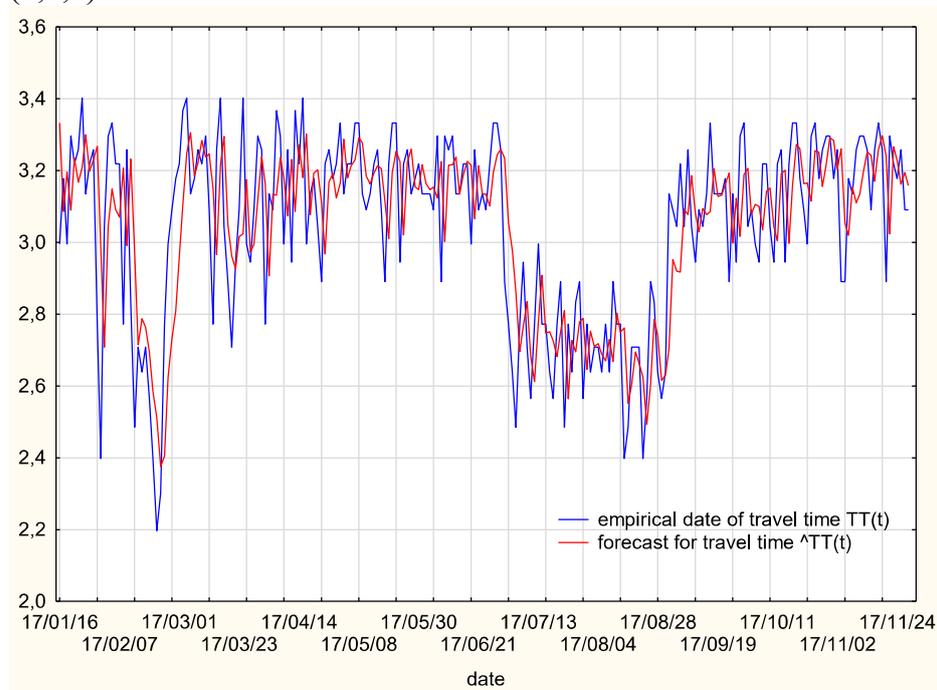
Source: own study

The regression model managed best with estimation on the days when traffic intensity was the highest, i.e. on Mondays and Fridays. On the remaining days of week, the ARIMA models proved better. If the capabilities of regression model and ARIMA (4,1,0) were combined, the average error of such a forecast would be just over 3% (Table 11).

Table 11. Average forecast errors for estimated models

	Regression model	ARIMA (4,1,0)	ARIMA (2,1,0)	Regression model / ARIMA (4,1,0)
Relative forecast mean error Ψ [%]	9,85	7,43	7,61	3,28

Source: own study

5. CONCLUSIONS

The problem of congestion regards many cities around the world. It is difficult to solve as it requires the involvement of many entities of the transport system, in particular road users. The best and latest developments in transport infrastructure and superstructure will not be effective without changing citizen's habits and mentality. Therefore, all available methods must be used to shape the right attitudes, to which this article also contributes. The mathematical analysis of short distance showed drastic differences between the driving times depending on the factor, which is a day of the week. Difference between the longest and the shortest time was as much as 21 minutes and it is a lost time, which could be used much more pleasantly and efficiently. The economic issue is also important. Getting stuck on crowded road is not only an increase in fuel costs, but also the cost of missed opportunities.

The article shows possibility of mathematical modelling of the congestion phenomenon. The versatility of chosen route allows the model to be applied also to other parts of the city, as well as to calculate the potential journey times. It is a good practice to determine the same phenomenon using many methods. In this way, the obtained results can be compared, first of all, in terms of their accuracy. For example, the simple regression model was found to better reflect the time travel of days with the highest traffic, while the more complex ARIMA model was proved for the remaining days of the week.

Therefore, the analyses carried out can be improved and developed using more advanced econometric models. (e.g. ARIMAX models). Moreover, it is worth to consider comparison of obtained results for passenger car journeys with alternative transport means, such as public transport (buses and trams) or the idea of city bikes, which is particularly preferred in times of sustainable transport development. There is no doubt that the congestion issue will be the subject of many scientific dissertations.

6. REFERENCES

Altshuler, A., Womack, J.P & Pucher, J.R (1979). *The Urban Transportation System Politics and Policy Innovation*. Cambridge Massachusetts: MIT Press.

Bielińska, E. (2007). *Forecasting temporary sequences*, Gliwice: Publishing Company of the Silesian University of Technology.

Box, G. E. P. & Jenkins, G.M. (1983). *Time-series analysis. Forecasting and control*, Warsaw: Scientific Publishing House PWN.

Ciesielski, M. (1986). *Costs of transport congestion in cities*. Poznań: Publishing Company of the University of Economics.

Cieślak M. (2005). *Economic forecasting*. Warsaw: Scientific Publishing House PWN.

Dębowska-Mróz, M.; Lis, P.; Szymanek, A. & Zawisza, T. (2017). The municipal bicycle as system components transport in cities, *Buses: technology, operation, transport systems*, (18)6, p. 1173-1182.

Dittmann, P., Szabela-Pasierbińska, E., Dittmann, I. Szpulak, A. (2011). *Forecasting in sales and financial management of a business*, Warsaw: Wolters Kluwer Polska Sp. Z o.o.

Dyr, T. & Kozłowska, M. (2018). Congestion costs in the European Union 2018, *Buses: technology, operation, transport systems*, 19 (1-2), p. 26-31.

Dyr, T.& Kozłowska, (2017). M. Congestion costs in the European Union, *Rail Transport Technique*, 24(7-8), p. 18-23.

Hugos, M. (2006). *Essentials of Supply Chain Management*, 2nd Edition, Hoboken, New Jersey: John Wiley & Sons, Inc.

Igliński, H. (2009). *Reducing transport congestion and sustainable urban development*, (Unpublished doctoral dissertation), Economic University In Poznań.

INRIX (2018). Global Traffic Scorecard [available at: http://inrix.com/scorecard/, access March 16, 2018]

Jacyna, M. (2001). *Multicriterial modelling in application for the assessment of transport systems*. Warsaw: Publishing Office of the Warsaw University of Technology.

Jacyna, M. (2009). *Modelling and evaluation of transport systems*. Warsaw: Publishing Office of the Warsaw University of Technology.

Jamroz, K. (2016). *Guidelines for the road extension to additional lanes depending on the expected road traffic volume*, *Part I – Analysis report*, Technical University of Gdańsk.

Kamińska, J. & Chalfen, M. (2017). Congestion versus travel time: modelling approach. *Buses: technology, operation, transport systems*, (18)6, p. 1212-1214.

Kamiński, T.; Niezgoda, M.; Siergiejczyk, M.; Oskarbski, J.; Świderski, A.& Filipek, P. (2016). Impact of the use of Intelligent Transport Systems services on the road safety level, *Scientific work - transport*, 113, p. 201-208.

Kornaszewski, M. & Gwiazda, A. (2017). Intelligent systems of transport as means of solving problems in transport congestion, *Buses: technology, operation, transport systems*, 18(6), p. 1429-1434.

Koźlak, A. (2015). Limitation of transport congestion in the cities as a major challenge for transport policy, *Logistics*, 3, p. 5386-5394.

Krajewska, R. & Łukasik, Z. (2017). Effective utilization of transport infrastructure in cities: examples of good practice, *Buses: technology, operation, transport systems*, (18)9, p. 203-211.

Kulińska, E.; Rut, J.& Partyka, P. (2014). Reduction of congestion with the use of flexible lanes, *Logistics*, 2, p. 26-34.

Lejda, K &, Siedlecka, S., (2016). Intelligent traffic systems in cities, *Buses: technology, operation, transport systems*, 17(12), p. 680-683.

Miklińska, J. (2017). Development Of The Intermodal Market In Poland From The Perspective Of Cargo Handling Operations In International Supply Chains; Selected Issues, *Business Logistics in Modern Management*, Osijek, p. 233-249.

Miklińska, J. (2017). Development Of The Market For Modern, Commercial Warehouse Spaces As The Logistics Infrastructure Facilities Of The Region; Experiences From Poland, *6th International Scientific Symposium Economy of Eastern Croatia; Vision and Growth*, Osijek, p. 880-890.

Nosal, K. & Starowicz, W. (2010). Some problems of mobility management, *Urban and Regional Transport*, 3, p. 26-31.

Osińska, M., Kośko, M. & Stempińska, J. (2007). *Modern econometrics*. Toruń: Publishing House - Scientific Society of the Organization and Management "Dom Organizatora".

Rabiej, D. (2012). *Statistics with the Statstica program*, Gliwice: Helion Publishing House.

Radzimski, A. (2012). Pedestrian and bicycle traffic as the elements of sustainable urban transport system in Copenhagen, *Urban and Regional Transport,* 2, p. 12 -20.

Rothenbarg, J. (1970). The economics of congestion and pollution: an integrated view. *The American Economic Review: Papers and Proceedings*, 60(2), p. 114-121.

Ruchaj, M., (2012). *Algorithms for controlling acyclic traffic lights in a congested road network*. (Unpublished doctoral dissertation). Opole University of Technology.

Selwon, A. & Roman, K. (2017). The impact of Intelligent Transport Systems to reduce congestion in cities, *Buses: technology, operation, transport systems*, (18)3, p. 28-32.

Sokołowski, A. 2016. Forecasting and analysis of time series. Training materials. StatSoft Polska. Cracow.

Szołtysek J., (2011). *Creating mobility of city residents*. Wolters Kluwer Polska Sp. z o.o.

Świderski, A. (2011). *Quality evaluation modelling of transport services*, Scientific work – transport, 81, Warsaw: Publishing Office of the Warsaw University of Technology.

Taylor, H. M. & Karlin, S. (1998) *An introduction to stochastic modelling*, New York: Academic Press.