

## FORECASTING DEMAND FOR TRANSPORT SERVICES ON THE EXAMPLE OF PASSENGER TRANSPORT

**Anna Borucka**

Military University of Technology, Poland

E-mail: [anna.borucka@wat.edu.pl](mailto:anna.borucka@wat.edu.pl)

Received: April 28, 2019

Received revised: July 5, 2019

Accepted for publishing: July 8, 2019

### *Abstract*

Success on the market is fostered not only by the quality of provided services, but also by the precise satisfaction of purchasers' needs. Therefore, demand forecasting is an important element of any company function, including transport. It also allows for appropriate shaping of the level and structure of inventories. It facilitates proper organization of processes and better management of resources. This is particularly important in the transport services industry, where vehicle readiness determines possibility of performing the task. Demand is influenced by a multitude of factors, which are often difficult to define and describe, therefore this article proposes the ARIMA model, in which the conducted study was based on the assumption that the dependent variable is affected only by its own value, lagged over time. The study was supplemented by the ARIMAX model, which additionally takes into account exogenous variables resulting from the diagnosed seasonality of the process.

The analysis was presented on the example of a Polish company (based in Warsaw) offering passenger transport services, for which the number of passengers was forecast. Such information allows not only for a more efficient use of the available human and technical resources, but also for an increase in the company's profit.

**Key words:** transport services, ARIMA model, ARIMAX model, demand, forecasting

### 1. INTRODUCTION

Passenger transport in literature is most frequently approached from the legislative perspective (Abramovic et al., 2017, Gładysz et al., 2016; Stimac & Vistica, 2018; Wesołowski, 2016), which results from the fact that passenger transport is a very complex issue, conditioned by a number of regulations. Many articles discuss the quality (Klopott & Miklińska, 2017; Świdorski, 2018) and customer preferences (Kozłowska & Cygan, 2018; Mikulska & Starowicz, 2016; Naletina et al., 2018) as they constitute the driving force behind the entire industry. Mathematical modeling of transport is less popular in literature, both in terms of theoretical methods and their practical application (Sikora & Borowski, 2011; Żurkowski, 2009). Estimating

demand for transport services is a particularly interesting issue in this area, therefore the aim of this article is to indicate selected, possible to apply methods of mathematical forecasting of this phenomenon.

The transport demand in cities is influenced by a multitude of factors, often difficult to predict. Some of them are life rhythm-based and can be modeled using variables, such as time of day or hour, others such as preferences or needs of potential customers are more difficult to identify. Therefore, companies offering passenger transport services need to adapt their strategy to the high dynamics of demand as well as plan drivers' work and vehicle availability in a way ensuring that as many of them as possible are available in a situation of high demand for services, and unjustified idleness at times of lower demand is avoided. Any unnecessary downtime generates costs, and lack of transport capability when there is a demand for it means a profit lost.

Mathematical tools and methods are helpful in identifying potential demand. A number of them (such as regression models) require information on factors affecting a given phenomenon, which are often difficult to identify or measure, in order to obtain reliable forecasts. Therefore, time series models are a good solution, which, in order to build a forecast, require information only about the value of the dependent variable lagged in time. Such models include, among others, autoregressive or moving average models, as well as combinations thereof, such as ARMA, ARIMA or SARIMA. The application and comparison of the selected model shall be presented in this article.

The author's intention was to show that it is possible to offer reliable forecasts of demand for transport services even in a situation of limited access to information, and thus to better plan the use of resources and adapt them to customer needs. Furthermore, the aim was to emphasize the utilitarian nature of such analyses when applied to create the company's strategy.

## 2. TIME SERIES FORECASTING MODELS

Analysis methods of sequences of chronologically ordered information, showing a certain dependence between individual observations, are called the time series analyses. Mathematical description of this dependence makes it possible, not only to determine the nature of the studied phenomenon, but also to forecast, i.e. to predict the future values of the time series. The so-called stationary models, which assume that the analyzed process is in balance with the constant average level, occupy a special place among the models used for describing real stochastic processes. However, since many economic phenomena are of non-stationary nature, such analysis methods have also been developed, the ARIMA processes in particular (Stimac & Vistica, 2018).

ARIMA (*Autoregressive integrated moving average model*) is a model created by integration of the autoregressive model – AR and the moving average model – MA. The AR model is based on the assumption that there is an autocorrelation between the current values of the forecast variable and its values lagged in time. The current value

of the process is expressed as a finite linear combination of the previous values of the process (1) (Box & Jenkins, 1983; Devon, 2016):

$$y_t = \theta_0 + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \dots + \theta_p y_{t-p} + \varepsilon_t \quad (1)$$

where:

$y_t, y_{t-1}, y_{t-2}, y_{t-p}$  - the value of the forecast variable at the time or in the period  $t, t - 1, t - 2, \dots, t - p$ ;

$\theta_0, \theta_1, \theta_2, \theta_p$  - model parameters;

$\varepsilon_t$  - model error (residual) for the moment or period  $t$ ;

$p$ - lag length.

Whereas in the MA (*moving average*) process, the values of the endogenous variable are expressed as a function of the lagged values of the stationary random component [Chaoqing et al., 2016; Chen et al., 2016]. The parameter  $q$  of this process, i.e. the order of the MA process, indicates the level of lags adopted for the model. The form of the MA model is as follows (2):

$$y_t = \phi_0 + \varepsilon_t - \phi_1 \varepsilon_{t-1} - \phi_2 \varepsilon_{t-2} - \dots - \phi_q \varepsilon_{t-q} \quad (2)$$

where:

$y_t, y_{t-1}, y_{t-2}, y_{t-q}$  - the value of the forecast variable at the time or in the period  $t, t - 1, t - q$ ;

$\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \varepsilon_{t-q}$  - errors, residuals of the model in the periods  $t, t - 1, t - 2, \dots, t - q$ ;

$\phi_0, \phi_1, \phi_2, \phi_q$ - model parameters;

$q$ - lag length.

Mixed autoregressive–moving average models allow for greater flexibility in fitting the model to the real time series [Xin, 2017]. The form of the ARIMA model is as follows (3):

$$y_t = \theta_0 + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \dots + \theta_p y_{t-p} + \phi_0 + \varepsilon_t - \phi_1 \varepsilon_{t-1} - \phi_2 \varepsilon_{t-2} - \dots - \phi_q \varepsilon_{t-q} \quad (3)$$

If there are clear seasonal variations in the process, the SARIMA (*Seasonal ARIMA*) model can be used. It is constructed by supplementing the ARIMA model with a seasonal component. This requires the determination of three additional parameters including such a component, i.e. P – the order of seasonal lags of the AR type, Q – the order of seasonal lags of the MA type, D – the seasonal differentiation parameter. ARIMA models can also be combined with classic regression models. The result of which is the ARIMAX (*Autoregressive integrated moving average with exogenous variables*) model, in which an additional exogenous variable is included in order to improve forecasting efficiency. Therefore, the form of the ARIMAX model is the ARIMA model supplemented by a set of exogenous regressors (4) [Box & Jenkins, 1983]:

$$y_t = c + \beta x_t + \theta_0 + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \dots + \theta_p y_{t-p} + \phi_0 + \varepsilon_t - \phi_1 \varepsilon_{t-1} - \phi_2 \varepsilon_{t-2} - \dots - \phi_q \varepsilon_{t-q} \quad (4)$$

where:

$\beta$ - coefficient of variation

$x_t$ - additional exogenous variable

The procedure for estimating the parameters of the above models was developed by George Box and Gwilym Jenkins in the 1970s. The proposed algorithm consists of

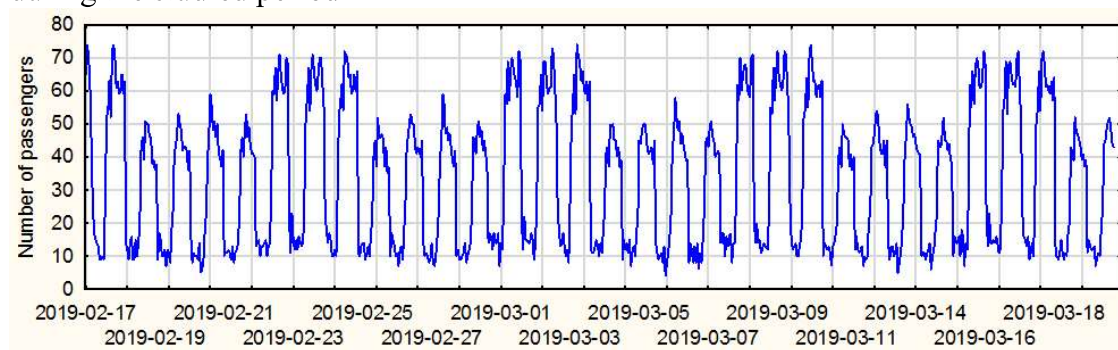
the following stages of time series analysis: identification, estimation, verification, forecast determination, according to which this analysis was conducted.

### 3. MODELING METHOD

#### 3.1. Arima Model

According to the methodology of statistical research (Bielińska, 2007; Taylor & Karlin, 1998), the first stage is the visual evaluation of the time series, presented in Fig. 11. The analyzed phenomenon is characterized by a clearly visible seasonality of the process, as a result of which the series is not stationary.

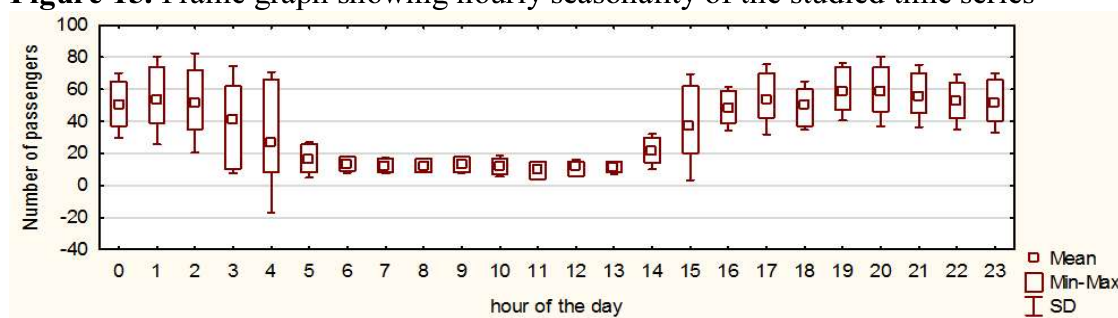
**Figure 14.** Graph presenting the number of passengers using transport services during the studied period



Source: own study

A detailed analysis revealed that this seasonality is primarily determined by the time of provision of the service. It turns out that at different times of day the demand for transport is different, which is shown in the frame graph in Fig. 12 on which the analysis of the examined variable was made depending on the time of its execution.

**Figure 15.** Frame graph showing hourly seasonality of the studied time series

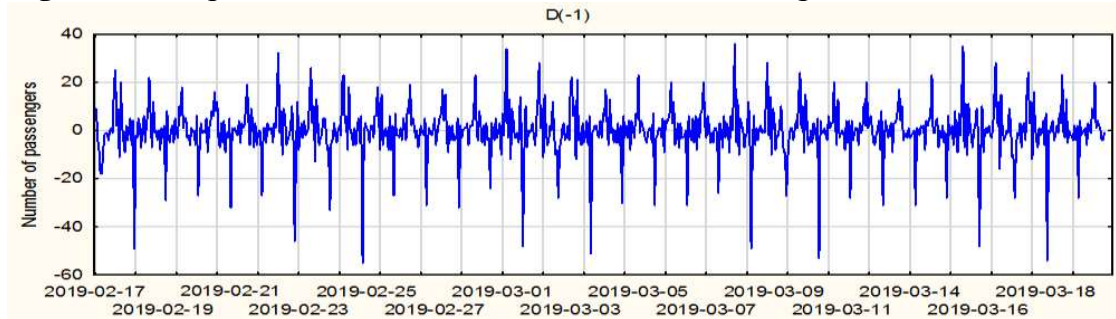


Source: own study

The variation in the number of people using the company's services changes over the course of the day, making the process non-stationary. Since ARIMA methods can only be used for stationary series or reduced to stationary series, it is necessary to achieve at least stationarity in a broader sense (unchangeability during the first and second moment) (Box & Jenkins, 1983). Therefore, in order to smoothen the expected

value and variance, a differentiation of the time series was applied. The graph of the variable after differentiation is shown in Fig. 13.

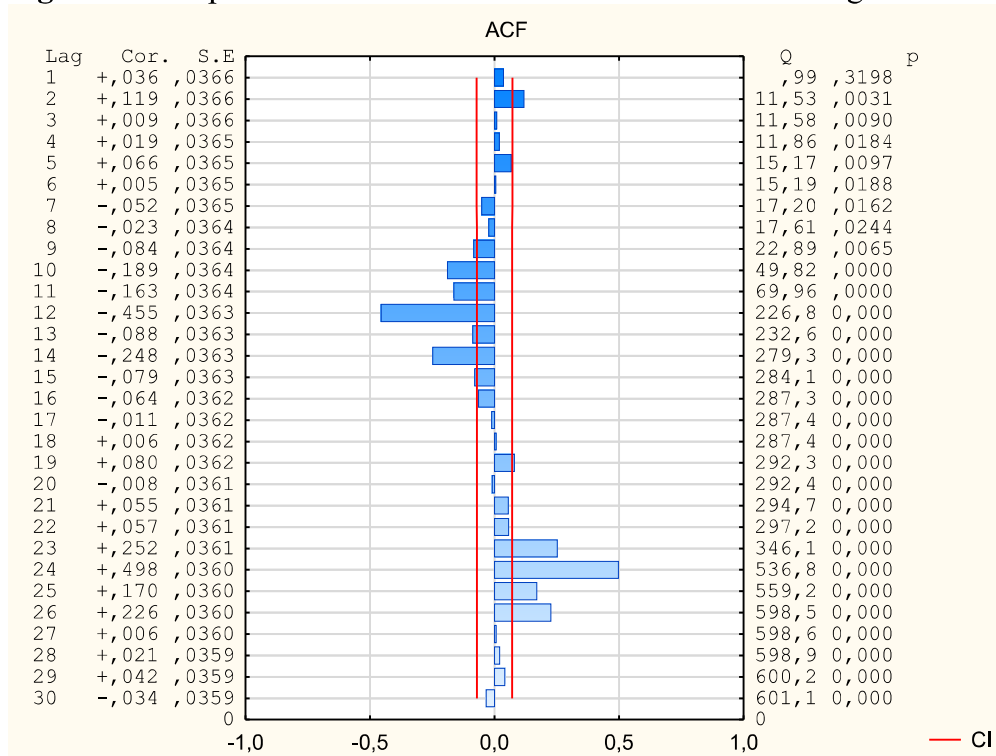
**Figure 16.** Graph of the series after differentiation with a lag d=1



Source: own study

The next step is to identify the appropriate subclass of ARIMA models by determining the initial values of their parameters. The basic tools in this respect are graphs of an autocorrelation function (ACF) and a partial autocorrelation function (PACF). The ACF values shown in the Fig. 14 disappear quickly for the initial lags and the significant value for lag 2 suggests that the value of the moving average parameter will be  $q=2$ . However, the values for the lag  $d=12$  and its multiplicity are clear, which indicates a strong seasonality of the process, necessary to be included in the model.

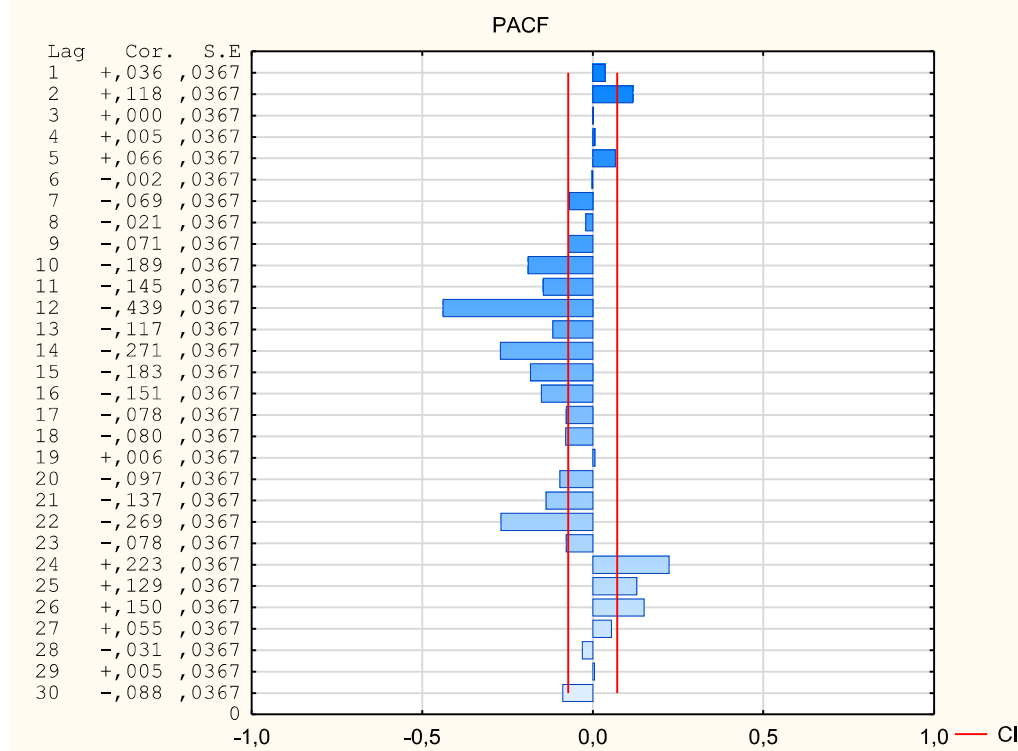
**Figure 17.** Graph of autocorrelation function with off-season lag d=1



Source: own study

The PACF graph (Fig. 15) leads to identical conclusions in terms of seasonality, also showing high indications for lag  $d=12$ . In addition, it suggests the value of the autoregressive element  $p=2$ .

**Figure 18.** Graph of partial autocorrelation function with nonseasonal lag  $d=1$



Source: own study

Finally, it was decided to estimate the parameters of the ARIMA model using the *testing down* method, assuming that it is a moving average process with a non-stationary parameter  $q = 2$  and a non-stationary autoregressive parameter  $p=2$ . Additionally, seasonal parameters were taken into account, which resulted in transformation of the assumed ARIMA model into a seasonal model, i.e. SARIMA. The two best models were selected from all tested models, for which all model parameters were statistically significantly different from zero. The results of the estimation are presented in Table 3.

**Table 3.** Results of the ARIMA model parameters estimation

Parameter	The ARIMA model (2,1,1)(2,0,0)	The ARIMA model (2,1,2)(2,0,0)
transformations	D(-1)	D(-1)
p(1)	0.56	-0.59
p(2)	0.13	0.35
q(1)	0.94	-0.21
q(2)		0.75
Ps(1)	-0.31	-0.35

Parameter	The ARIMA model (2,1,1)(2,0,0)	The ARIMA model (2,1,2)(2,0,0)
Ps(2)	0.48	0.49
Seasonal lag	12	12
MS	65.8	66.1
MRPE	27.5	28
AIC	5630.94	6699.33

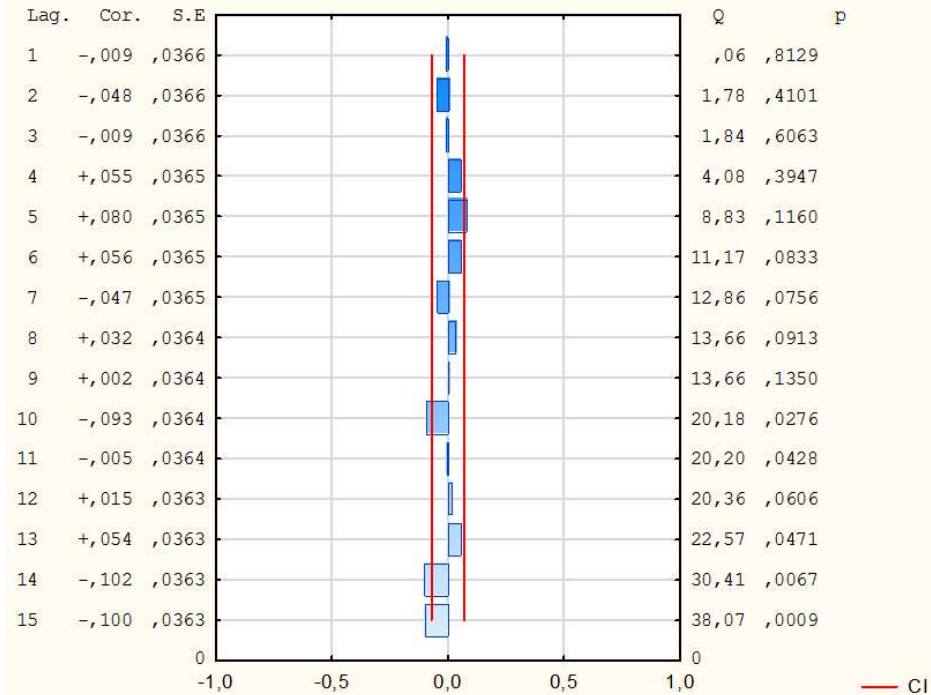
Source: own study

For the proposed models estimation errors and Akaike information criterion were calculated, and on the basis of these criteria (MS-Mean square error, MRPE - Mean relative prediction errors, AIC - Akaike Information Criterion) the best one was selected, which turned out to be the ARIMA model (2.1.1) (2.0.0) in the form (5):

$$y_t = 0.94 y_{t-1} + \varepsilon_t - 0.56 \varepsilon_{t-1} - 0.13 \varepsilon_{t-2} + 0.31 \varepsilon_{t-12} - 0.48 \varepsilon_{t-13} \quad (5)$$

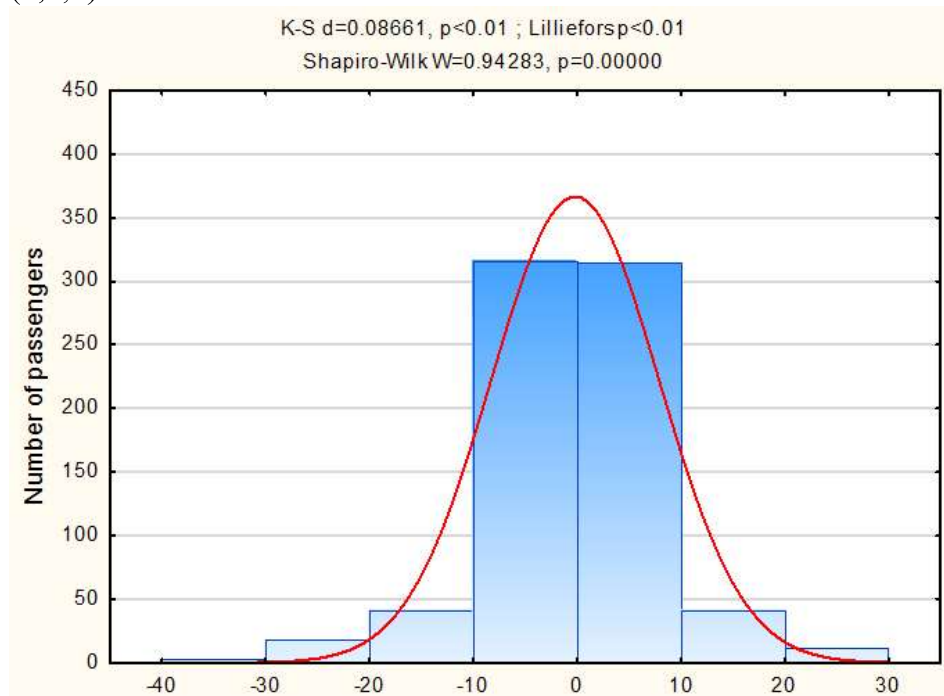
Then it was diagnosed by analyzing the distribution of residuals. In a correctly constructed model, the residuals should be random and symmetrical. In order to examine these features, the autocorelogram (Fig. 16) and the histogram of the distribution of residuals (Fig. 17) were drawn up. The ACF shows few statistically significant indications suggesting that not all dependencies were explained by the proposed model. Compatibility with normal distribution was not confirmed either. The test statistic in the Shapiro-Wilk test was  $W=0.94283$ , giving p-value  $p=0.0000$ .

**Figure 19.** ACF graphs for the residuals of the ARIMA model (2,1,1) (2,0,0)



Source: own study

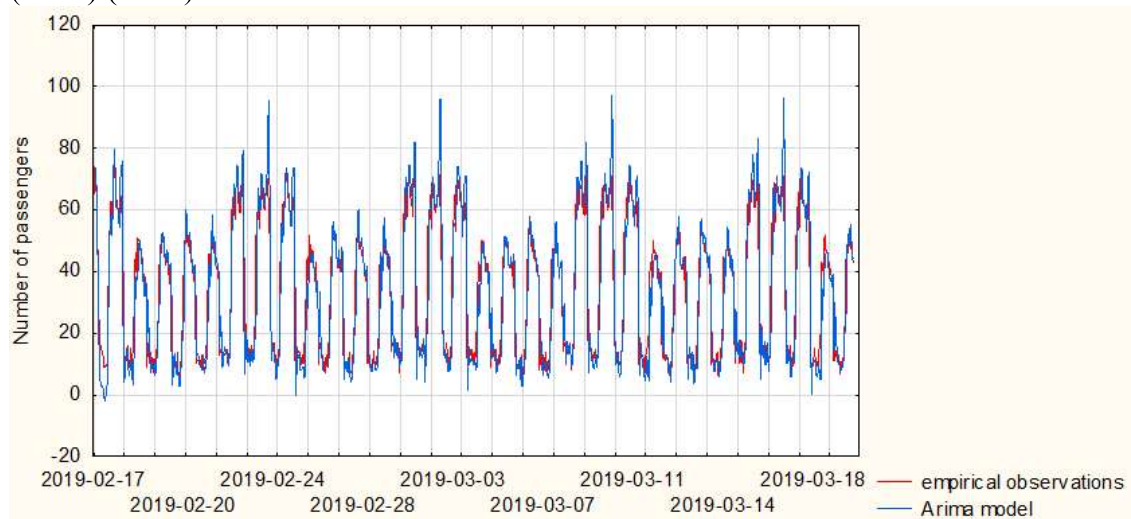
**Figure 20.** Histogram of the distribution of residuals of the ARIMA model (2,1,1) (2,0,0)



Source: own study

The above results explain why the forecast function differs from empirical observations. The largest errors relate to the maximum observation values achieved for peak transport service demand. The forecast for these hours is definitely overestimated 18.

**Figure 21.** Graph of the examined series and forecast according to the ARIMA model (2.1.1) (2.0.0)



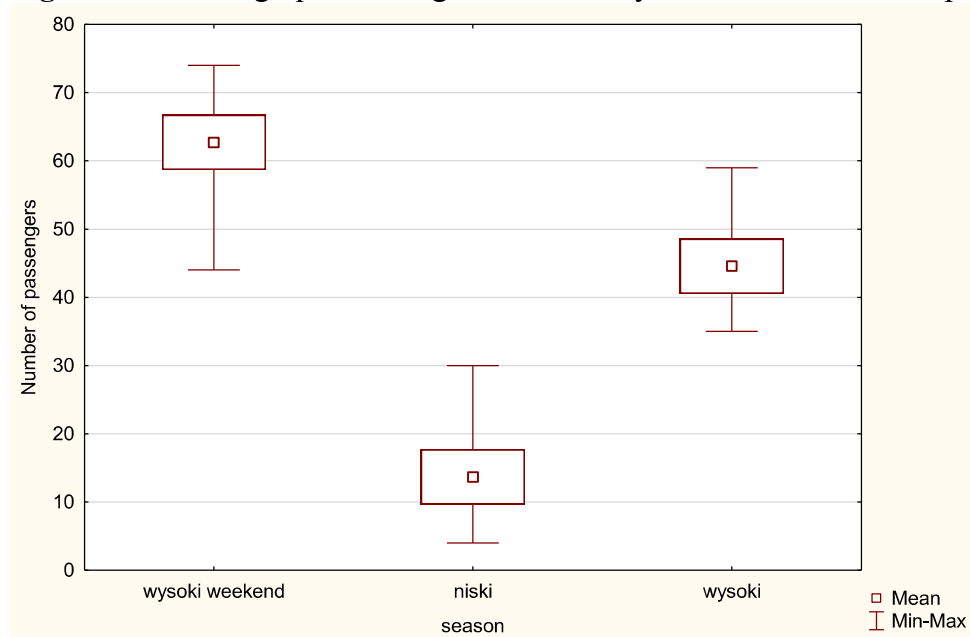
Source: own study



### 3.2. ARIMAX Model

In order to improve the obtained prognosis [Niematallah & Mototsugu, 2018], a modification of the model, allowing to include additional exogenous variables, has been proposed, i.e. the ARIMAX model [Sutthichaimethee & Ariyasajakorn, 2017; Wiwik, 2015]. Since the hourly variation shown in Figure 2 is clearly arranged into three separate periods, as confirmed by Fig. 19, it was divided into three levels, resulting from the diverse customer interest. First, low demand is defined as a group of hours during which the number of customers does not exceed 30, followed by a high season on weekdays with a maximum of 49 customers and a high season on weekends starting on Friday afternoons with a maximum of 70 customers.

**Figure 22.** Frame graph showing the seasonality of the examined transport services



Source: own study

The selected regressors are qualitative variables, thus it was necessary to re-code them into binary variables (zero-one values). The parameters of thusly constructed model were estimated. The results are presented in table 4.

**Table 4.** Results of the ARIMAX model parameters estimation

Parameter	ARIMAX model
C	-0.03
p(1)	0.22
q(1)	0.21
$\beta_1$	53.90
$\beta_2$	9.39
$\beta_3$	35.99

Parameter	ARIMAX model
MS	30.43
MRPE	17.7
AIC	4553.05

Source: own study

The model obtained this way has the following form

$$y_t = -0.03 + 53.9 x_1 + 9.39 x_2 + 35.99 x_3 + 0.22 y_{t-1} + \varepsilon_t - 0.21 \varepsilon_{t-1} \quad (6)$$

where

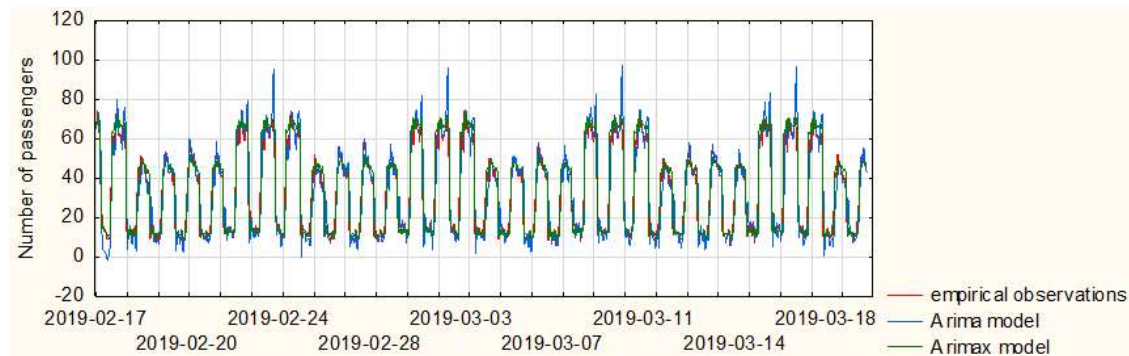
$x_1$ - the binary variable corresponding to high values at the weekend

$x_2$ - the binary variable corresponding to the low season on weekdays

$x_3$ - the binary variable corresponding to the high season on weekdays

According to the criteria adopted for verification (MS error, MRPE, AIC), the ARIMAX model achieved the most satisfactory results out of all the proposed models, as confirmed by Fig. 20, presenting empirical data and forecast functions according to ARIMA and ARIMAX models.

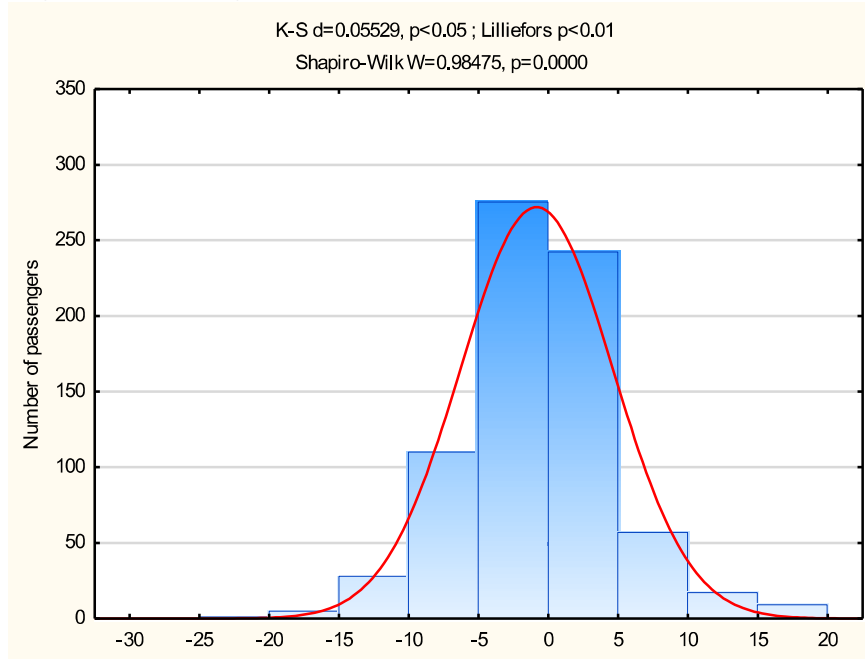
**Figure 23.** Graph of empirical values and forecast functions according to ARIMA and ARIMAX models



Source: own study

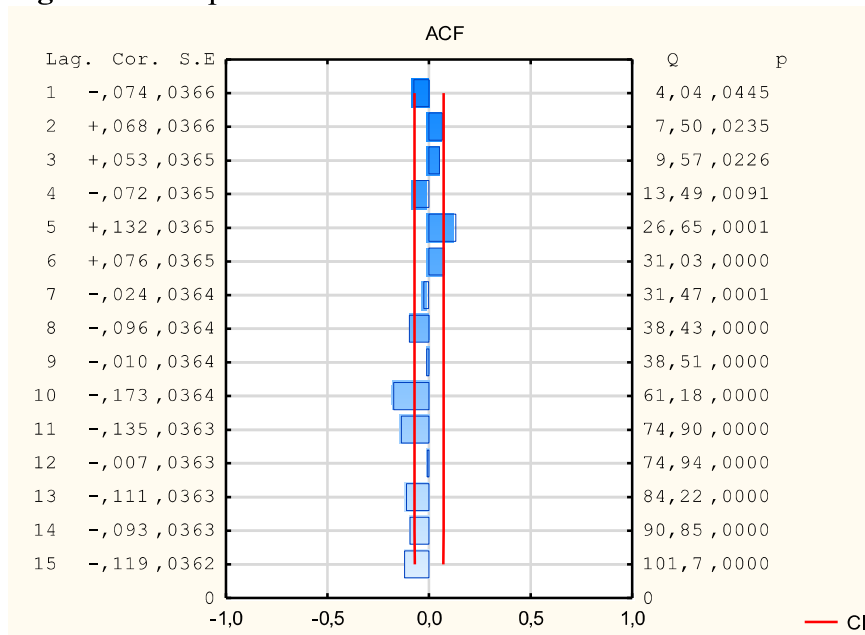
The ARIMAX model is better suited to real observations, but its diagnosis is not satisfactory because the distribution of residuals is still not close to the normal distribution (Fig. 21) and the graph of the autocorrelation function shows significant values of this function (Fig. 22).

**Figure 24.** Histogram of distribution of the residuals of the ARIMAX model



Source: own study

**Figure 25.** Graph of the autocorrelation function of the ARIMAX model residuals



Source: own study

### 3.3. Forecasting According To the ARIMA and ARIMAX Models

In the last stage of the study, the future values of the series were forecast and compared with the retained test observations, which did not participate in the estimation of model parameters. A relative forecast error was also calculated for each predicted value. The obtained results are presented in table 5.

**Table 5.** Forecasts according to the ARIMA and ARIMAX models

date/time(hour)	empirical observatio	ARIMA forecast	ARIMAX forecast	SE ARIMA	SE ARIMAX
20-03-19 0:00	41	40.3	45.1	1.6%	10.0%
20-03-19 1:00 AM	45	41.5	46.0	7.8%	2.3%
20-03-19 2:00 AM	40	36.6	46.2	8.4%	15.6%
20-03-19 3:00 AM	45	36.0	46.3	20.1%	2.9%
20-03-19 4:00 AM	17	16.3	19.7	4.3%	15.9%
20-03-19 5:00 AM	12	17.8	13.8	48.6%	14.7%
20-03-19 6:00 AM	13	15.6	12.4	19.9%	4.3%
20-03-19 7:00 AM	13	14.5	12.1	11.3%	6.6%
20-03-19 8:00 AM	12	14.1	12.1	17.8%	0.6%
20-03-19 9:00 AM	13	15.8	12.1	21.9%	7.2%
20-03-19 10:00 AM	14	14.2	12.1	1.5%	13.9%
20-03-19 11:00 AM	7	15.9	12.1	127.7	72.2%
20-03-19 12:00 PM	9	17.2	12.1	91.6%	34.0%
20-03-19 1:00 PM	11	16.4	12.1	49.1%	9.6%
20-03-19 2:00 PM	13	22.2	12.1	70.8%	7.3%
20-03-19 3:00 PM	20	24.8	12.1	24.1%	39.7%
20-03-19 4:00 PM	42	40.5	38.7	3.6%	8.0%
20-03-19 5:00 PM	45	40.5	44.6	10.0%	0.9%
20-03-19 6:00 PM	53	43.1	45.9	18.7%	13.4%
20-03-19 7:00 PM	52	44.4	46.2	14.6%	11.1%
20-03-19 8:00 PM	50	45.0	46.3	10.0%	7.4%
20-03-19 9:00 PM	48	42.5	46.3	11.4%	3.5%
20-03-19 10:00 PM	46	41.1	46.3	10.6%	0.7%
20-03-19 11:00 PM	46	40.1	46.3	12.8%	0.7%

Source: own study

The average forecast error for the ARIMA model was as high as 25.7%, while for the ARIMAX model it was much lower, at 12.6%. The higher effectiveness of the ARIMAX model is presented in Fig 23.

**Figure 26.** Graph of test observations and forecasts according to the ARIMA and ARIMAX models



Source: own study

#### 4. FINAL THOUGHTS

The presented process is characterized by high complexity, causing difficulties in estimation of parameters of econometric models. However, in companies similar to the examined one, where a limited number of vehicles and drivers providing transport services require the adjustment of their working hours to market demand, even an estimation with a certain error can prove to be very useful. Although the proposed models took into account only the dependence of the examined variable on its value lagged in time (ARIMA model) and additionally, the selected element of seasonality of provided services (ARIMAX model), they are a sufficient guidance to assess the development of demand and may determine the direction of the company's strategy.

The models presented here can be applied in forecasting demand in nearly every company. Depending on the needs, they can be extended with additional variables, increasing the forecast accuracy and reducing estimation errors. However, identifying and measuring all the factors that shape the demand for certain goods or services is not always possible. Often the process of collecting and processing such factors is difficult and sometimes even impossible. Nevertheless, as demonstrated in this article, even a simplified model can provide valuable guidance on the use of available resources (in this case vehicles and personnel) in order to respond more effectively to customer demand. This may translate into an increase in the quality of services provided and customer satisfaction and, as a result, into an increase in the company's profits.

#### 5. REFERENCES

Abramovic, B. & Nedeliakova, E., Panak, M. & Sipus, D. (2017). Synergy In Logistics Processes For Railway Transport, *Business Logistics in Modern Management, University of Osijek, Faculty of Economics, Croatia*, 17, p. 15-28.

- Barrow, D.K. (2016). Forecasting intraday call arrivals using the seasonal moving average method, *Journal of Business Research*, 69(12), p. 6088-6096.
- 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: PWN.
- Chen, S., Lei, L. & Tu, Y. (2016). Functional coefficient moving average model with applications to forecasting Chinese CPI. *Statistica Sinica*, 26(4), p. 1649-1672.
- Elamin, N. & Fukushige, M. (2018). Modeling and forecasting hourly electricity demand by SARIMAX with interactions, *Energy*, 165(B), p. 257-268.
- Gładysz, A., Czech, P., Łazarz, B., Cieśla, M. & Turoń, K. (2016). Legal conditions for regular passenger transport, *Autobusy : technika, eksploatacja, systemy transportowe*, 2016, 17(6), p. 1343-1347.
- Klopott, M. Miklińska, J. (2017). Initiatives and practices for greening logistics services on the polish market – status quo and development opportunities, *Business Logistics in Modern Management, University of Osijek, Faculty of Economics*, 17, p. 523-536.
- Kozłowska, M. & Cygan, G. (2018). Demand on the passenger transport market in the European Union, *TTS Technika Transportu Szybowego*, 4(25), p. 12-17.
- Mikulska, A. & Starowicz, W. (2016). Method for determining the annual number of passengers in public transport in Cracow based on the number of tickets sold, *Transport Miejski i Regionalny*, 12, p. 19-24.
- Naletina, D., Petljak, K. & Kapovic, A. (2018). Influence Of Liberalization On Air Transport Market: Research Into Consumers' Satisfaction With Low-Cost Airlines Services In The Republic Of Croatia, *Business Logistics in Modern Management, University of Osijek, Faculty of Economics, Croatia*, 18, p. 21-40.
- Sikora, M. & Borowski, Z. (2011). Assessment of urban logistics concerning public transport in Bydgoszcz, *Logistyka*, 2, p. 529-240.
- Stimac, H. & Vistica, J. (2018). Transport Policy From Perspectives Of Passengers, Cargo, Energy: Croatia Vs European Union Countries, *Business Logistics in Modern Management, University of Osijek, Faculty of Economics, Croatia*, 18, p. 55-72.
- Sutthichaimethee, P. & Ariyasajjakorn, D. (2017). Forecasting Energy Consumption in Short-Term and Long-Term Period by Using Arimax Model in the Construction and Materials Sector in Thailand. *Journal of Ecological Engineering*, 18(4), p. 52-59.
- Świdorski, A. (2018). *Inżynieria jakości w wybranych obszarach transportu*. Motor Transport Institute, Warsaw.
- Taylor, H. M. & Karlin, S. (1998). *An introduction to stochastic modelling*, New York: Academic Press.

Wesołowski, K. (2016). The Plurality of the Deciding Centers in the Law System Concerning Services of Passengers Carriage, *Marketing i Zarządzanie*, 42, p. 201-215.

Wiwik, A., Vinarti, R.A. & Kurniawati, Y.D. (2015). Performance Comparisons between Arima and Arimax Method in Moslem Kids Clothes Demand Forecasting: Case Study, *Procedia Computer Science*, 72, p. 630-637.

Xin, L., Bing, P., Rob, L. & Xiankai H. (2017). Forecasting tourism demand with composite search index, *Tourism Management*, 59, p. 57-66.

Yuan, C., Liu, S. & Fang, Z. (2016). Comparison of China's primary energy consumption forecasting by using ARIMA (the autoregressive integrated moving average) model and GM(1,1) model. *Energy*, 100, p. 384-390.

Żurkowski, A. (2009). Inter-Agglomeration Carriage Simulations, *Problemy Kolejnictwa*, 148, p. 5-47.