ISSN 1509-5878
Nr 3
3/2013
AUTOBUSY
Technika, Eksploatacja, Systemy Transportowe

Štefan LIŠČÁK, Lukáš HOLEŠA, Jacek CABAN

ANALYSIS OF IMPACT OF RESIDUAL COMPONENT WHEN FORECASTING THE DEMAND OF PASSENGERS BY MEANS OF UNIDIMENSIONAL TIME SEQUENCES OF ADDITIVE TYPE

Abstract

The article deals with the issue of mathematical – statistical modelling of passengers demand for suburban bus transport. It is based on time sequences of additive type with linear trend component with monthly season rate. We have managed to create in the paper a short-term prognosis for the selected type of transport, with the aim to point out the justification of adding the random (residual) component into the model for creating the prognosis. For practical reasons a specific mass public transport provider has been selected to contribute necessary data in order to elaborate the study.

INTRODUTION

The mass transport of people forms an important part of the transportation system as whole, national economy and is basically considered a material service for the public. It has an important and significant role in performing important functions of inhabited places and areas. It is important for the provision of large transport demands and at the same time it has no big demands to transport areas, it can provided better safety of travelling and has smaller negative impact on the environment as calculated per one transported person.

The role of the mass public transport itself is given by its properties in relation to the satisfaction of transport needs of the population living in the respective area, to environmental impacts and investment demands for the traffic infrastructure. Especially today, in the time of significant growth of individual motorization, we should consider possibilities for improvement especially of the mass transport quality, the technical equipment, technology and management organization of the same, on the grounds of which we can expect beneficial division of transportation labour in favour of the mass public transport and increase of passenger demand for it.

The analysis of data associated with the intensity of vehicle use is applied in the evaluation of a given transport system [2]. The following contribution deals with the issue of mathematical – statistical modelling of passenger demand using single-dimensional time sequences with the aim of the most accurate prediction possible for the future. This prediction is important for transporters from the already mentioned point of view of quality.

1. PROBLEM DEFINITION

In order to meet the objective of this contribution, which rests in the creation of a short-term prognosis of demand for the following period of one year with the subsequent examination of the impact of the non-systemic part in the proposed model, we have chosen a particular mass public transport operator. This transporter operates in the area of mass public transport, namely in following cities:

- Liptovský Mikuláš,
- Ružomberok,
- Dolný Kubín.

In addition to it, it also provides international regular and irregular bus transport and urban bus transport in following districts:

- Dolný Kubín,
- Námestovo,
- Ružomberok,
- Trstená,
- Liptovský Mikuláš.

Calendar-cleared empirical values of the indicator of the number of transported passengers for the period from 1 January 2006 until 31 December 2011 from the said transporter were used for practical elaboration. Values relate to numbers of transported passenger by urban bus transport in the district Dolný Kubín according to fare divided to student and full fare.

A method based on quantitative or mathematical – statistical point of view was chosen for the elaboration of the prognosis model. Usage of multi-criterion linear regressive model with artificial, (0-1) seasonability explaining variables, proved to be the most advantageous for the elaboration of the prognosis with regard to statistical significance of the whole model (determination coefficient), as well as individual parameters of the same.

Software program SAS 9.1.3 was used for mathematical expression of the model. Linear regression models for determination of point prognosis and interval estimate were created separately for time sequences of transported passenger travelling for student fare (SF) and separately for full fare. Further models for both fare types were created so as including in addition to trend and seasonal component also random (residual) components. Based on such additional modelling of the residual component and using decomposition approach, the total impact of such component in the model, as well as impact on original prognoses without it, could be assessed. Due to large extent of elaboration we include in the contribution only results of processed data for student fare for demonstrative purposes.

2. USED MATERIAL, METHODS AND EMPIRICAL DELIVERABLES SOLUTION

This chapter describes data inputs for decomposition task including performed modification, testing method of properties of examined time sequences, proposed models as well as empirical outputs.

2.1. Model of regression task for pupil travel

In order to choose a suitable type of regression task as specified in [1], it is necessary to know the given type of time sequence. Time sequences used in this contribution are additive time sequences with linear trend component with monthly seasonability. The chosen multi-criterion regression task using decomposition approach is based on the basic structure of short-term single-dimensional time sequences of additive type, where we subsequently transformed the model shape for our needs to:

$$\hat{y} = \hat{b}_0 + \hat{b}_1 \cdot t_k + \hat{b}_2 x_{2t} + \ldots + \hat{b}_k x_{kt} + \varepsilon_t \quad (1)$$

Where: $\hat{y}$ – modelled value of examined indicator [persons / month],

$\hat{b}_0$ – estimate of initial variable (specifies value of $\hat{y}$ in time $t = 0$),

$\hat{b}_k$ – parameter expressing the constant change of dependent variable $\hat{y}$ induced by the increase of value of respective time variables $k = 1, 2, 3, ...12$,

ε_t – random (residual) component,

t_k – order number of time period,

x_{kt} – artificial variables taking into account seasonability of time sequences.

The said regression model can be mathematically divided to systemic and non-systemic part (see Fig. 1).

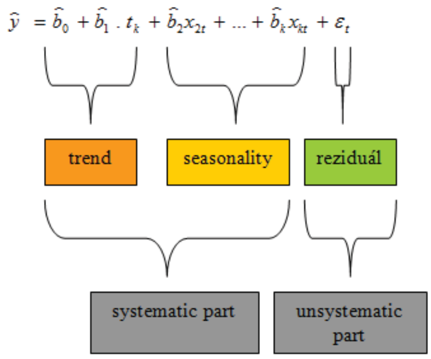

Fig. 1. The division of regression model

Source: Authors

Creating a model for pupils travel without a random component

Outputs from the SAS program (see tab. 1) show us the most important indicators relating to the said model. They include determination coefficient (R^2), which characterizes the overall quality of the proposed model, values of parameters (in table 1 – par 2 to 12), as well as the estimate of parameters of the trend component.

Tab. 1. Output from the software program SAS for SF

Model parameters	Values	Error middle value	T	Prob>|T|
Limits	94823	3524	26.9099	0.0001
Par 2	10636	4315	2.4651	0.0166
Par 3	3954	4310	0.9174	0.3627
Par 4	20489	4306	4.758	0.0001
Par 5	14457	4303	3.3601	0.0014
Par 6	17557	4299	4.0837	0.0001
Par 7	12381	4297	2.8816	0.0055
Par 8	-59470	4294	-13.8488	0.0001
Par 9	-57521	4292	-13.4008	0.0001
Par 10	16297	4291	3.7981	0.0003
Par 11	26816	4290	6.2511	0.0001
Par 12	19829	4289	4.623	0.0001
Linear trend	-236.02044	42.7188	-5.525	0.0001
R^2	0.94			

Source: Table elaborated by authors by means of software program [4]

On the grounds of knowledge of the most important parameters, as well as parameters (b_0, $b_1 \ldots b_{12}$) for modelling the trend and seasonable component of monthly intensity of transported people, the respective multi-criterion regression model characterizing the dependence $\hat{y}_{\dot{z}c}$ was determined:

$$\hat{y}_{SF} = 94\,823 - 236{,}02044\,.\,t + 10\,636x_{2t} + 3\,954x_{3t} + 20\,489x_{4t} + 14\,457x_{5t} + 17\,557x_{6t} + 12\,381x_{7t} - 59\,470x_{8t} - 57\,521x_{9t} + 16\,297x_{10t} + 26\,816x_{11t} + 19\,829x_{12t} \quad (2)$$

The said model describes, as is evident also from the relatively high value of the determination coefficient (R^2 = 0.94) app. 94% variability of used time sequences. The following diagram (Fig. 2) shows courses of curves of the real status of transported passengers (blue curve) for the period from 1 January 2006 until 31 December 2011 and modelled status (pink curve) created on the basis of relation 2. So-called artificial variables reflecting the seasonability, which occurs in time sequences (x_{nt}), were also incorporated in the relation.

The table 2 subsequently describes input data, which are in addition to time sequences of the number of transported passengers during individual time periods necessary for the estimate of modelled values, on the grounds of setting them in the relation 2. The relation 2 describes the trend component of the created model (94,823 – 236.02044.t), where values from the second column of the table are substituted for *t*. The said column (*t*) actually expresses numerical sequence of individual periods (months), during which passengers were transported. The number of monitored periods is for our case $n = 72$.

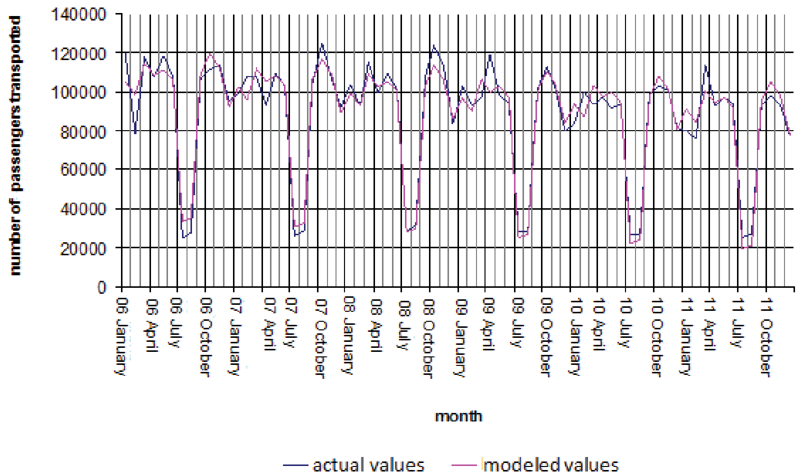

Fig. 2. Real and modelled number of passenger for SF

Source: Data provided by the companies concerned and the calculations of the model 2

Other columns (x_2, x_3, ... x_{12}) contain mentioned artificial variables, which reflect the monthly seasonability of time sequences of the number of transported passengers. The number of artificial variables added to the parameter *t* was chosen in the extent $x = 11$ with regard to creation of a short-term extrapolation to the period of one year (2012) or 12 months and due to the fact that the variable for the month December is a control variable *t* ($x_{12} = 0$). Artificial variables $x_k = 1$ express the numerical order from the column *t*. The said implies that analogically according to (Fig. 1) the seasonal component of the model is formed by its second part (see Relation 3).

Tab. 2. Input data (example) for the calculation of modelled values of SF

Month	t	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
January 06	1	1	0	0	0	0	0	0	0	0	0	0
February 06	2	0	1	0	0	0	0	0	0	0	0	0
March 06	3	0	0	1	0	0	0	0	0	0	0	0
April 06	4	0	0	0	1	0	0	0	0	0	0	0
May 06	5	0	0	0	0	1	0	0	0	0	0	0
June 06	6	0	0	0	0	0	1	0	0	0	0	0
July 06	7	0	0	0	0	0	0	1	0	0	0	0
August 06	8	0	0	0	0	0	0	0	1	0	0	0
September 06	9	0	0	0	0	0	0	0	0	1	0	0
October 06	10	0	0	0	0	0	0	0	0	0	1	0
November 06	11	0	0	0	0	0	0	0	0	0	0	1
December 06	12	0	0	0	0	0	0	0	0	0	0	0
January 07	13	1	0	0	0	0	0	0	0	0	0	0
February 07	14	0	1	0	0	0	0	0	0	0	0	0
March 07	15	0	0	1	0	0	0	0	0	0	0	0
April 07	16	0	0	0	1	0	0	0	0	0	0	0
May 07	17	0	0	0	0	1	0	0	0	0	0	0
June 07	18	0	0	0	0	0	1	0	0	0	0	0
July 07	19	0	0	0	0	0	0	1	0	0	0	0
August 07	20	0	0	0	0	0	0	0	1	0	0	0
September 07	21	0	0	0	0	0	0	0	0	1	0	0
October 07	22	0	0	0	0	0	0	0	0	0	1	0
November 07	23	0	0	0	0	0	0	0	0	0	0	1
December 06	24	0	0	0	0	0	0	0	0	0	0	0
January 08	25	1	0	0	0	0	0	0	0	0	0	0
February 08	26	0	1	0	0	0	0	0	0	0	0	0
March 08	27	0	0	1	0	0	0	0	0	0	0	0
April 08	28	0	0	0	1	0	0	0	0	0	0	0
May 08	29	0	0	0	0	1	0	0	0	0	0	0
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	
October 11	70	0	0	0	0	0	0	0	0	0	1	0
November 11	71	0	0	0	0	0	0	0	0	0	0	1
December 11	72	0	0	0	0	0	0	0	0	0	0	0

Source: Table compiled by authors

$$10\,636x_{2t} + 3\,954\,x_{3t} + 20\,489x_{4t} + 14\,457x_{5t} + 17\,557x_{6t} + 12\,381x_{7t} - \\ - 59\,470x_{8t} - 57\,521x_{9t} + 16\,297x_{10t} + 26\,816x_{11t} + 19\,829x_{12t} \qquad (3)$$

Statistical significance of the model as a whole (determination coefficient) can be verified by means of F – statistics, formula 4. It compares the calculated value F_r with the respective table value of the F – statistics. In general it applies that if F_r is bigger than the table value upon the chosen level of significance α and freedom degrees k and $n - (k + 1)$, the respective regression model is considered statistically significant and can be used for creating a prognosis (extrapolation) [3]. In our case, when F_r = 1250.75 and with freedom degrees 1 and 70 and chosen level of significance α = 0.05 (table value F = 3.9778), we can deem this condition (see relation 5) fulfilled. In other words, we refuse the zero hypothesis H_0 about the statistical insignificance of the determination coefficient (R^2). Individual parameters of this model as well as others were assessed on the grounds of p – value, what is described in detail in the following chapter.

$$F_r = \frac{\left(\frac{R^2}{k}\right)}{\left[\frac{(1-D^2)}{[n-(k+1)]}\right]} \tag{4}$$

Where: R^2 – determination coefficient [-],
n – number of observations (n=72) ,
k – number of explaining variables in the model (degrees of freedom).

$$F_r \rangle F_{\alpha,k,[n-(k+1)]} \qquad 1250{,}75 \rangle 3{,}9778 \tag{5}$$

Where: F_r – calculated value of testing statistics [-],
$F_{\alpha, k, [n-(k+1)]}$ – table value [-],
α – chosen level of significance (0.05),
k – number of degrees of freedom (1),
n – number of observed periods (n=72).

Factorial extrapolation of the created a model for student fare

Point prognosis and interval estimate of the number of transported passengers were determined for 2012 on the grounds of created model and verifications of the same. Both prognosis and estimate were made by leaving out the random (residual) component from the model. The extrapolation itself means mechanical prolongation of numerical values used before in the model (tab. 2) so as to logically follow one another.

Tab. 3. Prognosis of the model for SF for 2012 (persons/month)

Month	Composed of prognosis	U 95	L 95
January	88230	102790	73670
February	88231	95871	66752
March	88232	112170	83050
April	88233	105902	76783
May	88234	108767	79647
June	88235	103354	74235
July	88236	31267	2147
August	88237	32980	3861
September	88238	106562	77442
October	88239	116845	87725
November	88240	109622	80502
December	88241	89557	60437

Source: Table elaborated by authors on the grounds of model 2 and software program

So it was necessary to numerically prolong the column t of the table 2 by substituting order number above n = 72, together with substituting respective constants (0,1) of columns x_2 … x_{12}, so as they logically followed previous values. These prolonged values were then used in the relation 2 for creating the prognosis (see relation 6 for the month January and according to it analogically for other months). Prognosticated values together with the interval estimate for 2012 are calculated in table 3.

$$\hat{y}_{SF} = 94\,823 - 236{,}02044 \,.\, 73 + 10\,636 \,.\, 1 + 3\,954 \,.\, 0 + 20\,489 \,.\, 0 + 14\,457 \,.\, 0 + \\ + 17\,557 \,.\, 0 + 12\,381 \,.\, 0 - 59\,470 \,.\, 0 - 57\,521 \,.\, 0 + 16\,297 \,.\, 0 + 26\,816 \,.\, 0 + \\ + 19\,829 \,.\, 0 = 88\,230 \; persons \tag{6}$$

Fig. 3 on the following page shows courses of curves of prognosticated values (pink curve) for the period 2006 – 2012 and interval estimate generated by the software program SAS for the period 2012, where the red curve reflects the upper limit (U 95) and the green curve lower limit (L 95) of the interval prognosis of the number of transported passengers. The interval estimate implies that modelled transport performance will range in the given interval with 95% probability (see tab. 3, columns U 95 and L 95).

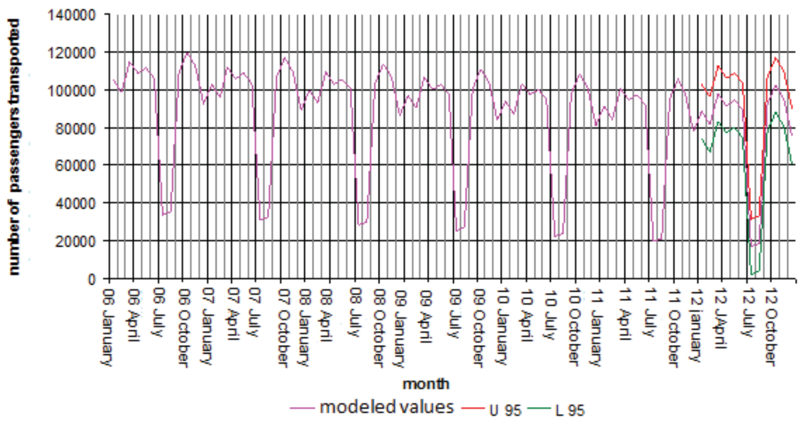

Fig. 3. Point and interval prognosis without the random component for SF for 2012

Source: Prepared on the basis of calculations of authors and software [4]

Inclusion of random (residual) component in the model for student fare

We considered in this sub-chapter in addition to the systemic part of the model (trend, seasonability) also its non-systemic part, namely the random (residual) component. The residue is basically the difference between the real and modelled value of the number of transported passengers. Mathematical record can be defined as follows:

$$\varepsilon_t = y_{\check{Z}C} - \hat{y}_{\check{Z}C} \tag{7}$$

Where: ε_t – residual member [persons],
$y_{\dot{z}c}$ – real value of the number of transported passengers [persons],
$\hat{y}_{\dot{z}c}$ – modelled value of the number of transported passengers [persons].

A model was proposed on the grounds of the software program SAS, which model connects these two parts and basically corresponds to the model 1 with regard to structure. This model includes:

– linear trend,
– artificial variables reflecting seasonability,
– autoregression of the sequence *p* (AR *n*).

In order to determine the number of coefficient of autocorrelation of residues autoregression equation of residues of sequence *p* (8) we used as basis the studies [1] and [5].

The Godfrey test of serial correlation was performed prior to quantification of coefficients of the autoregression equation, on the grounds of which the highest possible degree of serial correlation of residues of time sequences (admissible number of autocorrelation coefficients) was determined. The subsequent determination of the exact sequence of autocorrelation of residual values of time sequences and confirmation of either positive or negative autocorrelation was made on the grounds of Durbin – Watson (DW) tests of serial correlation (relation 9 is used as a basis). The autocorrelation means that each observation is statistically dependent on the previous one (meaning that if any value grows, the following one will be growing as well and vice versa - each value is linear dependent on the previous value).

The software program SAS determined for each autocorrelation coefficient (ρ) from the Godfrey test, what is basically the estimate of the parameter p, values of the testing criterion DW (relation 9). This parameter p is usually in the practice smaller than 1. The hypothesis about the negative autocorrelation of residues of the first sequence (p_1 = – 0.23445) was adopted on the grounds of assessment of DW test outputs, meaning that only one autocorrelation coefficient, which was suitable for the DW test, was incorporated into the model.

The auto-regression equation of the first series for each ε_t in the model can be expressed as:

$$\varepsilon_t = p \, . \, \varepsilon_{t-1} \tag{8}$$

Where: ε_{t-1} – previous value of ε_t,

p – parameter expressing the degree of autocorrelation of residues (p_1 = – 0.23445).

$$DW = \frac{\sum_{t=2}^{n} (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=2}^{n} \varepsilon_t^{\,2}} \tag{9}$$

The following table shows outputs from the software program SAS for the elaboration of regression model including important indicators, which were already mentioned in the previous sub-chapter, including the respective coefficient of autocorrelation (ρ) (see 2 line of column Values).

Tab. 4. Output from the software program SAS for SF with random component

Model parameters	Values	Error middle value	T	Prob>|T|
Limits	94069	3389	27,7553	0.0001
Autoregresion 1	-0.23445	0.1275	-1.8389	0.051
Par 2	11272	4742	2.3769	0.0208
Par 3	4579	4238	1.0804	0.2844
Par 4	21111	4360	4.8421	0.0001
Par 5	15075	4328	3.4828	0.001
Par 6	18172	4333	4.1937	0.0001
Par 7	12991	4329	3.001	0.004
Par 8	-58864	4329	-13.5963	0.0001
Par 9	-56921	4320	-13.1768	0.0001
Par 10	16903	4353	3.8833	0.0003
Par 11	27378	4204	6.5124	0.0001
Par 12	20556	4801	4.2814	0.0001
Linear trend	-232.02204	34.2327	-6.7778	0.0001
R^2	0.95			

Source: Table elaborated by authors by means of software program [4]

We elaborated and verified the model on the grounds of calculated parameters similarly as the model in the sub-chapter 3, only it takes into account in addition to the trend and seasonal component also the random (residual) component ε_t (see relation 8):

$$\hat{y}_{SF} = 94\,069 - 232{,}02204.\,t + 11\,272x_{2t} + 4\,579x_{3t} + 21\,111x_{4t} + 15\,075x_{5t} + 18\,172x_{6t} + 12\,991x_{7t} - 58\,864x_{8t} - 56\,921x_{9t} + 16\,903x_{10t} + 27\,378x_{11t} + 20\,556x_{12t} + \varepsilon_t \tag{10}$$

The figure 4 on the following page shows the course of curves of real (blue curve) and modelled values of time sequences of the number of transported passengers with the addition of the random component.

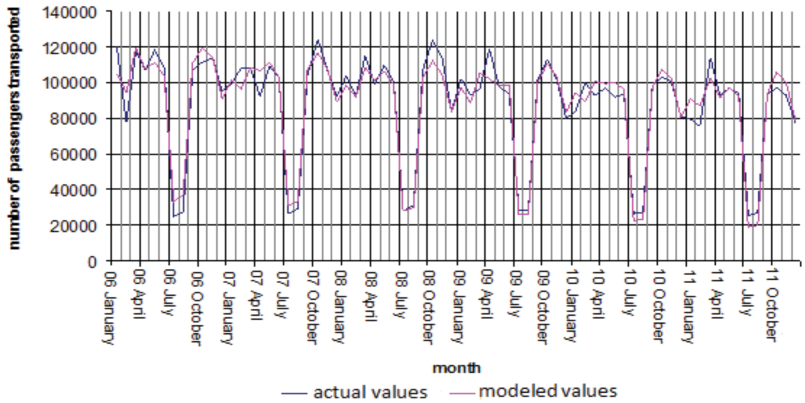

Fig. 4. Real and modelled number of passengers for SF with random component

Source: Data provided by the companies concerned and the calculations of the model 10

Factorial extrapolation of the created model for student fare with addition of the random component

Extrapolation in form of point prognosis and interval estimate for 2012 was made after the verification of the model. With regard to analogical procedure from the previous two sub-chapters, only empirical outputs of values are specified below.

$$\hat{y}_{SF} = 94\,069 - 232{,}02204.\,73 + 11\,272\,.\,1 + 4\,579\,.\,0 + 21\,111\,.\,0 + 15\,075\,.\,0 + 18\,172\,.\,0 + 12\,991\,.\,0 - 58\,864\,.\,0 - 56\,921\,.\,0 + 16\,903\,.\,0 + 27\,378\,.\,0 + 20\,556\,.\,0 + 327{,}52665 = 88\,440 \; \textit{persons} \tag{11}$$

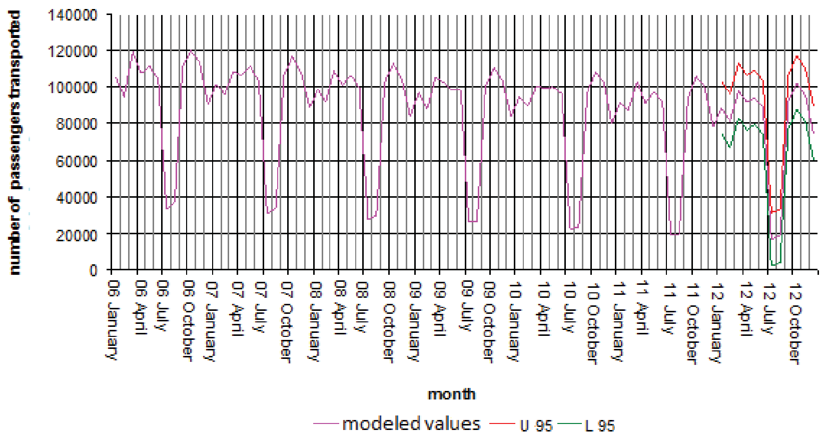

Fig. 5. Point and interval prognosis with added random component for SF for 2012

Source: Prepared on the basis of calculations of authors and software [4]

3. EVALUATION OF THE PROPOSED SOLUTIONS

The indicator p – value (p – value) was used to evaluate statistical significance of individual parameters and coefficients in the model. Values of this indicator enabled us to observe changes of the statistical significance of parameters (b_0, $b_1 ... b_{12}$), as well as coefficients of autocorrelation (p), with and without the random (residual) component. As can be seen on the outputs from the software program SAS (tab. 1, 4), values p – value are represented by columns Prob › |T|. Parameters complying with the value p – value ‹ α can be therefore on the grounds of the level of significance α = 0,05 chosen by us considered to be statistically significant.

With regard to student fare, it can be said on the grounds of results (column Prob › |T|) that all parameters except for the parameter b_3 = 0.3627 (par 3, see tab.1) can be considered statistically significant at the respectively chosen level of significance. This fact was influenced by adding the non-systemic part to the model. The addition of the random (residual) component succeeded in, even if not completely removing, then at least reduce the statistical insignificance of the parameter b_3 to more acceptable level b_3 = 0.2844 (see tab. 4). The statistical significance of the autocorrelation coefficient p_1 = 0,051 is in this case still at the acceptable level. Usage of this combined model (see relation 10) succeeded in the increase of the level of explanation of the variability of values of time sequence of the prognosticated indicator from R^2 = 0.94% to R^2 = 0.95%. Conclusions of testing for the full fare were similar to student fare.

It can be said also on the grounds of these facts that usage of models combining the systemic and non-systemic part with regard to student fare (relation 10) and full fare is the most suitable solution. Addition of the random (residual) component in models proved justified as it increased the statistical significance of individual parameters as well as whole models and given prognoses has thus become more relevant (see tab. 5).

Tab. 5. Prognosis of the model for SF for 2011 with addition of the random component

Month	Composed of prognosis	U 95	L 95
January	88440	102739	74142
February	81469	96156	66783
March	97780	112488	83073
April	91510	106218	76801
May	94375	109083	79666
June	88962	103671	74254
July	16875	31584	2167
August	18587	33295	3878
September	92178	106886	77469
October	102421	117129	87713
November	95368	110076	80659
December	74579	89288	59871

Source: Table elaborated by authors on the grounds of model 11 and software program [4]

To quantify the reliability of said prognoses for both the student and full fare it is possible to perform previously mentioned analysis ex – post. Both real and estimated values for 2011 for student fare which served as the basis are summarised in the tab. 6.

With regard to student fare, the deviation of prognosticated values from real ones for 2011 was only around 1.22% out of the total number of transported passengers, on the grounds of which it can be concluded that the respective combined model 10 is relatively accurate. Real values of transported passengers provided by the transporter show for years 2006 – 2011 with regard to SF an average year-on-year drop by 2.2%.

Tab. 6. Values for 2011 for SF (persons/month)

Month	Actually values	Prognosis
January	79647	90919
February	75750	86968
March	113982	102558
April	92886	91148
May	97007	97489
June	93832	91782
July	25636	19171
August	26923	19970
September	92800	93660
October	97466	105712
November	92859	99966
December	77207	78604

Source: Elaborated by authors on the grounds of outputs from the software program [4]

High seasonability of the previous demand of transported people (see Fig. 2) was confirmed especially for the student fare. It is characterized by repeated proportionality of transported students in the course of the year.

In relation to year-on-year drop of demand for urban bus transport, determined on the grounds of acquired data about the given indicator, the year-on-year drop with regard to SF by 5.54% can be stated also on the grounds of comparison of 2011 with the prognosticated year 2012 (see tab. 7).

Tab. 7. Comparison of transported persons in 2011 and 2012 with regard to SF (persons/month)

Month	2011	2012	Difference (%)
January	79647	88440	11.04
February	75750	81469	7.55
March	113982	97780	-14.21
April	92886	91510	-1.48
May	97007	94375	-2.71
June	93832	88962	-5.19
July	25636	16875	-34.17
August	26923	18587	-30.96
September	92800	92178	-0.67
October	97466	102421	5.08
November	92859	95368	2.70
December	77207	74579	-3.40
Σ	965995	942544	-5.54

Source: Elaborated by authors

CONCLUSION AND RECOMMENDATIONS FOR PRACTICE

When drafting a prognosis based on the theory of time sequence, it is very important to have sufficiently long time sequence of values (at least 5 years of previous development of the indicator in question). However, it should be remembered that the information ability of the prognosis falls with the growing number of prognosticated periods.

When using software program for creating a short-term prognosis or prognosis model, it is convenient to verify accuracy of the created model by prognosticating known values that are available (prognoses ex – post). And then subsequently perform prediction of the unknown future value. In practice it is good if the transport company, which needs very accurate prognosis of demand of passengers for a longer period of time, uses for the creation of prognosis professional statistical software and experienced analysts, because analysts can complete the prognosis created on the grounds of mathematical – statistical methods also with their own qualified estimate.

The contribution dealing with the determination of future demand of passengers for urban bus transport is based on methods based on time sequences. The subject of prognoses was time sequences of transported passengers travelling for student and full fare, where empirical outputs for the following period were obtained. Such method for prediction of data about the number of passengers can help the respective company to correctly plan its technical as well as staffing capacities in the future, or also pre-define its price or investment policy.

REFERENCES

1. Cyprich O., Liščák Š., *Modelovanie reziduálnej zložky časového radu a jej vplyv na kvalitu viackriteriálneho regresného ukazovateľa počtu prepravených cestujúcich*. Doprava a spoje 2010, roč. 6, č. 1, pp. 16-24, ISSN 1336-7676.
2. Droździel P., Komsta H., Krzywonos L., *Analiza intensywności użytkowania pojazdów (Część I)*. Logistyka 2012, nr 3, pp. 487-492, ISSN 1231-5478.
3. Dolinayová A., *Možnosti prognózovania vývoja osobnej prepravy v meniacich sa podmiankach*. Perner's Contacts 2009, roč. 4, č. 1, pp. 42-49, ISSN 1801-674X.
4. *SAS 9.1.3* [program on CD], Cary, NC: SAS Institute Inc., 2003, last update, 22 December 2008.
5. Trevin D.A., *Guide to Interpreting Time Series – Monitoring Trends*. ABS, Canberra 2003, pp. 148, ISBN 0-642-54286-4.

ANALIZA WPŁYWU SKŁADOWEJ REZYDUALNEJ PRZY PROGNOZOWANIU ZAPOTRZEBOWANIA PASAŻERÓW ZA POMOCĄ JEDNOWYMIAROWYCH SEKWENCJI CZASOWYCH TYPU ADDYTYWNEGO

Abstract

Niniejszy artykuł przedstawia matematyczno-statystyczne modelowanie zapotrzebowania pasażerów na podmiejski transport autobusowy. Analiza jest oparta na sekwencjach czasowych typu addytywnego ze składową trendu liniowego w miesięcznym przedziale czasowym. W pracy stworzono krótkoterminową prognozę dla wybranego środka transportu w celu uzasadnienia dodania losowej (rezydualnej) składowej do modelu dla potrzeb tworzenia prognozy. Dla celów praktycznych, autorzy wybrali konkretnego dostawcę usług transportu zbiorowego, który zapewnił dane niezbędne do przeprowadzenia analizy.

Authors:

Prof. Eng. **Štefan Liščák,** PhD – University of Žilina, The Faculty of Operation and Economics of Transport and Communications, Department of Road and Urban Transport, stefan.liscak@fpedas.uniza.sk

Eng. **Lukáš Holeša** – University of Žilina, The Faculty of Operation and Economics of Transport and Communications, Department of Road and Urban Transport, Lukas.Holesa@fpedas.uniza.sk

M.Sc. Eng. **Jacek Caban** – Lublin University of Technology, Faculty of Mechanical Engineering, Institute of Transport, Combustion Engines and Ecology, j.caban@pollub.pl