ONDREJ CYPRICH
E-mail: ocyprich@kia.sk
VLADIMÍR KONEČNÝ
E-mail: vladimir.konecny@fpedas.uniza.sk
KATARÍNA KILIANOVÁ
E-mail: katarina.kilianova@fpedas.uniza.sk
University of Žilina,
Faculty of Operation and Economics
of Transport and Communications
Univerzitná 1, SK-010 26 Žilina, Slovak Republic

Traffic Planning
Preliminary Communication
Accepted: Aug. 13, 2012
Approved: Sep. 24, 2013

SHORT-TERM PASSENGER DEMAND FORECASTING USING UNIVARIATE TIME SERIES THEORY

ABSTRACT

The purpose of the paper is to identify and analyse the forecasting performance of the model of passenger demand for suburban bus transport time series, which satisfies the statistical significance of its parameters and randomness of its residuals. Box-Jenkins, exponential smoothing and multiple linear regression models are used in order to design a more accurate and reliable model compared the ones used nowadays. Forecasting accuracy of the models is evaluated by comparative analysis of the calculated mean absolute percent errors of different approaches to forecasting. In accordance with the main goal of the paper was identified the ARIMA model, which fulfils almost all statistical criterions with an exception of the model residuals normality. In spite of the limitation, the best forecasting abilities of identified model have been proven in comparison with other approaches to forecasting in the paper. The published findings of research will have positive influence on increasing the forecasting accuracy in the process of passenger demand forecasting.

KEY WORDS

passenger demand, demand modelling, short-term demand forecasting, suburb bus transport

1. INTRODUCTION

The Slovak operators of suburban bus transport services are entitled to claim compensation connected with public service contract and reasonable profit by the authority. These services are economically inefficient from the aspect of operator of public passenger transport services. The Regulation (EC) No 1370/2007 of the European Parliament and of the Council of 23 October 2007 on public passenger transport services by rail and by road has been valid in Slovakia since December 2009. The Regulation determines the commercial principle. This principle means that the price should not be calculated as percentage of costs but rather the price must contain risk. The risk, stemming from the operation of transport services can be carried by the operator or by the authority according to contract requirements. One of possibilities of how to reduce the risk in the process of price determination, is the passenger demand forecasting by using univariate time series theory. Because of this fact, forecasting of passenger demand has become one of the most important methods in case of risk reduction and pricing from the point of contractual sides in the sphere of public passenger transport services.

Methods [1, 2], which are used for the purpose of passenger demand forecasting by the Slovak transportation companies at the present time, are considerably simplified, and what is more, they are neither considered to be accurate [3]. These limitations might be caused by insufficient research in this area over the last years. Statistical modelling and forecasting of the passenger demand by using univariate time series theory is probably one of the most common forecasting methods used for work with economic time series data. This theory is successfully applied in the sphere of freight modelling, i.e. of port services [4], intermodal (railway) transport services [5] and passenger demand modelling, i.e. of maritime transport [6], public transport [7]. Other very significant research activities focused on statistical modelling in the sphere of transportation are analysed in Karlaftis's [8] comprehensive survey. In the paper we are filling up the basis of established research focused on forecasting of periodic time series of passenger demand for suburban bus transport [9]. The main goal of this paper is to introduce a method of the statistical modelling of passenger demand (carried per full fares) by using univariate time series theory which appears to be more accurate and reliable alternative to automated or semi-automated forecasting methods [3]. The secondary aim of this paper is the evaluation of forecasting performance of chosen approaches to passenger demand forecasting. Three different methodologies - exponential smoothing, multiple linear regression and Box-Jenkins are used in order to identify more accurate and reliable statistical model compared with the nowadays used ones in the paper. The randomness of residuals of identified models is tested by using several linear independence, stationarity and normality tests. The forecasting accuracy of identified manually designed model is analyzed by comparative analysis of mean absolute percent errors of five different approaches to passenger demand forecasting. Autoregressive integrated moving average (ARIMA(0,1,0)(0,1,1)$_{12}$ with no intercept parameter) model of logarithmically transformed passenger demand time series was chosen for further forecasting and analytic purposes. The model with an exception of normality of its standardised residuals (ε_t) satisfies the statistical significance of model parameters and randomness of model standardised residuals.

An indicator, number of passengers carried per full fare is considered as an indicator of passenger demand in the paper. In terms of the Slovak transportation companies reporting system, the category of passenger carried per full fare consists of persons without any entitlement to reduced fares. In accordance with the main goal of the paper a statistical model which is suitable for short-term (forecast horizon $h \leq 1$ year) forecasting of passengers (carried per full fare) demand for suburban bus transport in Zilina self-governing region was designed. Most of the analyses, modelling and forecasting procedures of the time series mentioned in this paper were worked out by using SAS (Statistical Analysis System) LE 4.1 [10] and SAS 9.3.1 [11] software.

2. DATA AND METHODS

2.1 Properties and adjustments of input data

Input data of calculations presented in this paper were counts of passengers carried per full fares collected by the cooperating carrier. These values were aggregated by summing so that the output of the aggregation process was monthly time series of passenger demand carried per full fares $\{Q_p(t); 1 \leq t \leq 96\}$ (for period of months 1/2000-12/2007) in the Zilina region.

The values of the $Q_p(t)$ time series designed in such a manner were considered to be spatially and substantially homogeneous as the carrier had changed neither the geographic scope nor the transportation technology in the range affecting substantial and spatial aspects of the analysed time series within the specified period of months. "Trading day effects" were eliminated only in case of X-11-ARIMA and X-12-ARIMA by calendar adjustment procedures of these methods. The output of the calendar adjustment process was a fully homogeneous time series of passenger (carried per full fare) demand for suburban bus transport $\{Q(t); 1 \leq t \leq 96\}$.

At first the subjective methods identified and later the objective methods properly confirmed the downward linear trend, monthly multiplicative seasonality of $Q(t)$ time series in pre-forecasting analyses [12]. The analysed $Q(t)$ time series was logarithmically transformed prior to the modelling process for the purpose of eliminating its multiplicative properties. The models presented in this paper fully respect these properties.

2.2 Methods

2.2.1 Statistical modelling

Multiple regression, exponential smoothing and autoregressive models were used for statistical modelling of $Q(t)$ time series. The practices and principles of linear stochastic models designing [13, 14] were used in the process of developing and fitting of $Q(t)$ time series model by using the Box-Jenkins methodology. The moving average model (1) of seasonal time series ARIMA(0,1,0)(0,1,1)$_{12}$ (method A) was designed by the application of this methodology.

$$(1-B)Q_t = \Theta(B^s)\varepsilon_t \tag{1}$$

where:

- B – *the backshift operator, that is,* $BQ_t = Q_{t-1}$,
- $\Theta(B^s)$ – *the seasonal moving-average operator, represented as the polynomial in the backshift operator:* $\Theta(B^s) = 1 - \Theta_1 B^{12}$,
- ε_t – independent disturbance (random error) at time *t*.

Holt-Winters multiplicative (method B) model (2) was developed and fitted by using exponential smoothing methodology.

$$Q_t = (\mu_t + \beta_t \cdot t) \cdot s_p(t) + I_t \tag{2}$$

where μ_t represents time varying mean term estimated by smoothing level L_t (3).

$$L_t = \alpha(Q_t/S_{t-p}) + (1-\alpha)(L_{t-1} + T_{t-1}) \tag{3}$$

Time varying slope β_t is estimated by smoothed trend T_t equation (4).

$$T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)T_{t-1} \tag{4}$$

Time varying seasonal contribution for one of the p seasons $S_p(t)$ is estimated by seasonal factors S_t (5).

$$S_t = \delta(Q_t/L_t) + (1-\delta)S_{t-p} \tag{5}$$

The smoothing state at time $t = 0$ of the designed exponential smoothing model was obtained by Chatfield's backcasting method [15]. The smoothing weights (level α, trend γ, seasonal δ) were determined so as to minimize the sum of squared one-step-ahead prediction errors.

Simple variants of multiple regression models, namely, regression models of basic trend functions with the seasonal dummy variables (6), do not provide satisfactory results in terms of the statistical properties and therefore they were used in combinations with linear stochastic models; the general equation of the model can be seen in formula (7).

$$Q_t = \beta_0 + \beta_1 \cdot t + \sum_{t=1}^{11} \beta_j \cdot D_{jt} + \varepsilon_t \tag{6}$$

where β_0 is intercept parameter, β_1 is slope parameter, β_j are seasonal disturbance parameters and D_{jt} are seasonal dummies and ε_t is random error term. Each element of the seasonal dummy regressor is either one or zero, that means $D_{jt} = 1$ when t is in the j-th month and zero otherwise.

$$(1-B)^d \cdot (1-B^s)^D \cdot Q_t =$$
$$= \beta_0 + \beta_1 \cdot t + \sum_{i=1}^{11} \beta_j \cdot D_{jt} + \frac{\theta(B) \cdot \Theta(B^s)}{\phi(B) \cdot \Phi(B^s)} \cdot \varepsilon_t \tag{7}$$

- B^s – the seasonal backshift operator, that is, $B^s Q_t = Q_{t-s}$,
- $\theta(B)$ – the moving-average operator, in research represented as the polynomial in the backshift operator: $\theta(B) = 1 - \theta_1 B$,
- $\phi(B)$ – the autoregressive operator, in research represented as the polynomial in the backshift operator: $\phi(B) = 1 - \phi_1 B$,
- $\Phi(B^s)$ – the seasonal autoregressive operator, in research represented as the polynomial in the backshift operator: $\Phi(B) = 1 - \Phi_1 B^{12}$.

The models of linear trend with seasonal dummies combined with autoregressive moving average processes (method A - AR(1), method D – SARMA(1,0,0)$(1,0,0)_{12}$ with intercept and method E - SARMA(1,0,0)$(0,0,1)_{12}$ with intercept) were designed by applying the theory of combined linear regression and stochastic models.

Method F was a combination of linear trend model with autoregressive integrated moving average process of seasonal time series ARIMA(0,1,0)$(0,1,1)_{12}$. Logarithmically transformed $Q(t)$ time series in the statistical modelling of almost all processes were used. The only exception was method B where non-transformed data were applied.

2.2.2 Forecasts evaluation

The forecasting accuracy was evaluated by the comparison of the mean absolute percent errors ($MAPE_h$, formula 8) calculated from multi-step ahead forecasts of five different approaches to passenger demand forecasting, namely:

- manually designed model (ARIMA(0,1,0)$(0,1,1)_{12}$ of logarithmically transformed time series),
- multiplicative variant of X-11-ARIMA/88 [16] - namely ARIMA(0,1,2)$(0,1,1)_{12}$ with TDRegr model of trading day effects of logarithmically transformed time series,
- multiplicative variant of X-12-ARIMA [17,18] - namely ARIMA(1,0,1)$(0,1,1)_{12}$ + regARIMA model of trading day effects and additive outliers,
- semi-automatic forecasting - Holt-Winters additive model of logarithmically transformed data,
- semi-automatic forecasting using holdout samples [19] - Holt-Winters multiplicative model. (Semiautomatic forecasting using holdout samples is a forecasting method where candidate models are primarily fit to the data with the holdout sample excluded (period of months 1/2007-12/2007 in the paper). After model fitting, the one-step-ahead forecasts are made in the fit region (in-sample) or the multi-step-ahead forecasts are made in the holdout sample region (out-of-sample). The best forecasting model is selected from candidate models according to value of root mean square error (RMSE) calculated from multi-step-ahead forecasts and actual (values from holdout sample). The chosen model is then refitted to full range passenger demand data.

The best fitting models of both mentioned approaches to semi-automatic forecasting were chosen from enlarged model list according to RMSE (Root Mean Square Error) as a model selection criterion. It should be noted that the default model list of SAS 9.1.3 TSFS (Time Series Forecasting System) consists of 42 models and due to its identified insufficiency was enlarged with respect to the analysed time series statistical properties to a size of 242 exponential smoothing, regression and Box-Jenkins models.

$$MAPE_h = \frac{1}{h} \cdot \sum_{T=n+1}^{n+h} \left| Q_T - \hat{Q}_T \right| / Q_T \cdot 100\,[\%] \tag{8}$$

where:

- Q_T – actual value of passenger demand time series at time $T = n + h$; $h = 1, 2, \ldots, 12$,
- $\hat{Q}_T$ – estimated value of passenger demand at time $T = n + h$; $h = 1, 2, \ldots, 12$,
- h – forecast horizon, $h = 1, 2, \ldots, 12$,
- n – number of observations of $Q_p(t)$ time series used for model estimation, $n = 96$.

2.2.3 Testing of models

The manually designed statistical models presented in the paper were tested for compliance with the requirements imposed on mutual linear independence, stationarity and the normality of probability distribution of their standardized residuals (ε_t = 1,...,96). Mutual linear independence of ε_t was tested by Bartlett's test for autocorrelation [20] and Ljung-Box's χ^2 statistics [21]. Stationarity of the residual components was evaluated by augmented Dickey-Fuller's tests (ADF tests) [22] and Dickey-Fuller's unit root tests of seasonal time series (SDF tests) [23]. Normality of the standardized residuals probability distribution was in order to eliminate influence of weaknesses in process of normality testing and with respect to advice in literature [24] tested by the most commonly used normality tests, namely by Shapiro-Wilk's (SW) [24] and by Filiben's [20] described Kolmogorov-Smirnov's (K-S), Anderson-Darling's (A-D) and Cramér von Mises's (C-M) tests. The K-S test is distribution-free in the sense that the critical values do not depend on the specific distribution being tested. A-D test is a modification of the K-S test and gives more weight to the tails of the distribution than does the K-S and C-M test. Finally, the S-W test has done very well in comparison studies with other goodness-of-fit tests [20]. The statistical significance of the estimated parameters of the models was tested by the Student's t-test [25]. These tests were conducted at significance level α = 0.05 (except for the normality tests where significance levels α = 0.15 were used because of test detection abilities).

3. EMPIRICAL RESULTS

The outputs of the forecasting procedures (analyses, modelling, testing) presented in the paper are goodness-of-fit statistics (*Table 1*), outputs of the randomness tests as well as evaluation of statistical significance of estimated parameters of models. According to high volume of available outputs of computations they are presented only in considerably reduced form in the paper. Full outputs including estimates of model parameters and its statistical significance evaluation, goodness-of-fit statistics, point and interval forecasts are part of the dissertation thesis [12].

Six widely used in-sample and one extrapolational goodness-of-fit statistics were used in order to measure how well different models (methods A-F) fit the data and to eliminate the known difficulties concerning further comparative analyses. Traditional statistics used in the paper were non unit free and sensitive for outliers root mean square error (RMSE) and unit free good for comparisons mean absolute percent error – MAPE and mean percent error – MPE, which gave us information concerning underestimation or overestimation of the reality by models. Models with different number of parameters were used in the paper. Penalty goodness-of-fit statistics might be useful for comparative analyses of the models. Therefore, we computed Adjusted R^2 – AdjR^2, Akaike's information criterion – AIC and a large sample-sensitive Schwarz Bayesian information criterion – SBIC. In order to evaluate the extrapolational accuracy of the model was calculated by $MAPE_{12}$, which provides information about average absolute percent error in 12-month forecasting period (calculated from the forecast and real data). The computed values of these measures are contained in *Table 1*.

For further forecasting purposes, based on the forecasting accuracy as well as the results of tests for ε_t randomness, the models estimated by methods A and F were chosen. The choice was supported by the fact that according to the results of tests ε_t of both models are mutually independent and stationary, and what is more, the estimated parameters of the models are almost statistically significant (statistical insignificance of one of the parameters estimated by method F should not be the difficulty for further forecasting processes [26]). After considering the forecasting accuracy (*Table 1* row $MAPE_{12}$) and the statistical significance of its parameters (*Table 2* column statistical significance) we can say that model estimated by method A is more suitable for forecasting in comparison with one estimated by method F. Model (2) estimated by method A does not show very well fitting ability for actual data by its forecasts compared with other ones,

Table 1 - Computed values of the goodness-of-fit statistics

Statistics	Unit	Method					
		A	B	C	D	E	F
RMSE	[thous. of pass.]	42.91	39.51	34.97	35.06	35.11	40.38
MAPE	[%]	1.39	1.61	1.46	1.45	1.45	1.59
Adj R_2	[-]	0.98	0.99	0.99	0.99	0.99	0.98
AIC	[-]	626.01	711.88	710.44	712.97	713.23	629.32
SBIC	[-]	628.43	719.57	746.34	751.43	751.7	639.05
MPE	[%]	-0.14	-0.07	-0.04	-0.04	-0.04	-0.05
$MAPE_{12}$	[%]	5.08	6.64	5.89	5.72	5.67	5.14

Note: RMSE, MAPE, AdjR^2, AIC, SBIC and MPE were computed by using the actual and forecast values of observations in the period of evaluation (for a period of months 1/2000 – 12/2007), parameters of the models used for forecasting were estimated by applying observations from the same period of time. $MAPE_{12}$ were computed by using the actual and forecast values of observations in the period of evaluation (for a period of months 1/2008 -12/2008 – $MAPE_{12}$), parameters of the model used for forecasting were estimated by applying observations for a period of months 1/2000 – 12/2007 – $MAPE_{12}$.

Table 2 - Evaluation of tests for randomness of ε_t and statistical significance of estimated parameters of models

Method	Statistical significance	Linear independence			Stationarity		Normality			
		BT-ACF	BT-PACF	LB	ADF	SDF	S-W	K-S	C-M	A-D
A	0/1	+	+	+	+	+	-	-	-	-
B	0/3	+	+	+/-	+	+	-	-	-	-
C	2/14	+	+	+	+	+	-	-	-	-
D	3/14	+	+	+	+	+	-	-	-	-
E	3/14	+	+	+	+	+	-	-	-	-
F	1/4	+	+	+	+	+	-	-	-	-

Note: Statistical tests provided "+"- satisfactory "-"- unsatisfactory "+/-"-boundary (satisfactory) results.

Table 3 - Estimates of SMA model parameter and output of its statistical significance test

Model parameter	Estimate	Standard error	t-test criterion	p-value
SMA(1) - Θ_1	0.4829	0.1420	3.4002	0.0010

but it is the only model which satisfies all statistical criteria from the set of developed models. The estimated value of its parameter with standard error and output of parameter statistical significance test is presented in *Table 3*.

Further use of values estimated by method A requires consideration of the fact that model (method A) systematically overestimates the reality. This accrues from the value of mean percentage error (MPE = -0.137%) of this model. True values $Q(t)$; $t = n+1,\dots,n+h$ are likely lower than by model estimated values.

Graphical output of modelling and forecasting by using method A is seen in Figures 1 and 2. The estimated values are expressed by smooth curve and empirical values by black points in *Figure 1*. The graphical interpretation of the actual (empirical) and forecast values shows that this model accurately describes the variability of empirical values of $Q(t)$. This fact is also supported by low levels of residuals (displayed by the bar diagram in *Figure 1*). There are confidence intervals drawn in *Figure 1* and in more detail in *Figure 2*.

Note: In order to protect the interests of cooperating bus transport company the values in Figures 1 and 2 are presented in the form of simple index numbers ($I_{n/0}$). Reference value of the variable Q(t) is expressed in the base 1.0 in reference situation (t = 1, January 2000).

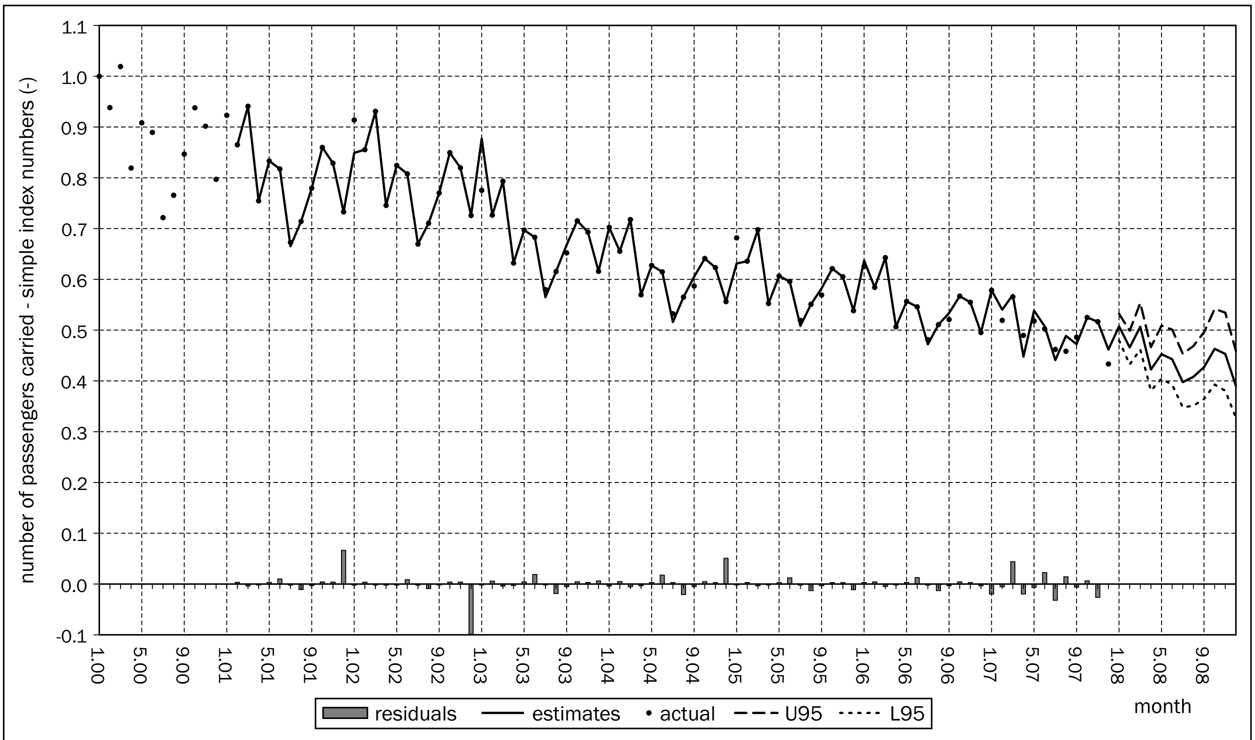

Figure 1 - Actual and estimated values of Q(t) time series

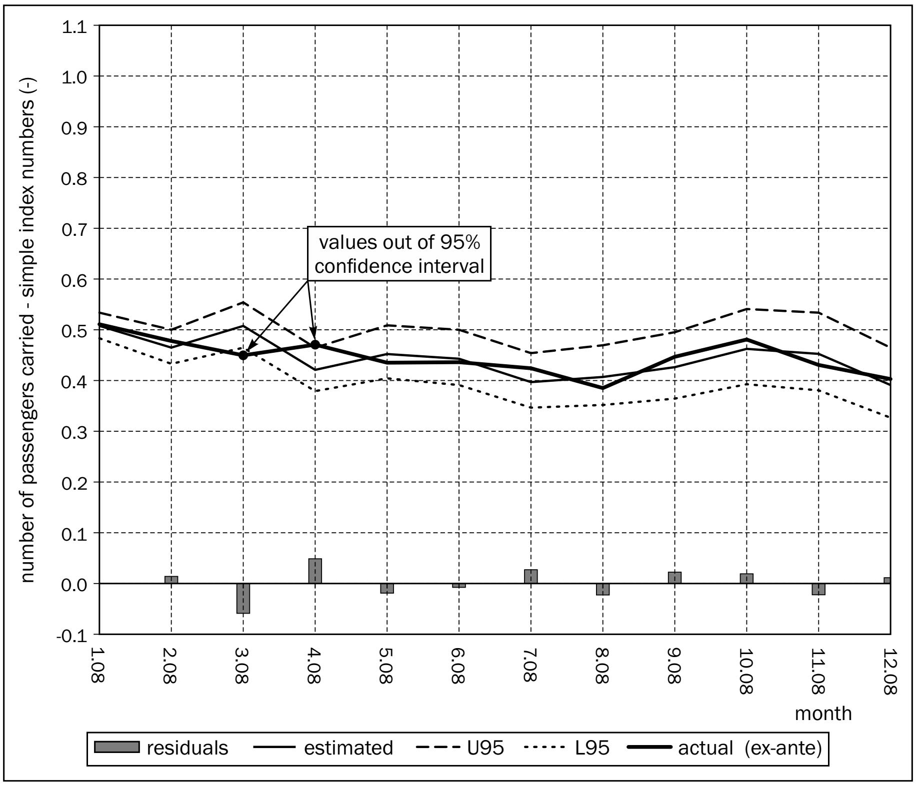

Figure 2 - Ex-ante evaluation of forecasts of Q(t) time series

4. DISCUSSION

Six models were designed by manual approach, among which ARIMA(0,1,0)(0,1,1)$_{12}$ model (with no intercept parameter) of logarithmically transformed $Q(t)$ was chosen for further forecasting purposes. This model with an exception of normality of its standardised residuals (ε_t) satisfies the statistical significance of model parameters and randomness of its ε_t. The non-normality of ε_t of the model estimated by method A confirms the results of goodness-of-fit tests (*Table 2* section normality), although it is evident from the histogram and probability plot (*Figure 3*) of model ε_t. However, it is necessary to note that the normality is typical for neither manually designed models (A-F) nor alternative approaches (X-11-ARIMA, X-12-ARIMA, semi-automatic procedures) to passenger demand time series modelling. These findings support Arlt's [14] theory, who says that non-normality of ε_t probability distribution is a common and problematically eliminable phenomenon when linear stochastic models are used for modelling of economic time series. The problem of non-normality in case of identified model caused reducing of 95% confidence interval width; therefore, the real interval will be probably wider.

According to the calculated values of MAPE$_{12}$ the chosen forecasting model despite the identified limitation gives the most accurate multi-step ahead fore-

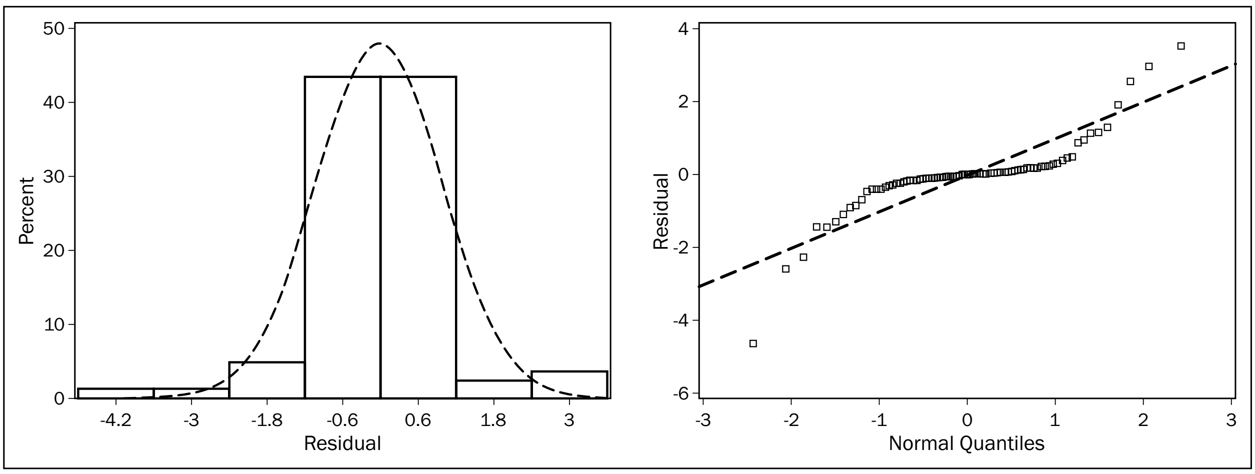

Figure 3 - Distribution analysis of standardised residuals of log ARIMA(0,1,0)(0,1,1)$_{12}$ model

Table 4 - Comparison of forecasting accuracy of chosen forecasting methods

Statistic \ Method	Manual estimation			Semi-automatic forecasting	
	TSFS	X-11-ARIMA	X-12-ARIMA	TSFS	TSFS holdout
$MAPE_{12}$ [%]	5.0830	5.1469	7.8637	5.5280	5.4229

casts of $Q(t)$ in comparison with other manually designed models (*Table 1* row $MAPE_{12}$) and alternative approaches to passenger demand forecasting (*Table 4* row $MAPE_{12}$). An interesting finding is the fact, that the chosen manually designed model (method A) has the worst fitting properties compared with other manually designed models (*Table 1* row RMSE), and what is more, it does not influence the model ability to generate accurate forecasts. On the other hand, model with the best values of RMSE (*Table 1* method C) gives the least accurate multi-step ahead ex-ante forecasts of $Q(t)$ time series. These findings support the hypothesis that very good fitting characteristics (goodness-of-fit statistics) of models do not guarantee its very good forecasting accuracy as well.

The analysis of models MAPEs (*Table 4*) also refers to comparatively accurate forecasting accuracy of X-11-ARIMA procedure ($MAPE_{12}$ = 5.1469%) and nearly 0.5% worse predictive accuracy of semi-automated approaches to passenger demand forecasting. The analysis also identified poor forecasting ability of X-12-ARIMA procedure. Poor forecasting accuracy of the procedure probably results from poor statistical properties of trading day variations (regARIMA) model of analysed time series. There was proved statistical insignificance of regARIMA model (F-test p-value = 0.074) and 6 out of 7 regARIMA model parameters. Similar findings result from outputs of X-11-ARIMA procedure which provided more accurate forecasting outputs without trading day regression (TDRegr) model of trading day effects compared to one with TDRegr model. Taking into account the proved non-uniformity of weekly seasonal pattern of passenger demand [27] the problems with TDRegr and regARIMA models clearly indicate ineffective modelling of trading day effects. The mentioned problem of trading day effects modelling has not been solved even by the latest variant of class X-11 methods, X-12-ARIMA. It has been found, that X-12-ARIMA models have better statistical properties in comparison with X-11-ARIMA, but the problem of statistical significance of its parameters is not solved [3].

According to outputs of analyses presented in the paper it is clear that manual approach to passenger demand time series forecasting by using univariate time series theory provides the most accurate forecasting results and also improves the statistical properties of forecasting models. On the other hand it should be obvious that this approach is not suitable when analysts do forecasts of a large number of passenger demand time series. In such cases the analysts should apply semi-automatic forecasting methods for low-valued forecasts and manual approach for high-valued or problematic time series forecasting.

5. CONCLUSION

Outputs of the statistical tests of standardized residuals randomness of the model and computed values of goodness-of-fit statistics proved that ARIMA(0,1,0)(0,1,1)$_{12}$ model (with no intercept parameter) of logarithmically transformed $Q(t)$ fulfils all requirements for the statistical significance of its parameters, and what is more, mutual linear independence and stationarity of its residuals. On the other hand this model does not satisfy the requirements for normality of probability distribution of its standardized residuals. The elimination of this phenomenon by using the Box-Cox's transformations, non-linear stochastic models or extension of original multiple regression model by GARCH model of its residuals will be the object of our further research. In case of class X-11 decomposition methods it is necessary to focus further research activities onto analysis of modelling with predefined parameters of trading day regression models. The outcomes of this research could bring further improvement of forecasting accuracy, reliability and statistical properties of designed models too. The implementation of the planned research requires daily time series of passenger demand, but neither transportation companies nor other stakeholders have provided such data up to date.

The ARIMA(0,1,0)(0,1,1)$_{12}$ model (with no intercept parameter) of logarithmically transformed passenger demand time series presented in this paper despite the abovementioned restriction represents more reliable and more accurate passenger demand forecasting method in comparison with up to this time used ones and what is more, the alternative methods such as X-11-ARIMA, X-12-ARIMA and semi-automatic forecasting, too. The attendant phenomenon of applying the model described in this paper in a relevant transport company management is the reduction of manager's decisions uncertainty, and, moreover, it can result in increase of company revenues. This model with respect to cross-regional differences should not be considered as universally applicable in all regions or counties without further analysis of the properties of relevant time series. The model applicability was successfully proved in seven other regions of the Slovak Republic and it is now used in forecasting studies of the Slovak self-governing regions and should be applicable in case of time series with linear trend, monthly multiplicative seasonality.

ACKNOWLEDGEMENT

This contribution/publication is the result of the implementation of the project:

Centre of Excellence for Systems and Services of Intelligent Transport II.,

ITMS 26220120050 supported by the Research & Development Operational Programme funded by the ERDF.

"Podporujeme výskumné aktivity na Slovensku/ Projekt je spolufinancovaný zo zdrojov EÚ"

ONDREJ CYPRICH
E-mail: ocyprich@kia.sk
VLADIMÍR KONEČNÝ
E-mail: vladimir.konecny@fpedas.uniza.sk
KATARÍNA KILIANOVÁ
E-mail: katarina.kilianova@fpedas.uniza.sk
Žilinská univerzita v Žiline,
Fakulta prevádzky a ekonomiky dopravy a spojov
Univerzitná 1, 010 26 Žilina, Slovenská republika

ABSTRAKT

KRÁTKODOBÉ PROGNÓZOVANIE DOPYTU CESTUJÚCICH TEÓRIOU JEDNOROZMERNÝCH ČASOVÝCH RADOV

Účelom príspevku je identifikovať model ukazovateľa dopytu cestujúcich po prímestskej autobusovej doprave a analyzovať vhodnosť jeho použitia na prognostické účely, ktorý spĺňa požiadavky kladené na štatistickú významnosť parametrov ako i náhodnosť reziduálnej zložky skúmaného časového radu. Za účelom identifikácie konštrukcie presnejšieho a spoľahlivejšieho modelu v porovnaní s tými používanými v súčasnosti sú použité modely exponenciálneho vyrovnávania a viacnásobnej regresnej úlohy ako i Boxova-Jenkinsova metodológia. Prognostická presnosť modelov je vyhodnocovaná porovnaním hodnôt priemerných absolútnych percentuálnych chýb jednotlivých prístupov k modelovaniu. V súlade s hlavným cieľom príspevku bol identifikovaný ARIMA model, ktorý s výnimkou normality rezíduí spĺňa štatistické kritériá. Napriek tejto limitácii bola pre identifikovaný model preukázaná najlepšia prognostická presnosť v porovnaní s ostatnými navrhnutými modelmi. Zistenia prameniace z výstupov výskumnej aktivity prezentovaných v príspevku budú mať priaznivý vplyv na zvyšovanie presnosti a spoľahlivosti prognostických modelov dopytu cestujúcich.

KĽÚČOVÉ SLOVÁ

dopyt cestujúcich, modelovanie dopytu, krátkodobé prognózovanie dopytu, prímestská autobusová doprava

REFERENCES

[1] **Gnap, J.**, **Poliak, M.**, **Konečný, V.**: 2008a. *Prognóza vývoja pre okresy Žilinského kraja obsluhované SAD Žilina*. Žilina: FPEDaS ŽU v Žiline; 2008

[2] **Gnap, J.**, **Poliak, M.**, **Konečný, V.**: 2008b. *Prognóza vývoja pre okresy Žilinského kraja obsluhované SAD Liptovský Mikuláš*. Žilina: FPEDaS ŽU v Žiline; 2008

[3] **Cyprich, O.**: *Modelovanie dopytu cestujúcich po prímestskej autobusovej doprave*. Žilina: Žilinská univerzita v Žiline, Fakulta prevádzky a ekonomiky dopravy a spojov, Katedra cestnej a mestskej dopravy; 2012

[4] **Jugović, A.**, **Hess, S.**, **Jugovic, T.P.**: *Traffic demand forecasting for port services*. Promet - Traffic&Transportation [Internet]. 2011 [cited 2012 April 14]; 23(1): 59-69. Available from: http://www.fpz.unizg.hr/traffic/index.php/PROMTT/article/view/149/56

[5] **Brnjac, N.**, **Abramović, B.**, **Maslarić, M.**: *Forecasting intermodal transport requirements on corridor X*. Promet - Traffic&Transportation [Internet]. 2010 [cited 2012 April 20]; 22(4): 303-307. Available from: http://www.fpz.unizg.hr/traffic/index.php/PROMTT/article/view/195/100

[6] **Krasić, D.**, **Gatti, P.**: *Forecasting methodology of maritime passenger demand in a tourist destination*. Promet - Traffic&Transportation [Internet]. 2009 [cited 2012 April 20]; 21(3): 183-190. Available from: http://www.fpz.unizg.hr/traffic/index.php/PROMTT/article/view/224/129

[7] **Dicová, J.**, **Ondruš, J.**: *Trend of public mass transport indicators – as a tool of transport management and development of regions*: Communications – Scientific Letters of the University of Žilina. 2010; 12(3A): 121-126

[8] **Karlaftiss, M.G.**, **Vlahogianni, E.I.**: *Statistical methods versus neural networks in transportation research: Differences, similarities and some insights*. Transportation Research Part C [Internet]. 2011 [cited 2012 April 13; 19 (3): 387-399. Available from: http://www.sciencedirect.com/science/article/pii/S0968090X10001610

[9] **Cyprich, O.**: *Application of Univariate Time Series Theory to Passenger Demand Forecasting*: Communications – Scientific Letters of the University of Žilina. 2011

[10] SAS LE 4.1 [software]. Cary, NC : SAS Institute Inc. 2006

[11] SAS 9.1.3 [software]. Cary, NC : SAS Institute Inc. 2003

[12] **Cyprich, O.**: *Modelovanie vývoja vybraných kvantitatívnych ukazovateľov ako nástroja riadenia dopravnej spoločnosti*, Ph.D. thesis concept. Žilina: University of Žilina; 2010

[13] **Cipra, T.**: *Analýza časových řad s aplikacemi v ekonomii*. Praha/Bratislava: STNL/ALFA; 1986

[14] **Arlt, J.**, **Arltová, M.**: *Ekonomické časové řady*. Praha: Professional Publishing; 2009

[15] **Chatfield, Ch.**, **Yar, M.**: *Holt-Winters forecasting: some practical issues*: The Statistician; 1988

[16] **Dagum, E.B.**: *The X-11-ARIMA/88 Seasonal Adjustment Method*: Foundations and User´s Manual, Statistics Canada. Ottawa; 1988

[17] U.S. Bureau of the Census: *X-12-ARIMA Seasonal Adjustment Program - Version 0.2.8*, U.S. Bureau of the Census. Washington; 2001

[18] U.S. Bureau of the Census: *X-12-ARIMA Reference Manual - Version 0.2.8*, U.S. Bureau of the Census. Washington; 2001

[19] **Leonard, M.**: *Large-Scale Automatic Forecasting*. Millions of Forecasts [Internet]. 2002 [cited 2012 April 28]. Available from: https://support.sas.com/rnd/app/papers/largescale.pdf

[20] **Filiben, J.J.**, **Heckert, A.**: *Exploratory data analysis*. NIST/SEMATECH e-Handbook of Statistical Methods, NIST/SEMATECH, [Internet]. 2003 [cited 2012 April 15]; Available from: http://www.itl.nist.gov/div898/handbook/

[21] **Ljung, G.M.**, Box GEP. *On the measure of lack fit in time series models*: Biometrika. 1978

[22] **Hamilton, J.D.**: *Time Series Analysis*. Princeton: Princeton University Press; 1994

[23] **Dickey, D.A.**, **Hasza, D.P.**, **Fuller, W.A.**: *Testing for unit roots in seasonal time series*: Journal of the American Statistical Association. 1984

[24] **Shapiro, S.S.**, **Wilk, M.B.**: *An analysis of variance test for normality* (complete samples): Biometrika. 1965

[25] **Marček, D.**, **Marček, M.**: *Analýza, modelovanie a prognózovanie časových radov s aplikáciami v ekonomike*. Žilina: EDIS; 2001

[26] Duke University: *What to look for in regression model output* [online], Duke university, Durham, [cited 2012 April 16] available from: http://www.duke.edu/~rnau/411regou.htm

[27] **Cyprich, O.**, **Holeša, L.**: *Analýza použiteľnosti metódy X-12-ARIMA pri prognózovaní a dekompozícii časových radov dopytu cestujúcich*: Perners Contacts, 2012; 7(1): 13-25