Application of an Econometric Method for Forecasting Cellular Mobile Traffic

Suzana Miladic[1], Goran Markovic[2] and Valentina Radojicic[3]

Abstract – **This paper presents an econometric traffic forecasting model that is applied to the cellular mobile network of one telecom operator in the Republic of Srpska (BiH). As a suitable metric of traffic volume we chose the total monthly traffic minutes. The model deals with several explanatory variables affecting the changes to the traffic volume. These variables are measured by using available historical data and by applying regression analysis for estimation the model parameters. Through the numerical results the short-term forecasted results of cellular mobile traffic are given.**

Keywords – **Forecasting, econometric model, cellular mobile traffic.**

I. Introduction

A correct forecast of the future traffic needs, based on reliable and relevant input historical data, is the basis for the planning and upgrading of telecommunication networks. Generally, forecasting in the field of telecommunications involves estimation the number of subscribers as well as the traffic volume.

Forecasting as a process involves several steps: first it is necessary to define the problem or purpose of forecasting, then to collect the data, choose the suitable method of forecasting, estimation of necessary parameters and finally make the analysis and perform necessary calculations. Today, changes in market dynamics, especially rapid growth of the Internet and the number of mobile telephony users encourages telecommunications operators to respond the technological challenges. Consequently, the network planning methodology has to consider the market and demand forecasting in the first stage.

If the forecast is not reliable as a result may appear insufficient or excessive dimensioning of certain network capacity, resulting in operator losses and reducing the quality of services for the end users [1]. One of the most important parameters for the process of market and demand forecasting is the forecasted network traffic.

There are many forecasting methods. Quantitative or statistical and qualitative or non-statistical methods represent one possible classification. Quantitative methods process the data using statistical principles and have a numerical output while qualitative methods are based on other principles such as extrapolation, reasoning, etc., so the output may not be numerical. This paper deals with one quantitative traffic forecasting method and its application on real traffic data.

The methods of traffic forecasting are defined by the ITU-T recommendations E.506 and E.507 [2], [3]. Some important methods include the following: curve fitting models (linear, parabolic, exponential, logistic, Gompertz trend), smoothing models (moving average, Holt, Holt-Winter), autoregression models, ARIMA (*Autoregressive Integrated Moving Average)*, regression, Kalman and econometric models. Models for total traffic forecasting based on increasing the number of subscribers and models for point-to-point traffic forecasting are given in [4]. Forecasting network traffic by means of neural networks and linear models is presented in [5].

The paper is organized as follows. In the second section we formulate the problem, describe the forecasting model and discuss its parameters. The model is applied to real data in the third section, and the obtained traffic forecast for the cellular mobile operator is presented. Concluding remarks and further work are given in the forth section.

II. The forecasting model

Inspired by [2], [3] and [6] we selected an econometric model to perform forecasts of total mobile telephony traffic for one telecom operator in the Republic of Srpska (BiH). Econometric model defines a set of explaining variables from which the unit of interest (traffic) is calculated according to given mathematical model. The model parameters have been determined through regression analysis. As a suitable metric of the traffic volume we chose monthly traffic minutes (denoted by T) which refers to the number of subscribers times the minutes of use per subscriber and month (denoted by MoU). The relevant historical data and the number of observations required for this model are obtained from sources [7] and [8].

A. Modelling the variables

The model considers following three independent or explanatory variables: the spending power S in the market (the market is specified over country area or network traffic area), the actual service cost C and the market penetration P. These variables affect the changes to the traffic volume T, which represents the dependent variable [6].

It is shown that the demand for mobile telecommunications services appears to be relatively elastic. It means that subscribers are highly sensitive to service price, which has

[1]Suzana Miladic (Ph.D student) is with the Faculty of Transport and Traffic Engineering at University of East Sarajevo, Bosnia and Herzegovina, e-mail: miladics@hotmail.com

[2]Goran Markovic is with the Faculty of Transport and Traffic Engineering at University of Belgrade, Serbia, e-mail: g.markovic@sf.bg.ac.rs

[3]Valentina Radojicic is with the Faculty of Transport and Traffic Engineering at University of Belgrade, Serbia, e-mail: valentin@sf.bg.ac.rs

direct impact on traffic volume. Obviously, the traffic volume depends on the spending power of population. Usually, as a measure of a spending power, the Gross National Product (GDP) per capita is used.

The costs of mobile service, C_t includes various parameters such as: subscription fees, usage tariffs, the cost of a terminal, etc. In this paper, we made a simplified assumption that the operator's revenues of telephony traffic service are based on paid traffic per minute. The structure of the traffic includes pre-paid, post-paid and VPN mobile users. The revenues are measured as average revenues per user and month (ARPU) and have been expressed in national monetary units (BAM).

The mobile service penetration rate, α, is the percentage of the relevant country's population that has purchased a mobile service during the time under study. Service life cycles may go through several stages: Introduction, Growth and Maturity stage. After the rapid sales during the Growth stage, the market starts to be saturated as there are fewer new subscribers. When service sales reach peak values and market becomes saturated, the penetration rate becomes very low. It could be modeled by the penetration function $P(\alpha)$ as follows [6]:

$$P(\alpha) = \left[1 - \cos\left(\alpha^2 \pi\right)\right] / 2 \qquad (1)$$

B. Mathematical background

Considering all variables indicated above, the main form of the traffic forecasting model could be described as follows [6]:

$$T_t = e^{b_0} \cdot S_{t-1}^{b_1} \cdot C_t^{b_2} \cdot P(\alpha_t)^{b_3} \qquad (2)$$

Transforming Eq. 2 into linear form, the model thus becomes:

$$\ln T_t = b_0 + b_1 \ln S_{t-1} + b_2 \ln C_t + b_3 \ln P(\alpha_t) \qquad (3)$$

Finally, if we introduce shift variables, it is obtained:

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 \qquad (4)$$

The variables are indexed by time, t, and the model parameters b_0, b_1, b_2, b_3 are estimated using the ordinary least square (OLS) technique. The parameters b_1, b_2, b_3 can be interpreted as elasticity parameters indicating the impact of each explanatory variable or how much they attribute to the variation of dependent variable. This could be shown by index of determination R^2 and adjusted coefficient of determination R^2 (adj) which provides better estimate of the actual value of R^2, if we talk about small sample. These coefficients range from [-1,1]. The closer the marginal values, better the model is. The multiple linear regressions have been applied here, since the model is nonlinear in variables (because of the logarithms) and linear in parameters.

The model provides forecasts of the dependent variable value based on the changes in the values of explanatory variables.

III. Numerical Results

In this section, we have applied the model described above to the mobile network of one telecom operator in the Republic of Srpska. The historical data considering the GDP values and population size are given in Table I. The relevant statistical data for the chosen mobile operator obtained from their annual reports, are given in Table II. MoU and ARPU are given in real terms while the total monthly mobile telephony traffic (minutes) and subscribers are given in millions.

TABLE I
RELEVANT DATA FOR THE REPUBLIC OF SRPSKA [8]

Year	Republic of Srpska	
	GDP	Population
2009	5 739	1 435 179
2010	5 805	1 433 038
2011	6 073	1 429 668
2012	6 006	1 429 290
2013	6 146	1 425 549

TABLE II
RELEVANT DATA FOR THE CHOSEN OPERATOR [7]

Year	OPERATOR				
	Traffic	MoU	ARPU	Sub.	P(α)
2009	50.74	41.2	15.4	1.231	0.78
2010	59.94	44.6	16.2	1.344	0.82
2011	61.15	44.4	15.3	1.377	0.80
2012	72.01	50.9	13.7	1.415	0.82
2013	77.98	54.34	13.4	1.435	1

We used the direct method to determine the model parameters. Since we made assumption that traffic variable T depends on several significant factors, it is necessary to forecast these factors (explanatory variables), firstly. Considering the historical data and by applying the OLS technique, the regression equations could be written, which are the basis for computing these values for the next few years (Table III). Therefore, we have: $x_1=5.650+0.101t$; $x_2=16.75-0.65t$; $x_3=0.712+0.044t$.

TABLE III
FORECASTED VALUES OF EXPLANATORY VARIABLES

Year	t_i	GDP ($x10^3$)	ARPU	P(α)
2014	6	6.256	12.85	0.98
2015	7	6.357	12.20	1.02
2016	8	6.458	11.55	1.06
2017	9	6.559	10.90	1.11

Forecasting of monthly traffic minutes for mobile telephony, considering the forecasted values of significant factors, is done by using the Eq. 4, whereby the shift variables are: $lnT=y$; $lnS=x_1$; $lnC=x_2$ and $lnP(\alpha)=x_3$. The parameters b_0, b_1, b_2, b_3 are estimated using the system of normal equations:

$$nb_0 + b_1\sum_{i=1}^{n} x_{1i} + b_2\sum_{i=1}^{n} x_{2i} + b_3\sum_{i=1}^{n} x_{3i} = \sum_{i=1}^{n} y_i \qquad (5)$$

$$b_0\sum_{i=1}^{n} x_{1i} + b_1\sum_{i=1}^{n} x_{1i}^2 + b_2\sum_{i=1}^{n} x_{1i}x_{2i} + b_3\sum_{i=1}^{n} x_{1i}x_{3i} = \sum_{i=1}^{n} x_{1i}y_i \qquad (6)$$

$$b_0\sum_{i=1}^{n} x_{2i} + b_1\sum_{i=1}^{n} x_{1i}x_{2i} + b_2\sum_{i=1}^{n} x_{2i}^2 + b_3\sum_{i=1}^{n} x_{2i}x_{3i} = \sum_{i=1}^{n} x_{2i}y_i \qquad (7)$$

$$b_0\sum_{i=1}^{n} x_{3i} + b_1\sum_{i=1}^{n} x_{1i}x_{3i} + b_2\sum_{i=1}^{n} x_{2i}x_{3i} + b_3\sum_{i=1}^{n} x_{3i}^2 = \sum_{i=1}^{n} x_{3i}y_i \qquad (8)$$

After deduction of sums and incorporating them into the equations, we have got:

$$5b_0 + 8,92b_1 + 13,47b_2 - 0,87b_3 = 20,77 \qquad (9)$$

$$8,92b_0 + 15,91b_1 + 24,02b_2 - 1,65b_3 = 37,08 \qquad (10)$$

$$13,47b_0 + 24,02b_1 + 36,3b_2 - 2,36b_3 = 55,91 \qquad (11)$$

$$-0,87b_0 - 1,65b_1 - 2,36b_2 + 0,19b_3 = -3,56 \qquad (12)$$

Solving this system of equations, we have determined the necessarily model parameters. Using the SPSS Statistics 17.0 tool, we have calculated the coefficient of determination R^2 and adjusted coefficient of determination R^2 (adj). The obtained results are given in Table IV.

TABLE IV
FITTED MODEL PARAMETERS

	b_0	b_1	b_2	b_3	R^2	R^2(adj)
	Const.	S_{t-1}	C_t	$P(\alpha_t)$		
Republic of Srpska Operator	13.568	0.077	-3.539	0.107	0.967	0.942

In terms of the model parameters, the first one, b_1 shows that, if the GDP is increased for 1 % the traffic will increase for 0.077 million of minutes. The second one, b_2 shows that, if the ARPU is changed for 1 % the traffic will change for 3.539 %. This value is inversely proportional to the traffic value and therefore the Pearson correlation is negative. Finally, the third one, b_3 shows that, if the penetration effect is higher for 1 %, the traffic value will be higher, too. Constant b_0 shows the value of T if there were no independent variables.

It could be noted that R^2 (adj) is more than satisfactory for the presented model and explain the 94,2 % of the variation in total demand (traffic) concerning with independent variables (elasticities denoted by b_1, b_2, b_3).

In the next step, we used the obtained parameters to forecast the monthly traffic minutes, using the historical data (2009-2013) to fit the model. If we apply this model for forecasting the monthly traffic minutes of mobile telephony, for the years 2014, 2015 and 2016, we will get the following values (in millions of minutes) 98.54; 126.17 and 153.95, respectively. Fig. 1 summarizes the results.

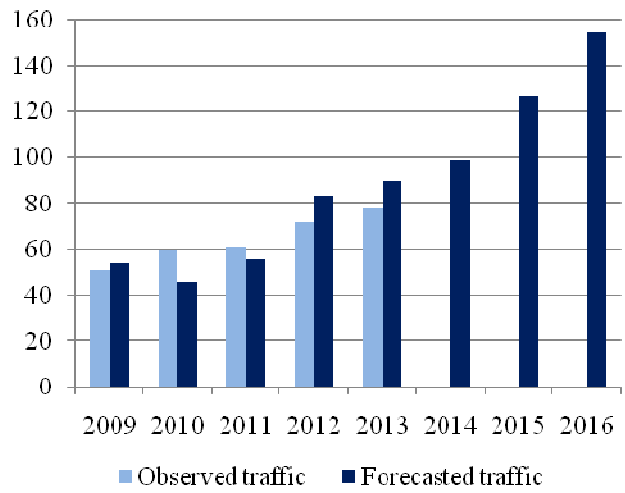

Fig. 1. Forecasted and observed monthly traffic volume (in millions of minutes)

It could be seen, that the forecasted traffic values are very similar to the obseved traffic values, which means that the proposed model can be used for forecasting the cellular mobile traffic. For a number of observations, using SPSS, we have determined the regression line with a deviations, shown in Fig. 2.

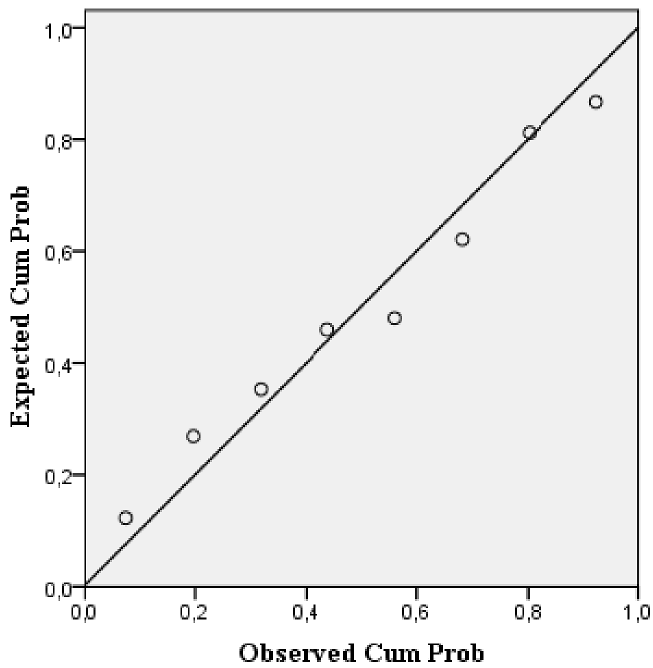

Fig. 2. The plot area of a regression

The observed data from 2009-2013 show that the revenues of the operator regarding to mobile telephony, since 2011, are slightly reduced due to the presence of smart phones on the mobile market. The forecasted values for 2014-2016 show that the volume of mobile cellular traffic is constantly increasing since the number of subscribers is increased every day.

IV. Conclusion

Traffic forecasting is one of the main tasks that precedes to telecommunication network dimensioning or upgrading and management of network resources. This paper presents one econometric model for traffic forecasting. It demonstrates that it is possible to identify a few essential variables which could be used to obtain the reliable forecast of the cellular mobile traffic. The spending power, service costs and market penetration are certainly the main factors affecting the changes to the traffic volume.

The model gives accurate forecast and could be used for some other traffic types, for example: Internet traffic, fixed telephony etc. Besides this, future work could include the data from more operators, since BiH has three dominant operators which provide mobile telephony service. At this moment, only the data for one operator were publicly available.

Additionally, it could be considered more operators from different countries, in order to do some comparative analysis of the influencing factors on the predicted traffic value.

It is important to mention, that the quality of obtained results depends on the reliable historical data, the right choice of forecasting model as well as the precise estimation of relevant model parameters.

Acknowledgement

This paper is partially supported by the Serbian Ministry of Education, Science and Technological Development by the research project TR-32025.

References

[1] V. Radojicic, A. Kostic-Ljubisavljevic, “Telecommunication network design under traffic forecast uncertainty”, PosTel, Conference Proceedings, pp. 281-291, Belgrade, Serbia, 2005.

[2] ITU-T Recommendation E.506 (rev.1): “Forecasting international traffic”, ITU-T Geneva, 1992.

[3] ITU-T Recommendation E.507: “Models for forecasting international traffic”, ITU-T Geneva, 1993.

[4] H. Leijon, “Basic forecasting theories: A brief introduction”, ITU, 1998.

[5] I. Klevecka, “Forecasting network traffic: a comparison of neural networks and linear models”, Proceedings of the 9th International Conference (RelStat’09), pp. 170-177, Riga, Latvia, 2009.

[6] A. Arvidsson et al., “Forecasting cellular traffic: an econometric approach”, International Teletraffic Congress, Beijing, Blekinge Institute of Technology, 2006 available on: http://www.i-teletraffic.org/fileadmin/ITCBibDatabase/2005/arvidsson05.pdf

[7] Mtel, *Annual Report*, Banja Luka, 2009-2013, available at: http://www.mtel.ba/article/819

[8] Institute of statistics, *Statistical bulletin: National Accounts*, no. 7, Banja Luka, 2014, available at: http://www2.rzs.rs.ba/static/uploads/bilteni/nacionalni_racuni/Bilten_Nacionalni_Racuni_7.pdf