Australian Journal of Basic and Applied Sciences, 1(4): 386-394, 2007
ISSN 1991-8178

A Time - Series Forecasting of Average Daily Traffic Volume

[1]M. Sabry , [1]H. Abd – El - Latif, [2]S. Yousef and [3]N. Badra

[1]Public Works Department, Faculty of Engineering, Ain Shams University, Cairo, Egypt.
[2]Department of Structural Engineering, Faculty of Engineering, Ain Shams University, Cairo, Egypt.
[3]Dep. of Physics and Engineering mathematics, Faculty of Eng., Ain Shams University, Cairo, Egypt.

Abstract: This paper presents a procedure for forecasting average daily traffic using a time-series analysis. The procedure uses a time-series analysis that models traffic volume as a logistic regression to model traffic volume over a period of years. Model parameters are estimated using least –square regression method. The model parameters are estimated for a certain Egyptian intercity road. A total of 14-year traffic data for the selected station in each direction (from 1990 to 2003) were used in the development of the model to predict traffic volume. All of the data sets were best represented by logistic regression model. The data set containing average annual, average monthly and average weekly daily traffic volume for years 1990 through 2000 was used to fit a time series model. The resulting models were used to forecast traffic volumes for year 2001 to 2003. The forecasted traffic volumes for the models were then compared with the actual traffic volumes for the period from January 2001 to December 2003 for a certain Egyptian intercity road on both directions.

Keywords: Traffic volume, Regression model, Logistic function, forecasting method.

INTRODUCTION

Studies on time-series and forecasting procedures have been developed (J Kim, 2003., C. Chen., 2001. C. Davis, 2002., B. Smith, 2002, *et al.,* B. Williams, 2001). Forecasting plays a major role in traffic engineering. Time series is a set of statistical observations, arranged in chronological order. One of the problems created in forecasting is the selection of a suitable model for a specific purpose. Selection of model depends on several factors including the context of forecasts, the availability of historic data, the degree of desirable accuracy, the cost of evaluation and the time to be forecasted. The standard demand forecasting method centers on the well-known UTPS computer package that is maintained by the U.S. Department of Transportation. The UTPS package models demand in five basic steps. A good review of the theory underlying the UTPS models is presented by Kanafani (1983). The UTPS package relies heavily on detailed survey data for parameter estimation, which is typically large scale origin destination interzonal travel behavior. A generation of disaggregate models was developed in order to reduce these data that looked at travel behavior in even greater detail. A summary of disaggregate approaches is found in Stopher and Meyburg (1975). These approaches were extended to examine the basic decision processes of travelers by using attitudinal techniques. A summary of these techniques is presented in Bemamin and Sen, (1982). Although many of the travel forecasting models are not based on econometric techniques, there is a direct correspondence between econometric forecasting techniques and the standard travel demand forecasting procedure. There are basically two approaches to econometric forecasting: Structural models that are permitted direct analysis to make decisions based on an evaluation of some criterion variable, and Time – series models that only examine the relationship of the criterion variable to time. These models permit only answers to questions concerning where the criterion variable will be in the future without regard to specific inputs. In travel demand, the criterion variable is average daily traffic volume. Past emphasis has been on developing better structural models with increasing attention to trip making decisions. The structural process, however, leads to the development and estimation of parameters for a series of models in each case, these estimates are subject to assumptions and measurement errors that compound each other progressively. Benjamin (1986) suggested that instead of a structural approach, under

Corresponding Author: N. Badra, Department of Physics and Engineering mathematics, Faculty of Engineering, Ain Shams University, Cairo, Egypt.

proper circumstances, only traffic trends need to be modeled. Such models only use average daily traffic data. The proposed technique requires substantially fewer observations and is demonstrated to be accurate for much longer periods. The availability and convenience of average traffic information is emphasized by Carey, Hendrickson and Siddharthan (1981). Nihan and Holmesland (1980) suggested the direct use of traffic volumes for traffic forecasting but only for short-term forecasts. The Box-Jenkins techniques used only for a 1-year forecast and require more than 60 periodic observations. Mahajan and Muller (1979) were summarized a set of logistic growth curves, including exponential, Gompertz and a family of dynamic growth curves. The characteristics and estimation of the logistic function parameters are discussed by Kruskal and Tanur (1977).

This paper uses Benjamin (1986) procedure to model traffic volume for a Tanta – Mansoura road (Station 9) in both directions DIR1 and DIR2 from a period of January 1990 to December 2003. The procedure assumes a logistic function to model traffic volume. Model parameters are estimated using ordinary least-squares regression in a period from January 1990 to December 2000. The model is tested in forecasting both average annual, average monthly, and average weekly daily traffic in a period from January 2001 to December 2003 in a Tanta – Mansoura road on both directions and compared with the actual traffic volume.

Methodology and Logistic Function:

Let V_{max} to be the maximum growth in traffic volume, V_{min} equal the minimum traffic volume. Solve the current average daily traffic volume V as a function of time t.

The proportion of growth p is defined by :

$$p = \frac{V - V_{min}}{V_{max}} \qquad (1)$$

This indicates that the current average daily traffic volume is defined by :

$$V = V_{min} + p\, V_{max} \qquad (2)$$

The potential growth 1 – p is defined by :

$$1 - p = \frac{V_{max} + V_{min} - V}{V_{max}} \qquad (3)$$

This indicates that the current average daily traffic volume is defined by :

$$V = V_{min} + V_{max} + (1 - p)\, V_{max} \qquad (4)$$

From (2) and (4), the rate of change of traffic volume at any time t is proportional to the proportion of growth p and the potential growth (1 – p). This indicates that :

$$\frac{dV}{dt} = V_{max} \frac{dp}{dt} \qquad (5)$$

The rate of change of traffic volume $\frac{dp}{dt}$ at any time t is assumed to be proportional to the proportion of growth p, the potential growth (1 – p) and an overall growth rate A_1.

This indicates that:

$$\frac{dp}{dt} = A_1 . p\,(1 - p) \qquad (6)$$

From (5) and (6), the following is obtained :

$$V_{max} \frac{dp}{dt} = V_{max} . A_1 . p\,(1 - p) \qquad (7)$$

Hence,

$$\frac{dp}{dt} = A_1 . p (1 - p) \qquad (8)$$

Separating the variables, it is found that:

$$\frac{dp}{p(1-p)} = A_1 \, dt \qquad (9\text{-a})$$

$$[\frac{1}{p} + \frac{1}{1-p}] dp = A_1 \, dt \qquad (9\text{b})$$

Integrating both sides, the following equation is obtained:

$$\ln p - \ln (1 - p) = A_1 t + A_o \qquad (10\text{-a})$$

$$\ln (\frac{p}{1-p}) \; A_1 t + A_0 \qquad (10\text{-b})$$

Substitute from (1) and (3) by p and (1 – p) in (10-b) :

$$\ln (\frac{V - V_{min}}{V_{max} + V_{min} - V}) \; A_1 t + A_0 \qquad (11)$$

Which is the well-known logistic function that will be used in this paper in estimating and forecasting traffic volume. Moreover, the exponent of the logistic function is not necessary to be a linear function of t but may be a polynomial function of t. Thus, the general form of (11) may be represented as follows :

$$\ln \; (\frac{V - V_{min}}{V_{max} + V_{min} - V}) = A_0 + A_1 t + A_2 t^2 + A_3 t^3 + A_4 t^4 + \ldots . \quad (12)$$

Table 1: Forecasted and observed average annual daily traffic volume form 2001 to 2003.

	DIR1	DIR2	TOTAL
$V_{min} + V_{max}$	12703	18091	30811
V_{2001} (forecast)	8514	9263	18394
V_{2001} (actual)	9376	9485	18861
% Error 2001=$((V_{actual} - V_{forecast})/V_{actual})x100$	9.2	2.34	2.48
Growth factor = $(V - V_{min})/V_{max}$	1.492	1.953	1.675
% Developed land = $V_{forecast} /(V_{min}+V_{max})$	67.02	51.2	59.7
V_{2002} (forecast)	9007	9834	19419
V_{2002} (actual)	9267	9482	18760
% Error 2002=$((V_{actual} - V_{forecast})/V_{actual})x100$	2.81	3.71	3.51
Growth factor = $(V - V_{min})/V_{max}$	1.41	1.84	1.59
% Developed land = $V_{forecast} /(V_{min}+V_{max})$	70.9	54.36	63.03
V_{2003} (forecast)	9482	10409	20423
V_{2003} (actual)	9390	9599	18988
% Error 2003=$((V_{actual} - V_{forecast})/V_{actual})x100$	0.98	8.44	7.58
Growth factor = $(V - V_{min})/V_{max}$	1.34	1.74	1.51
% Developed land = $V_{forecast} /(V_{min}+V_{max})$	74.64	57.54	66.28

Table 2: Results of regression analysis for average annual daily traffic from 1990 to 2000.

Parameters		DIR1	DIR2	TOTAL
Asymptotes:	V_{min}	2073	2325	4422
	V_{max}	10630	15766	26389
Coefficients:	A_0	-1.552	-1.993	-1.766
	A_1	0.134	0.146	0.157
	A_2	0.002598	-------	-------
R^2		0.8185	0.9027	0.9538

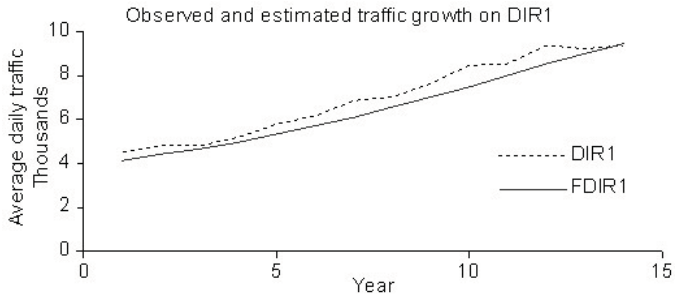

Fig. 1: Observed and estimated average annual daily traffic volumes on DIR1.

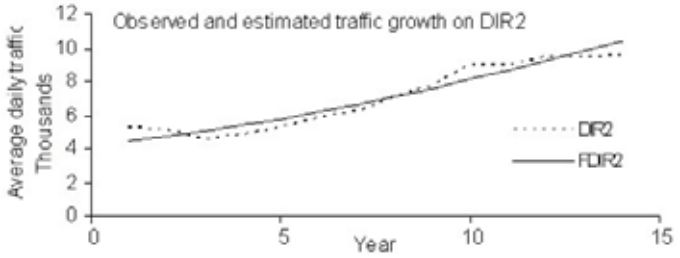

Fig. 2: Observed and estimated average annual daily traffic volumes on DIR2.

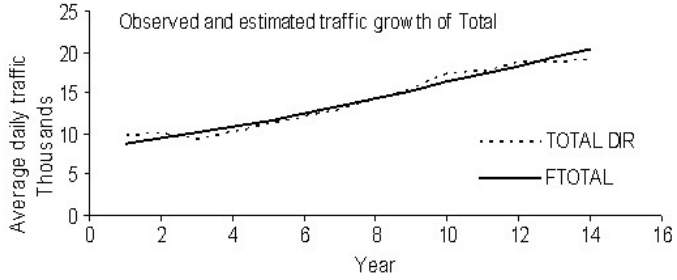

Fig. 3: Observed and estimated average annual daily traffic volumes of TOTAL.

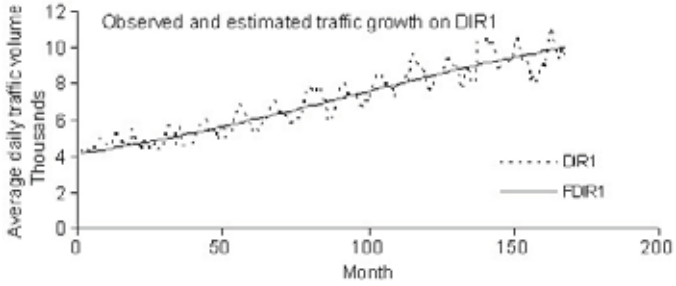

Fig. 4: Observed and estimated average monthly daily traffic volumes on DIR1.

The parameters A_o, A_1, A_2, A_3 and A_4 … are calculated by using the least squares regression analysis with $\ln (V - V_{min} / V_{min} + V_{max} - V)$ as the dependent variable, and t, t^2, t^3 and t^4 … as the independent variables. The statistical package SPSS was used for the analysis work of this paper. Goodness of fit measures available for the least squares regression analysis such as R^2 is also estimated. Use the data available in a Tanta – Mansoura road on both directions in a period from January 1990 to December 2000 for both average annual, average monthly and average weekly daily traffic volume. The time-series forecasts compared with observed traffic volume and orecasts observations in a period from January 2001 to December 2003.

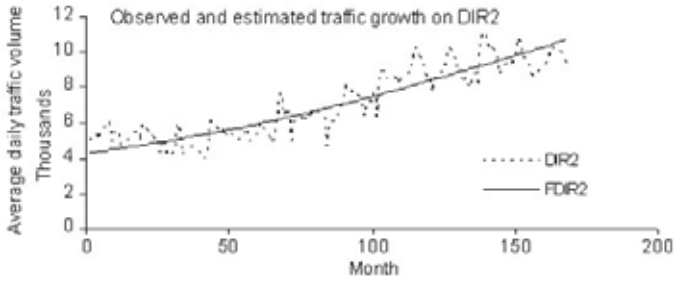

Fig. 5: Observed and estimated average monthly daily traffic volumes on DIR2.

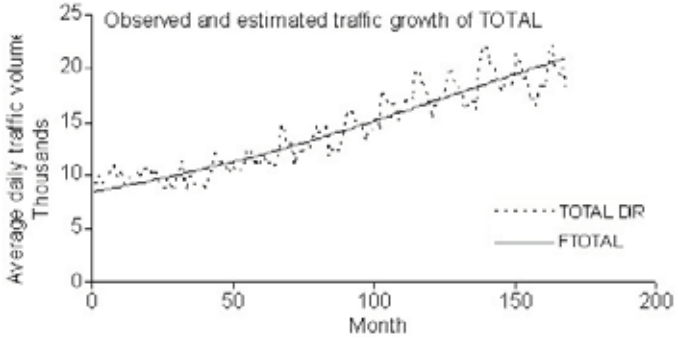

Fig. 6: Observed and estimated average monthly daily traffic volumes of TOTAL.

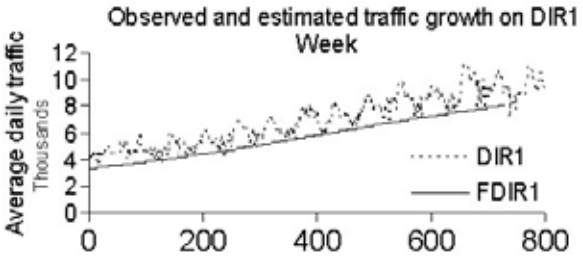

Fig. 7: Observed and estimated average weekly daily traffic volumes on DIR1.

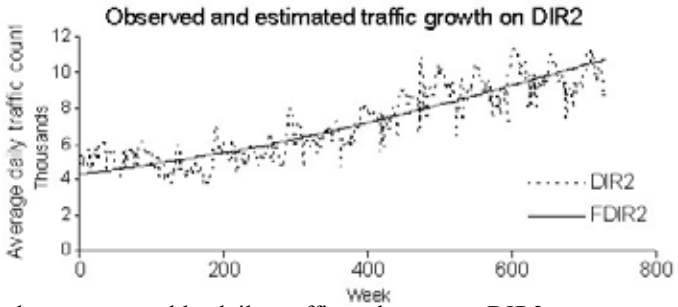

Fig. 8: Observed and estimated average weekly daily traffic volumes on DIR2.

Graphical and Functional Analysis:

Use the data available of a Tanta – Mansoura road on both directions in a period from January 1990 to December 2003 for both average annual, average monthly and average weekly daily traffic volume. Results of the forecasting logistic function for average annual daily traffic are presented in Figures 1, 2 and 3. Also, results of forecasting logistic function for average monthly daily traffic are plotted for the same

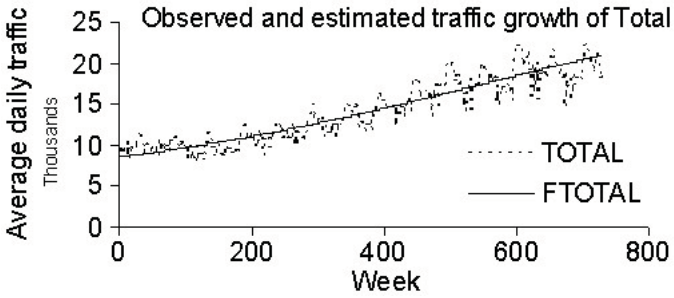

Fig. 9: Observed and estimated average weekly daily traffic volumes of TOTAL

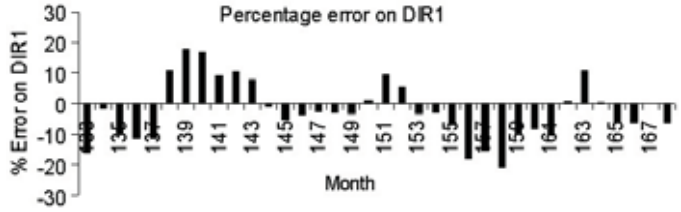

Fig. 10: % errors in forecasting average monthly daily traffic from 2001 to 2003 on DIR1.

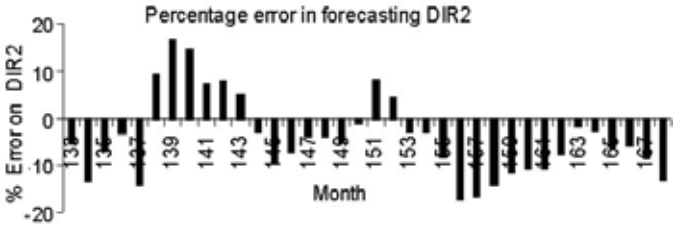

Fig. 11: % errors in forecasting average monthly daily traffic from 2001 to 2003 on DIR2.

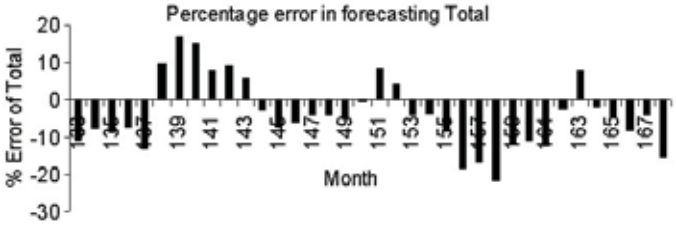

Fig. 12: % errors in forecasting average monthly daily traffic in 2001 to 2003 of TOTAL.

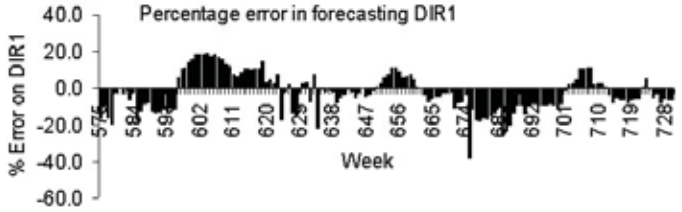

Fig. 13: % errors in forecasting average weekly daily traffic from 2001 to 2003 on DIR1.

period are presented in Figures 4, 5 and 6. While, results of forecasting logistic function for average weekly daily traffic are presented in Figures 7, 8 and 9. Results of the percentage of errors in forecasting average monthly traffic volume on both directions are shown in Figures 10, 11 and 12. Forecasts traffic volumes are compared with observed traffic volumes in a period from January 2001 to December 2003 and are summarized in Table 1.

Table 3: Logistic forecasted and observed traffic from 2001 to 2003.

Year	Traffic Volume	DIR1	DIR2	TOTAL
2001	Observed Traffic	9376	9485	18861
	Logistic point estimate	8514	9263	18394
2002	Observed Traffic	9267	9482	18760
	Logistic point estimate	9007	9834	19419
2003	Observed Traffic	9390	9599	18988
	Logistic point estimate	9482	10409	20423

Table 4: Comparison between the proposed and ordinary logistic forecasts traffic in 2003.

		Logistic function	Ordinary logistic function
DIR1	Estimated traffic 2003	9482	10796
	% Error 2003	0.98	14.97
DIR2	Estimated traffic 2003	10409	10631
	% Error 2003	8.44	10.75
TOTAL	Estimated traffic 2003	20423	21396
	% Error 2003	7.58	12.68

Table 5: Results of regression analysis for average monthly daily traffic from 1990 to 2000.

Parameters		DIR1	DIR2	TOTAL
Asymptotes:	V_{min}	2073	2325	4422
	V_{max}	10630	15766	26389
Coefficients:	A_0	-1.438	-1.952	-1.712
	A_1	0.01511	0.01234	0.843
R^2		0.87	0.711	0.843

Table 6: Results of regression analysis for average weekly daily traffic from 1990 to 2000.

Parameters		DIR1	DIR2	TOTAL
Asymptotes:	V_{min}	2073	2325	4422
	V_{max}	10630	15766	26389
Coefficients:	A_0	-1.437	-1.952	-1.710
	A_1	0.003488	0.00285	0.003053
R^2		0.85	0.689	0.822

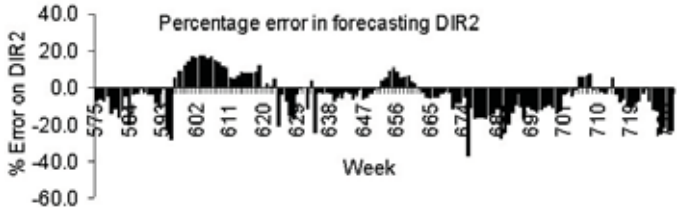

Fig. 14: % errors in forecasting average weekly daily traffic from 2001 to 2003 on DIR2.

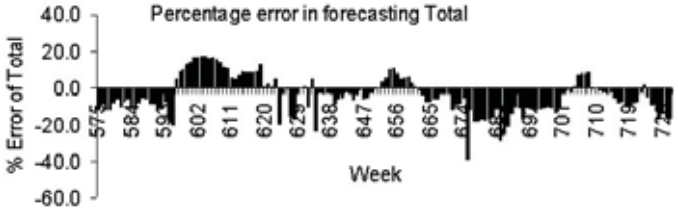

Fig. 15: % errors in forecasting average weekly daily traffic in 2001 to 2003 of TOTAL.

Model parameters for the available data for the average annual daily traffic volume are estimated using the regression analysis. The values of the parameters are shown in Table 2.

Comparison between Logistic Forecasts and Observed Traffic:

Results of observed and estimated values of average annual daily traffic volume using the logistic function analysis in the period from 2001 to 2003 are summarized in Table 3.

Comparison between Two Logistic Forecasted and Observed Traffic:

Two logistic functions are used in forecasting traffic volumes in 2003 using the regression analysis. The first logistic function is the logistic function presented in this paper and the second function is the original logistic function ln V as the dependent variable and t, t^2, t^3 and t^4 … as the independent variables. The proposed logistic function is considered the best according to the error estimates. The results are summarized in Table 4.

Logistic Forecasts for Average Monthly Daily Traffic:

Model parameters for the available data for the average monthly daily traffic volume are estimated using the regression analysis. The values of the parameters are shown in Table 5.

Logistic Forecasts for Average Weekly Daily Traffic:

Model parameters for the available data for the average weekly daily traffic volume are estimated using the regression analysis. The values of the parameters are shown in Table 6.

Percentage of errors in forecasts traffic volumes of average weekly daily traffic volume are estimated in the period from January 2001 to December 2003. Results of the percentage errors in forecasting average weekly traffic volumes on both directions are shown in Figures 13, 14 and 15. It has been found that the percentage of errors in forecasting the average weekly daily traffic volume was relatively higher compared with the forecasting average annual daily traffic volume. This is because of the nature of the logistic function is always increasing when time increasing which is not true in practical.

Conclusion:

This paper presents a procedure for forecasting average annual, monthly and weekly daily traffic volume using logistic regression analysis. The procedure assumes a logistic function to model traffic volume. The proposed logistic function uses Benjamin, procedure to model traffic volume for a Tanta – Mansoura road in both directions DIR1 and DIR2 from a period of January 1990 to December 2003. The model parameters are estimated for a Tanta – Mansoura road in each direction using ordinary least-squares regression in the period from January 1990 to December 2000. Comparison between logistic forecasts and observed traffic are presented for the Street on both directions for the period from January 2001 to December 2003. The model is tested in forecasting both average annual, average monthly, and average weekly daily traffic in a period from January 2001 to December 2003 in the street on both directions and compared with the actual traffic volume. Two logistic functions forecasts are estimated using the regression analysis for the data in 2003. The first logistic function is the logistic function presented in this paper and the second function is the original logistic function ln V as the dependent variable and t, t^2, t^3 and t^4 … as the independent variables. In comparing forecasted and observed traffic of the Street on both directions, the logistic function is considered the best in forecasting average annual daily traffic. The technique used in this paper requires fewer observations and is accurate for much longer periods. While, the proposed technique is not accurate in forecasting short periods such as average monthly and average weekly daily traffic since the nature of the logistic function is always increasing when time period is increasing which does not accurate in practice. Another major disadvantage of these models is that they are not sensitive to large changes. The statistical package SPSS was used for the analysis work of this paper. It has been found that the percentage of errors in forecasting the average monthly daily traffic volume was relatively higher compared with the forecasting average annual daily traffic volume. This is because of the nature of the logistic function is always increasing when time increasing which is not true in practical.

REFERENCES

J. Kim, 2003. " Forecasting Autoregressive Time Series with Bais- corrected Parameter Estimators", International Journal of Forecasting, 19: 493-502.

C. Chen and S. Grant, 2001. "Use of Sequential Learning for Short Term Traffic Flow Forecasting", Transportation Research, 9: 319-336.

C. Davis, 2002. " Statistical Methods for the Analysis of Repeated Measurements ", New York.

B. Smith, B. Williams, and R. Oswald, 2002. " Comparison of Parametric and Nonparametric Models for Traffic Flow Forecasting ", Transportation Research, 10: 257-321.

B. Williams, 2001. “Multivariate Vehicular Traffic Flow Prediction: Evaluation of ARIMAX Modeling”, Transportation Research Board, Washington, pp: 194-200.

Kanafani A., 1983. “Transportation Demand Analysis”, Mc Grow-Hill, New York.

Stopher P. and A. Meyburg, 1975, “Urban Transportation Modeling and Planning ”, Lexington Books.

Benjamin J. and L. Sen, 1982. “ A Manual of Procedures to Analyze Attitudes Towards Transportation ”, U. S. Departement of Transportation, Washington, DC.

Benjamin J., 1986. “ A Time – Series Forecast of Average Daily Traffic Volume ”, Transportation Research, 20: 51-60.

Carey M., C. Hendrickson and K. Siddharthan, 1981. “A Method for Direct Estimation of Origin/Destination Trip Matrices”, Transportation Science, 15: 32-49.

Nihan N. and K. Holmesland, 1980. “Use of the Box and Jenkins Time Series Techniques in Traffic Forecasting”, Transportation, 9: 125-143.

Mahajan V. and E. Muller, 1979, “Innovation Diffusion and New Product Growth Models in Marketing”, Marketing Journal, 43: 55-68.

Kruskal W. and J. Tanur, 1977. “International Encyclopedia of Statistics”, New York.