Univariate Time Series Modeling for Traffic Volume Estimation

Presented by:

KARTIKEYA JHA
&
SHRINIWAS S. ARKATKAR
Civil Engineering Department
Birla Institute of Technology & Science, Pilani, Rajasthan

Traffic Forecasting

Process of estimating the number of vehicles or people that will use various transportation facilities in the future.

- Collection of data on current traffic → Combined with other known data, such as population, economic growth rate, employment rate etc. → Feeding it with predicted data for chosen explanatory variables → Estimates of future traffic.

- In this study, forecasting analysis has been done for total vehicular population in India (to develop an insight) and Time Series Analysis on AADT data from PeMS, DOT, California.

Methods Generally Adopted for Traffic Forecasting

❖ Predicts a dynamic variable. So, a no. of approaches may be adopted for traffic forecasting depending upon the situation at hand.

- ✓ **Trend line analysis**
- ✓ **Trip Generation**
- ✓ **Econometric Demand Models**
- ✓ **Turning Movement Calculation**
- ✓ **Pattern Matching Technique**
- ✓ **Advanced Traveler Information System (ATIS)**
- ✓ **Time Series Analysis**

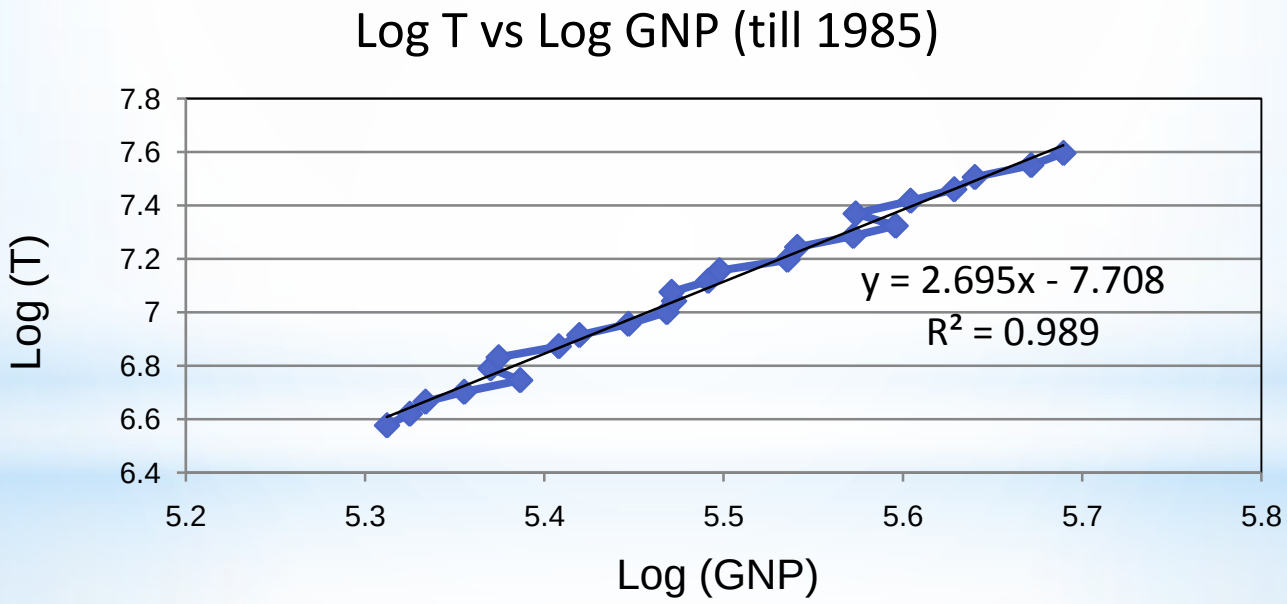
Methodology of Approach

- Time Series analysis done for IRC data for comparison with 2 other methods; separate TS analysis for AADT data from PeMS, DOT, California.

- Data cited from “Time Series Data on Road Transport Passenger and Freight Movement (1951-1991)”, Special Publication 45, Indian Roads Congress, New Delhi, 1996 and PeMS, DOT, California.

Trend Line Analysis

IRC data for vehicular population has been used to establish the traditional relationship “Log T= a + b Log GNP”. The data used for analysis is for the years 1961-1985 (25 years) and estimation has been done for the year 1996 (11 years ahead in future).



Trend Line Analysis..

From this plot, the following results have been obtained:

Estimated value for 1996 = 209702058
Actual value (IRC figure for 1996) = 108336195
Error= 93.57 %

Thus, we have an unacceptably high error value, resulting from an absurdly high overestimation. This invokes the need for a method that lends more dependability and is more logical to arrive at more acceptable results.

Econometric Analysis

• The demographic/econometric indicators chosen for analysis are Population and Per Capita Income. The data for these have been sourced from www.indiastat.com

• Regression Analysis on SPSS; done for the same period as that for Trend Line Analysis.

Results:

From SPSS:

Equation: Log T= -5.613 + 0.414 Log PCI + 4.121 Log P

Estimated value for 1996= 115055054

Actual value for 1996= 108336195

Error= 6.202%

Time Series Analysis

For univariate time series analysis on STATA, same data (from IRC) from years 1951-1985 (35 years) has been used for analysis. The estimation has been done for the same target year 1996.

The Box & Jenkins methodology has been used and ARIMA (Auto-Regressive Integrated Moving Average) technique has been adopted for analysis.

✓ Available data had to be differenced twice to achieve stationarity (a pre-requisite for Time Series Analysis).
✓ The Dickey Fuller and Philip Perron tests were conducted to confirm stationarity.
✓ Out of 9 prospective models, the model ARIMA (7,2,3) was chosen based on least RMSE (Root Mean Square Error) and Maximum Likelihood Rule.
✓ The table showing the reason for this choice is shown on subsequent slides.

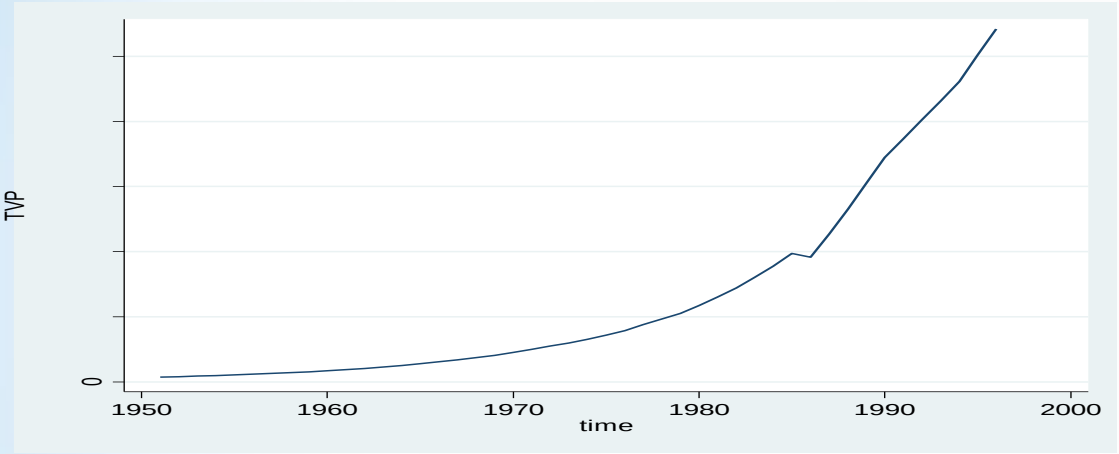
Time Series Analysis…

• **Box and Jenkins Methodology-** An iterative three-stage process of model selection, parameter estimation and model checking. The five broad steps include the following:

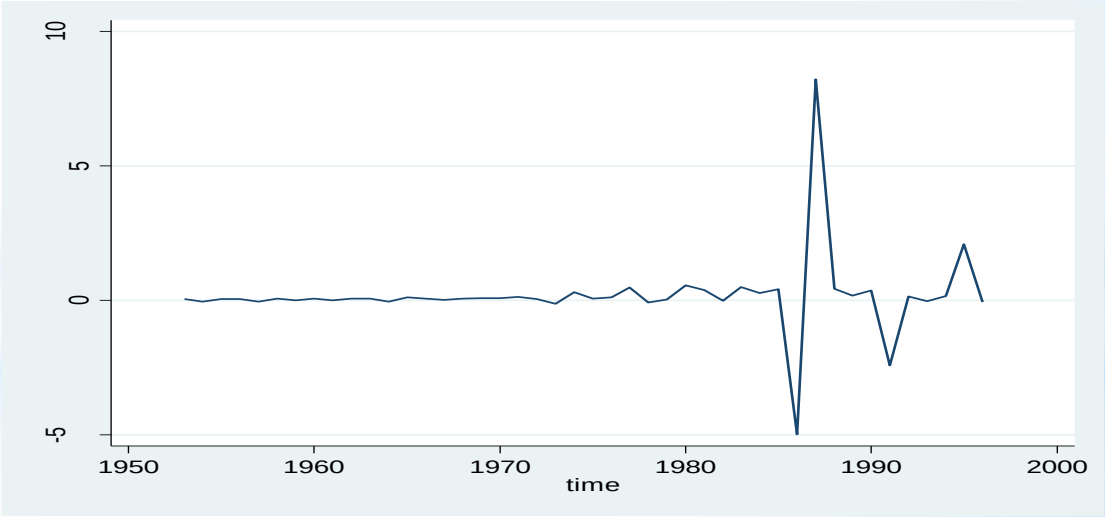
➢ Checking for stationarity and transforming the data set such that variance of the data set is stabilised and the mean becomes constant.

➢ Identification of the parameters of the model: To get the order of **AR** and **MA** process, the autocorrelation function and partial autocorrelation function are studied. An autoregressive process is a function of **lagged dependent variables** and a moving average process a function of **lagged error terms**. If a series needs to be differenced d times before it is stationary, the series is said to be integrated to degree d**.** This d gives us the variable **I.**

Time Series Analysis…



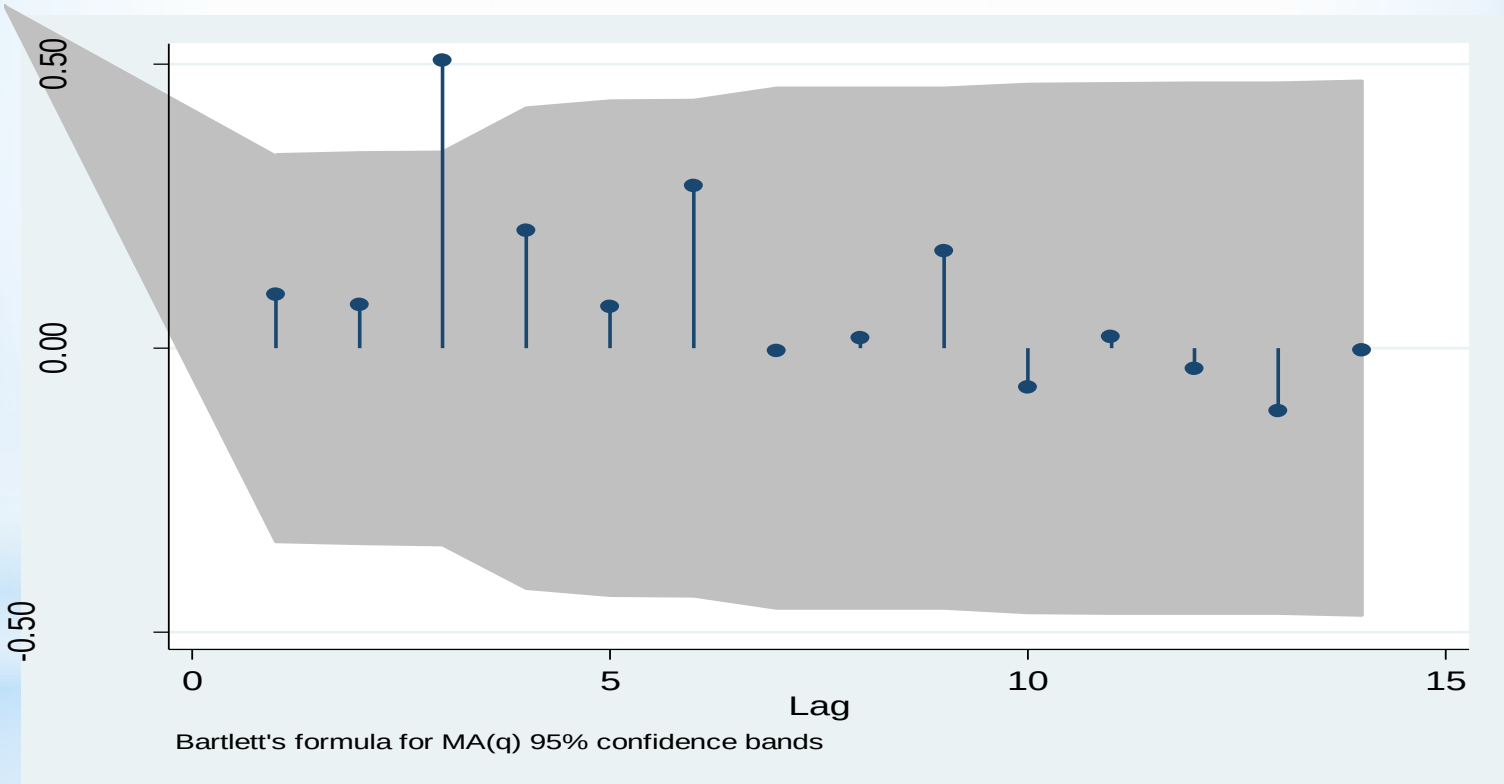
Non-stationary data



Stationary data

Time Series Analysis…

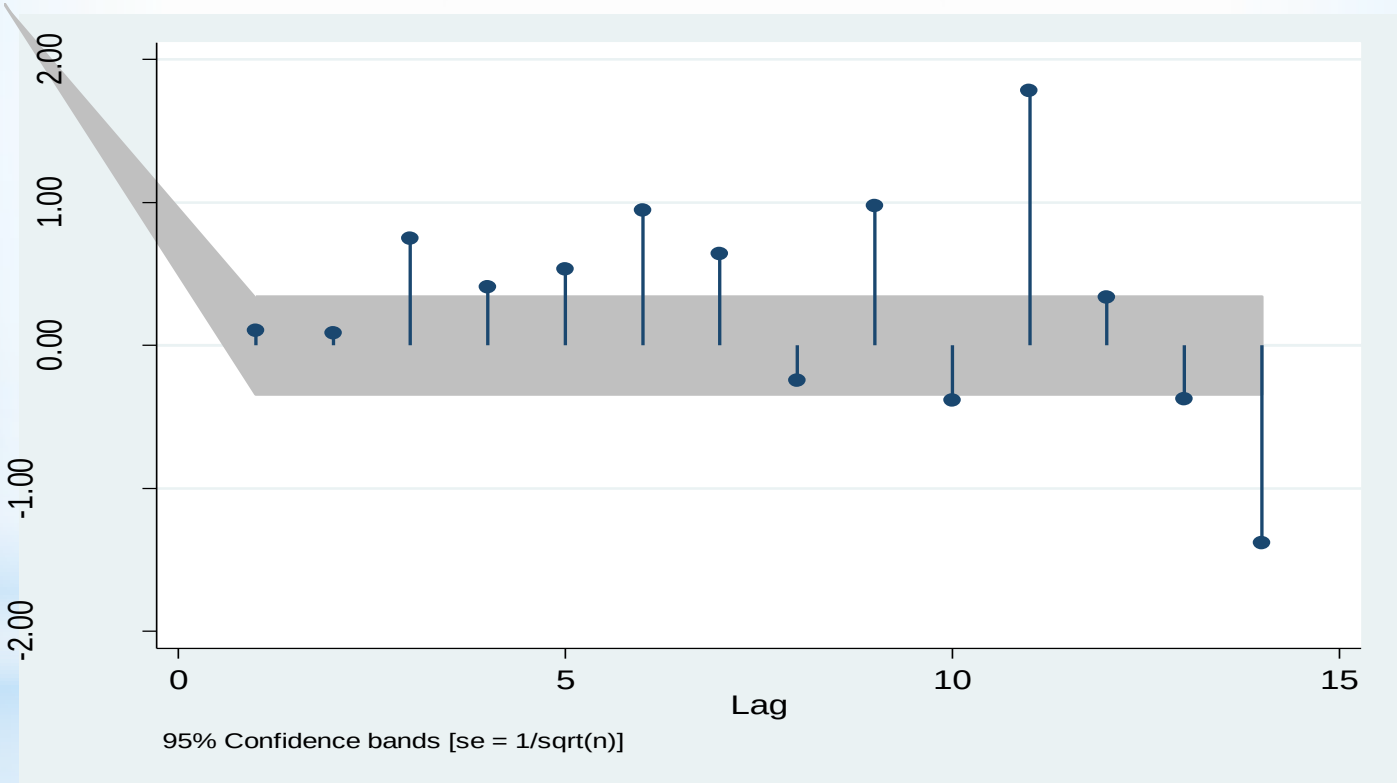
The results of ACF (Auto-correlation Function) for estimating MA part of ARIMA in the time series analysis model are shown below:



ACF Plot for estimating MA

Time Series Analysis…

The results of PACF (Partial Auto-correlation Function) for estimating AR part in ARIMA of the time series analysis model tests are shown below:



PACF Plot estimating AR

Time Series Analysis…

- Performing diagnostic checks: The residuals of the model should be uncorrelated. In other words, should be a white noise. One way to test this is to get **a portmanteau test statistic.** This is also called the **White Noise Test.**

For ARIMA(7,2,3), Portmanteau (Q) statistic as 3.3464 which is less than the critical value of 23.7 at 5% level of significance.

- Forecasting: Dynamic forecasting has been done in this analysis. For the purpose of forecasting, the period from 1986-1996 has been kept aside taken as the 'forecasting window'. These observed values will then be compared with the forecasted values to calculate the root mean forecasting error.

Time Series Analysis…

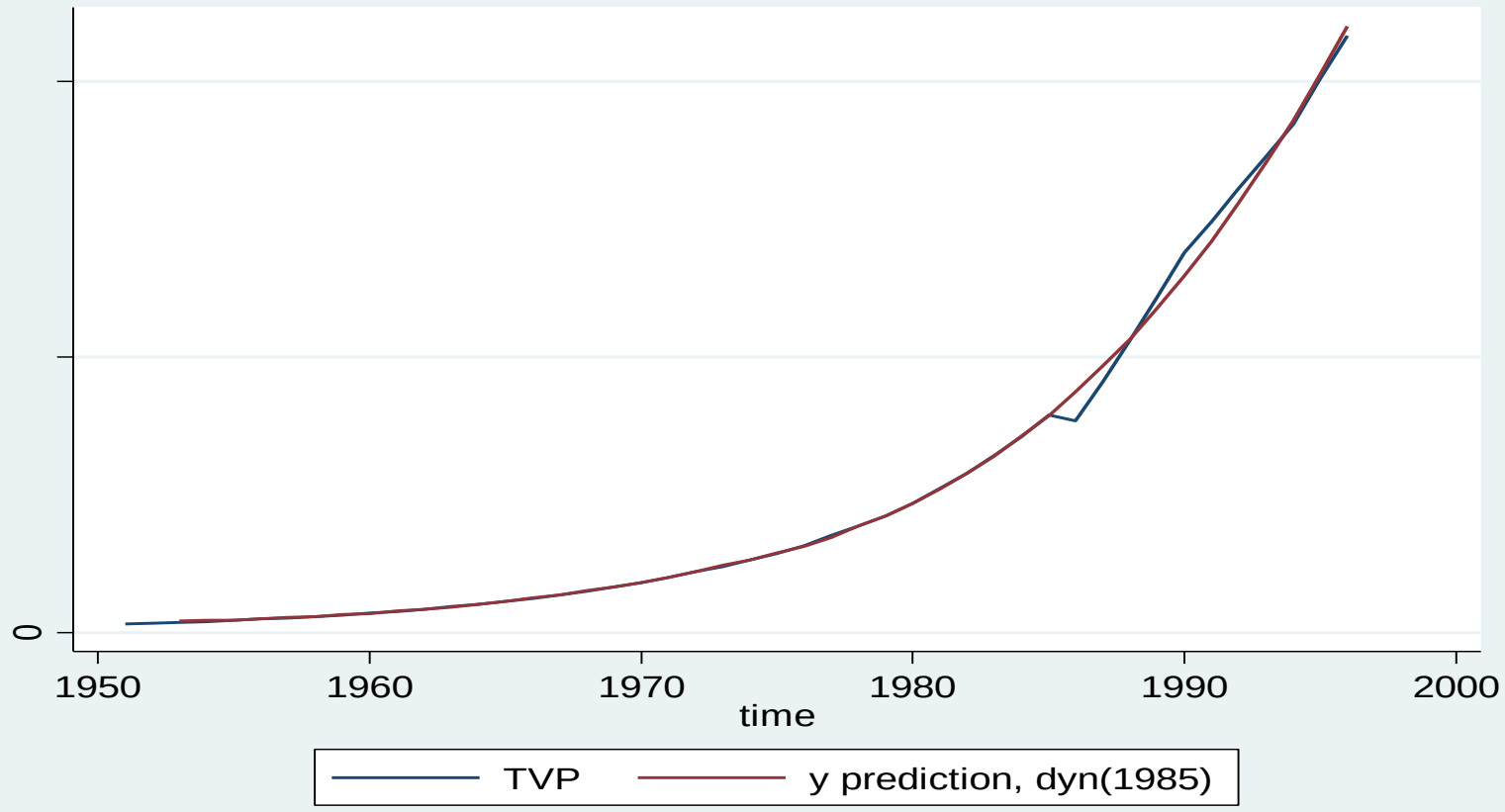
Based on these tests, the following models were found suitable for comparison and analysis. The model-wise test results are given below:

MODEL	P (White Noise Test)	AIC	BIC	RMSE
ARIMA (3,2,3)	0.932	882.41	894.39	2572936
ARIMA (4,2,3)	0.832	889.61	903.072	3674235
ARIMA (5,2,3)	0.906	885.04	900.00	1603122
ARIMA (6,2,3)	0.909	886.82	903.281	1593738
ARIMA (7,2,3)	0.998	882.59	900.55	1374773

On the basis of the above results, ARIMA (7,2,3) has been considered best for analysis.

Time Series Analysis…

The graph showing both the predicted values and the actual values for remaining years is below. The graph above shows the **out-of-sample forecast** for the time period from 1986 to 1996.



Time Series Analysis…

From Time Series Modeling: ARIMA (7,2,3)
Predicted Value for 1996 = 109951968
Actual Value=108336195
Error= 1.491%

Thus we see that the minimum level of error in approximation is observed with time series analysis. The error obtained from this analysis (1.491%) is certainly more acceptable than the same figures from Econometric Analysis (6.202%) and Trend Line Analysis (93.57 %).

Time Series Analysis…

Analysis on AADT data from PeMS, DOT, California:

- Analysis has been done for location “Lark Ellen” on I10(W) in district 7.

Model	P (White Noise)	AIC	BIC	RMSE
ARIMA (1,1,1)	0.703	1523.31	1533.77	948.004
ARIMA (1,1,2)	0.847	1523.86	1536.94	953.260
ARIMA (2,1,2)	0.846	1523.05	1538.74	948.004
ARIMA (2,1,1)	0.825	1524.07	1537.15	956.340

Time Series Analysis…

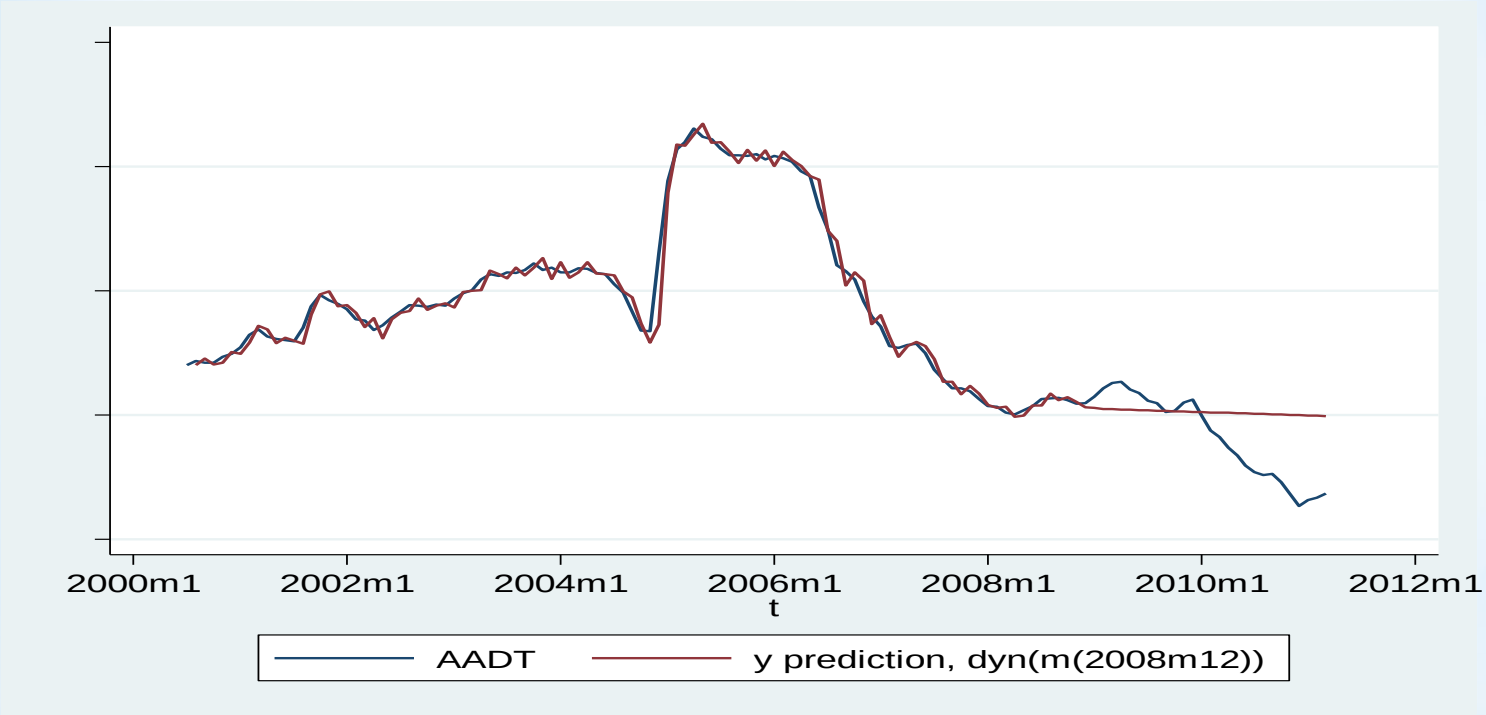
ARIMA regression

Sample: 2000m8 - 2008m12 — Number of obs = 101

Log likelihood = -755.5266 — Wald chi2(4) = 133.68; Prob > chi2 = 0.0000

D.AADT	Coef.	OPG Std. Err.	z	P>|z|	[95% Conf.	Interval]
AADT						
_cons	-11.0944	126.6279	-0.09	0.930	-259.2806	237.0918
ARMA						
ar						
L1.	-.5027332	.215281	-2.34	0.020	-.9246762	-.0807903
L2.	.2845574	.216591	1.31	0.189	-.1399531	.7090679
ma						
L1.	1.308261	.2071327	6.32	0.000	.9022883	1.714234
L2.	.4858782	.1615852	3.01	0.003	.169177	.8025794
/sigma	427.2118	15.14778	28.20	0.000	397.5227	456.9009

Time Series Analysis…



Results for Lark Ellen:
Predicted Value for Mar. 2011= 114959
Actual AADT for Mar. 2011= 111834
Error= 2.794%

Conclusions

- As can be derived from the results obtained in this study, Trend Line Analysis usually results in high overestimation of future traffic volume. This may lead to wasteful use of scarce resources like land and money.

- The study suggests that the use of more logical, dependable and advanced methods of analysis like Econometric Analysis and Time Series Analysis results in more acceptable results.

- Time Series Analysis deserves a special mention. This method has been in use for short term forecasting in fields of finance and econometrics for a long time now and an understanding of its use in transportation engineering must be developed.

- If the limitation of high and rich data requirement for this method is overcome by implementation of proper technology over time, it should contribute favorably towards accurate traffic forecasting in times to come.

Thank You Very Much!

Queries can be directed to:

itskartikatbits@gmail.com
sarkatkar@gmail.com