STATISTICAL ANALYSIS OF INCOME AND VEHICULAR GROWTH IN INDIA

PROJECT WORK SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF M.Sc. IN LIEU OF PAPER MST-403 OF M.Sc.-IV SEMESTER EXAM-2020

DEPARTMENT OF STATISTICS

UNIVERSITY OF RAJASTHAN

JAIPUR

Supervised by: -
Dr. Pankaj Nagar

Dept. Of Statistics
(2019-2020)

Submitted by:-
Divi Joshi
M.Sc. (IV - Sem)
Dept. Of Statistics
(2019-2020)

CERTIFICATE

This is to certify that Miss Divi Joshi, a student of M.Sc. Statistics, Department of Statistics, University of Rajasthan has prepared this survey report based on secondary data collected on “STATISTICAL ANALYSIS OF INCOME AND VEHICULAR GROWTH IN INDIA” under my supervision.

This survey report has been prepared under the course work of M.Sc. Statistics Semester IV of University of Rajasthan.

Dr. Pankaj Nagar

Dept. Of Statistics,

University of Rajasthan

ACKNOWLEDGEMENT

I would like to express my special thanks of gratitude to my professor, Dr. Pankaj Nagar, for his engrossing guidance, propitious assistance, incessant encouragement and kind and gracious patronage during the entire course of investigation and preparation of this manuscript.

I am also grateful to my family, friends and seniors for their constructive suggestions and rendering generous amount of time in finalizing this project within the limited time frame.

With deep reverence, I bow my head before The Almighty for they provide me with energy and path to pursue my dreams.

DIVI JOSHI

M.Sc. Statistics

(IV Semester)

PREFACE

This study has been done to analyse the statistical relationship between vehicular and income growth in India. For this purpose, data was collected regarding GDP per capita, total no. of road accidents and yearly vehicular sale by type. The study can be categorised in the following steps:

(i) Collection of secondary data from various sources.
(ii) Finding out the trend values.
(iii) Performing stationarity test on the data in SPSS.
(iv) Eliminating trend from all data sets.
(v) Performing co-integration test on various models.
(vi) Reaching the conclusion based on regression modelling between the variables.

A positive relationship between personal and overall sales of vehicles in India was found while in case of total no. of road accidents it was observed that it had negative impact on the overall sales for some types of vehicles.

INDEX

LIST OF TABLES

TOPICS **PAGE NO.**

LIST OF FIGURES

TOPICS **PAGE NO.**

INTRODUCTION

India is one among the fastest growing economies in the world with an average growth rate of 8% of Gross Domestic Product in the last decade. The impact of increase in GDP can be observed in many sectors of the economy and transport is not devoid of it. Micro-economic theories firmly established the relationship between income and consumption having direct and positive impact. This can be observed in case of India's per capita income and vehicular growth.

This study makes an attempt to find out if increasing no of road accidents and personal income affect the sales of different vehicles.

The tremendous growth of vehicular population, particularly in the mega cities, is directly influenced by the growth of personalized income. There is a strong relationship among growth in per capita income and total sales of different vehicles with approximately 1.2 billion vehicles in the world, excluding construction vehicles (IOMVM, 2014). The same pattern of growth in per capita income and vehicular growth is observed in India as well.

The growth linkage between GDP and transport sector is very strong in recent decades. During the last decade, the annual average growth rate of road transport sector when compared to GDP stood at 8.8% which was greater than the overall GDP growth of 7.6% between 2001 and 2011. This clearly signifies that the road transport sector is one of the fastest growing sectors in the economy. Further, the contribution of the entire transport sector to the total GDP is 4.8 percent.

BACKGROUND OF TIME SERIES ANALYSIS

Meaning of time series:-

A time series is simply a series of data points ordered in time. This time interval can be daily, weekly, monthly, quarterly, half yearly, yearly, or any other time measure. Time series are influenced by a variety of forces. Some are continuously effected, other make themselves felt at recurring time intervals, and still others are non-recurring or random in nature.

Components of a time series:-

1. **Basic or Secular or Long-time trend** - Basic trend underlines the tendency to grow or decline over a period of years. It is the movement that the series would have taken, had there been no seasonal, cyclical or erratic factors. There are all kinds of trends, two general classes that we may think about are:

- **Deterministic Trends**: These are trends that consistently increase or decrease.
- **Stochastic Trends**: These are trends that increase and decrease inconsistently.

In general, deterministic trends are easier to identify and remove.
In order to identify trend in a series,

- We can plot time series data to see if a trend is obvious or not.
- Create line plots of our data and inspect the plots for obvious trends.
- Add linear and nonlinear trend lines to our plots and see if a trend is obvious.

2. **Seasonal variations** - The two principal factors liable for seasonal changes are the climate or weather and customs. Seasonal variations are measured as a percentage of the trend rather than in absolute quantities. The seasonal index for any month (week, quarter etc.) may be defined as the ratio of the normally expected value (excluding the business cycle and erratic movements) to the corresponding trend value.

3. **Business cycles or cyclical movement** - Because of the persistent tendency for business to prosper, decline, stagnate, recover and prosper again, the third characteristic movement in economic time series is called the business cycle. The business cycle does not recur regularly like seasonal movement, but moves in response to causes which develop intermittently out of complex combinations of economic and other considerations. Its measurement becomes a process of contrast occurrences with a normal estimate arrived at by combining the calculated trend and seasonal movements. The measurement of the variations from normal may be made in terms of actual quantities or it may

be made in such terms as percentage deviations, which is generally more satisfactory method as it places the measure of cyclical tendencies on comparable base throughout the entire period under analysis.

4. **Erratic or Irregular fluctuations** - These movements are exceedingly difficult to dissociate quantitatively from the business cycle. Their causes are such irregular and unpredictable happenings such as wars, droughts, floods, fires, pestilence, fads and fashions which operate as spurs or deterrents upon the progress of the cycle. The common denominator of every random factor it that does not come about as a result of the ordinary operation of the business system and does not recur in any meaningful manner.

Finding trend values:-

A time series may not be affected by all type of variations. Some of these components may affect a few time series, while the other series may be effected by all of them.
In classical time series analysis, it is assumed that any given observation is made up of trend, seasonal, cyclical and irregular movements and these four components have multiplicative or arithmetic relationship i.e.

$O = T \times S \times C \times I$ or $O = T + S + C + I$

Where, O = original data
T = trend
S = seasonal variations
C = to cyclical variations
I = irregular variations.

The additive model is useful when the seasonal variation is relatively constant over time and the multiplicative model is useful when the seasonal variation increases over time.

The methods used here to find trend values are;

- **Ordinary least square method** – a straight line is fitted to the data by the method of least squares. This method is designed to accomplish two results i.e. the sum of the vertical deviations from the straight line must be equal to zero and the sum of the squares of all deviations must be less than the sum of the squares for any other conceivable straight line. The formula for a straight-line trend can most simply be expressed as

$$Y = a + bX \qquad \text{… (i)}$$

Where, X = time variable

Y = dependent variable for which trend values are to be calculated
a,b = constants
n = total no. of observations

To obtain the values of constants a and b by using the Principle of Ordinary Least Squares, we have to solve simultaneously the following two equations.

$$\Sigma Y = na + b\, \Sigma Y \qquad \text{... (ii)}$$

$$\Sigma XY = a\, \Sigma X + b\, \Sigma X^2 \qquad \text{... (iii)}$$

The resulting values are then put in equation (i) and $\hat{Y}$ values are computed for different X values. Finally trend is eliminated by computing another series by the operation (Y -$\hat{Y}$).

- **<u>Moving average method</u>** - It consists in measurement of trend by smoothing out the fluctuations of the data by means of a moving average of extent 'm' by making a series of successive average, where first term is the mean of the first 'm' terms; the second is the mean of m terms from 2nd to (m+1)th terms, third is the mean of m terms from 3rd to (m+2)th terms and so on.

  If m is odd= (2k+1) say, moving average is placed against the mid-value of the time interval it covers, i.e., against t = k+1 and,

  If m is even = (2k) say, it is placed between the two values of the time interval it covers.

  In the latter case, the moving average does not coincide with an original time period and an attempt is made to synchronise the moving average and the original data by centring the moving averages which consists in taking a moving average of extent two, of these moving averages and putting the first of these value against values t = k+1.

  The graph obtained on plotting the moving average value against the corresponding time value gives the trend curve.

  It has been established mathematically that if the fluctuations are regular and periodic then the moving average completely eliminates the oscillatory movement provided:

a) The extent of moving average is exactly equal to (or a multiple of) the period of oscillation, and
b) The trend is linear.

<u>Stationarity test and de-trending:-</u>

A time series with a trend is called non-stationary. An identified trend can be modeled. Once modeled, it can be removed from the time series dataset. This is called *de-trending* the time series. There are two methods to de-trend a given time series:

a) **De-trend by Differencing:** Perhaps the simplest method to de-trend a time series is by differencing. Specifically, a new series is constructed where the value at the current time step is calculated as the difference between the original observation and the observation at the previous time step and then the time series is plotted. This has the effect of removing a trend from a time series dataset. Because no difference value can be created for the first observation (there is nothing for it to be subtracted from), the new dataset contains one less record. This approach works well for data with a linear trend. If the trend is quadratic (the change in the trend also increases or decreases), then a difference of the already-differenced dataset can be taken, a second level of differencing. This process can be further repeated if needed.

b) **De-trend by Model Fitting:** A trend is often easily visualized as a line through the observations. Linear trends can be summarized by a linear model, and nonlinear trends may be best summarized using a polynomial or other curve-fitting method. In addition to being used as a trend identification tool, these fit models can also be used to de-trend a time series. For example, a linear model can be fit on the time index to predict the observation. The predictions from this model will form a straight line that can be taken as the trend line for the dataset. These predictions can also be subtracted from the original time series to provide a de-trended version of the dataset. The residuals from the fit of the model are a de-trended form of the dataset.

A time series with a trend is non-stationary i.e. its statistical properties such as mean, variance and covariance are dependent on time. In this case, it's not going to be possible to estimate any parameters because they are changing according to time. A time series is said to be stationary if its mean and variance are constant over time and the value of covariance between the two time periods depends only on the distance or gap or lag between the two time periods and not the actual time at which the covariance is computed. Such a time series is known as *weakly stationary or covariance stationary*. Stationarity is a requirement for carrying out forecasts and to conduct studies with time series because useful time series theories apply only to stationary variables.

When we have different time series and are interested in finding out if they are stationary or not, we use Augmented Dickey- Fuller (ADF) test which assumes that the original error terms are correlated. The ADF test consists of estimating the following regression:

$$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \alpha_1 \Delta Y_{t-1} + \alpha_2 \Delta Y_{t-2} + \alpha_3 \Delta Y_{t-3} + \alpha_4 \Delta Y_{t-4} + e_t$$

Where Y_t = the variable to be tested for stationarity

α_i ,β_i = constants

e_t = error term

We have the following hypothesis;

H_0: given series is not stationary
H_1: given series is stationary

The τ value of Y_{t-1} coefficient, i.e. τ value of δ, is compared with critical τ value at specified level of significance (here LOS is taken as 5%). If the absolute computed value is greater than the critical value at 5% level of significance then we reject H_0 at 5% LOS.

Co-integration test:-

After performing ADF test, we check if all the series under consideration are integrated of order one, i.e. I(1), or two, i.e. I(2) etc. If they are they integrated of order one or two etc. they can be made stationary after taking first order difference or second order difference of the series.

The regression of a non-stationary time series on another non-stationary time series may produce a spurious regression. However, it is also possible that the two series share the same common trend so that the regression of one on the other will not be necessarily spurious.

To test this, first, a regression model is fitted between the variables under consideration and then their residuals are computed by taking the difference between original and predicted observations. Further, unit root test is applied on the resultant residuals to check for stationarity. If the linear combination comes out to be I(0), i.e. stationary, even though the variables themselves were first order integrated i.e. I(1) or second order integrated i.e. I(2), then the variables are said to be *co-integrated* and their regression is known as *co-integrating regression*

EXPERIMENTAL ANALYSIS

We are basically interested in finding out whether an increase or decrease in GDP per capita (i.e. Personal income) and total no. of road accidents impact the sales of different type of vehicles or not.

For this study, the yearly estimates of GDP of India (in US$) from 2004 – 2016 were taken from World Bank website, estimates of GDP per capita (current US$) for same time period were taken from annual publications of Reserve Bank of India, the government's open portal data.gov.in provided data for total registered vehicles wherein the bifurcation was done on the basis of type and mode of transportation and lastly, the data for total no. of road accidents in India along with information about the share of fatal accidents and no. of persons killed and severely injured in those accidents were taken from the Road Transport Year Book released yearly by the Transport Research Wing, Ministry of Road Transport and Highways.

TABLE -1: Total registered vehicles in India from 2004 – 2016

YEAR	BUSES	TAXIS	CARS	TWO WHEELER	JEEP	total
2004-2005	678521	939738	8072650	58799702	1307926	**69798537**
2005-2006	762341	1039845	9109855	64743126	1376744	**77031911**
2006-2007	1098422	1042347	10146468	69128762	1460364	**82876363**
2007-2008	1156568	1201862	11200142	75336026	1547825	**90442423**
2008-2009	1205793	1307805	12365806	82402105	1638975	**98920484**
2009-2010	176642	3615086	13749406	91597791	1760428	**110899353**
2010-2011	1238245	1789417	15467473	101864582	1974253	**122333970**
2011-2012	1296764	2011022	17569546	115419175	1987098	**138283605**
2012-2013	1418763	2216453	20503389	132550294	2132893	**158821792**
2013-2014	1468010	2109348	21671515	139409778	2216888	**166875539**
2014-2015	1527396	2256619	23807986	154297746	2546731	**184436478**
2015-2016	1384740	2341375	25634824	168975300	2265488	**200601727**
total	**13412205**	**21870917**	**189299060**	**1254524387**	**22215613**	**1501322182**

Source: data.gov.in

Sales for all types of vehicles are observed to be increasing year by year, with the sale of two wheelers being the highest amongst all.

TABLE -2: Registered vehicles per 1000 persons in India from 2004 – 2016

vehicle per 1000 persons						
YEAR	**BUSES**	**TAXIS**	**CARS**	**TWO WHEELER**	**JEEP**	**total**
2004-2005	0.59125	0.81887	7.03432	51.23666	1.13970	**60.82079**
2005-2006	0.65410	0.89220	7.81636	55.55031	1.18126	**66.09422**
2006-2007	0.92834	0.88095	8.57538	58.42479	1.23424	**70.04369**
2007-2008	0.96327	1.00099	9.32825	62.74500	1.28913	**75.32664**
2008-2009	0.99020	1.07397	10.15483	67.66883	1.34593	**81.23376**
2009-2010	0.14311	2.92890	11.13961	74.21145	1.42628	**89.84934**
2010-2011	0.99037	1.43120	12.37112	81.47285	1.57904	**97.84458**
2011-2012	1.02448	1.58876	13.88038	91.18403	1.56986	**109.24750**
2012-2013	1.10768	1.73046	16.00769	103.48651	1.66522	**123.99756**
2013-2014	1.13307	1.62808	16.72696	107.60214	1.71108	**128.80133**
2014-2015	1.16582	1.72241	18.17192	117.77084	1.94384	**140.77483**
2015-2016	1.04547	1.76773	19.35420	127.57575	1.71044	**151.45359**
total	**10.73715**	**17.46452**	**150.56100**	**998.92914**	**17.79602**	**1195.48783**

TABLE -3: GDP and population of India from 2004 – 2016

YEAR	GDP in US$	POPULATION	GDP Per capita (current US$)
2004-2005	820381595512.90	1147609927	714.8610135
2005-2006	940259888792.14	1165486291	806.7532806
2006-2007	1216735441524.86	1183209472	1028.334771
2007-2008	1198895582137.51	1200669765	998.522339
2008-2009	1341886602798.69	1217726215	1101.96084
2009-2010	1675615335600.56	1234281170	1357.563719
2010-2011	1823050405350.42	1250288729	1458.103527
2011-2012	1827637859135.70	1265782790	1443.879529
2012-2013	1856722121394.53	1280846129	1449.605912
2013-2014	2039127446298.55	1295604184	1573.881492
2014-2015	2103587813812.75	1310152403	1605.605431
2015-2016	2290432075123.75	1324509589	1729.268021

Source: databank.worldbank.org

An increasing trend is observed in both GDP and population growth.

TABLE -4: Profile of Road Accidents in India from 2004 – 2016

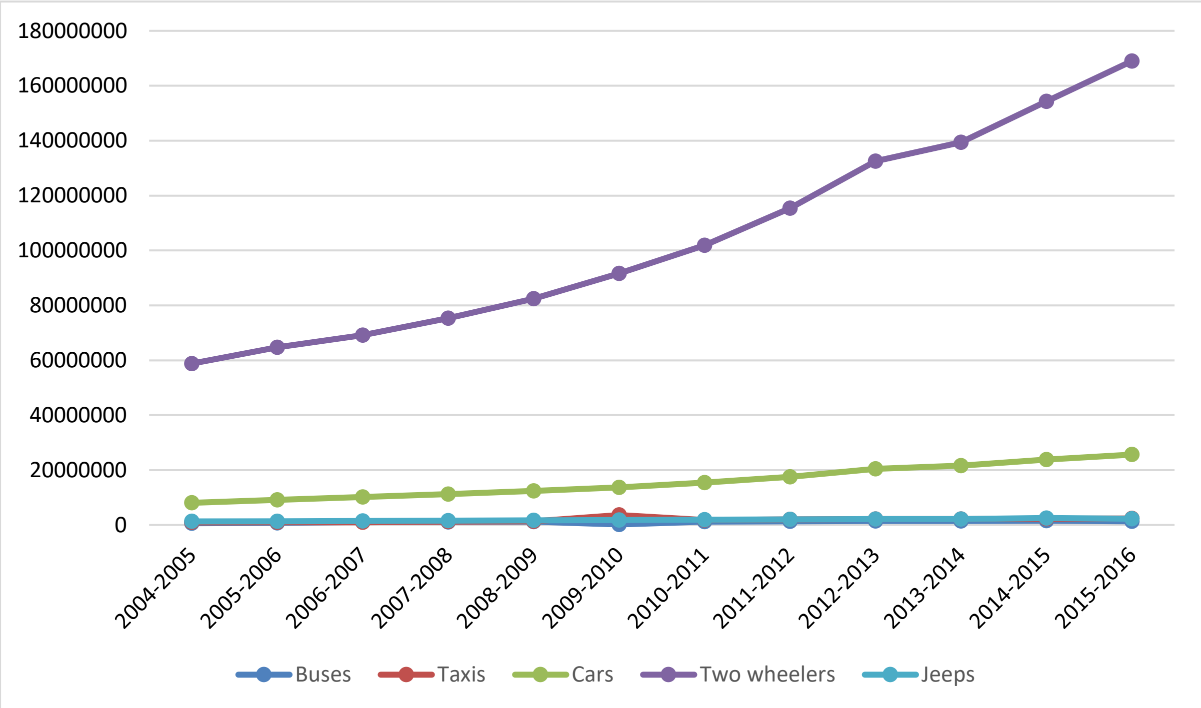
Year	Number of Accidents			Number of Persons		Accident Severity
	Total	Fatal	% share of Fatal accidents in total accidents	Killed	Injured	
2004-2005	4,39,255	83,491	19	94,968	465282	21.6
2005-2006	4,60,920	93,917	20.4	105,749	496,481	22.9
2006-2007	4,79,216	1,01,161	21.1	114,444	513,340	23.9
2007-2008	4,84,704	1,06,591	22	119,860	523,193	24.7
2008-2009	4,86,384	1,10,993	22.8	125,660	515,458	25.8
2009-2010	4,99,628	1,19,558	23.9	134,513	527,512	26.9
2010-2011	4,97,686	1,21,618	24.4	1,42,485	5,11,394	28.6
2011-2012	4,90,383	1,23,093	25.1	1,38,258	5,09,667	28.2
2012-2013	4,86,476	1,22,589	25.2	1,37,572	4,94,893	28.3
2013-2014	4,89,400	1,25,828	25.7	1,39,671	4,93,474	28.5
2014-2015	5,01,423	1,31,726	26.3	1,46,133	5,00,279	29.1
2015-2016	4,80,652	1,36,071	28.3	1,50,785	4,94,624	31.4

Source: Road transport yearbook, Ministry of Road Transport and Highways

Total accidents per year seems to be positively correlated with the number of registered vehicles given the evidence from the data (see table-1).

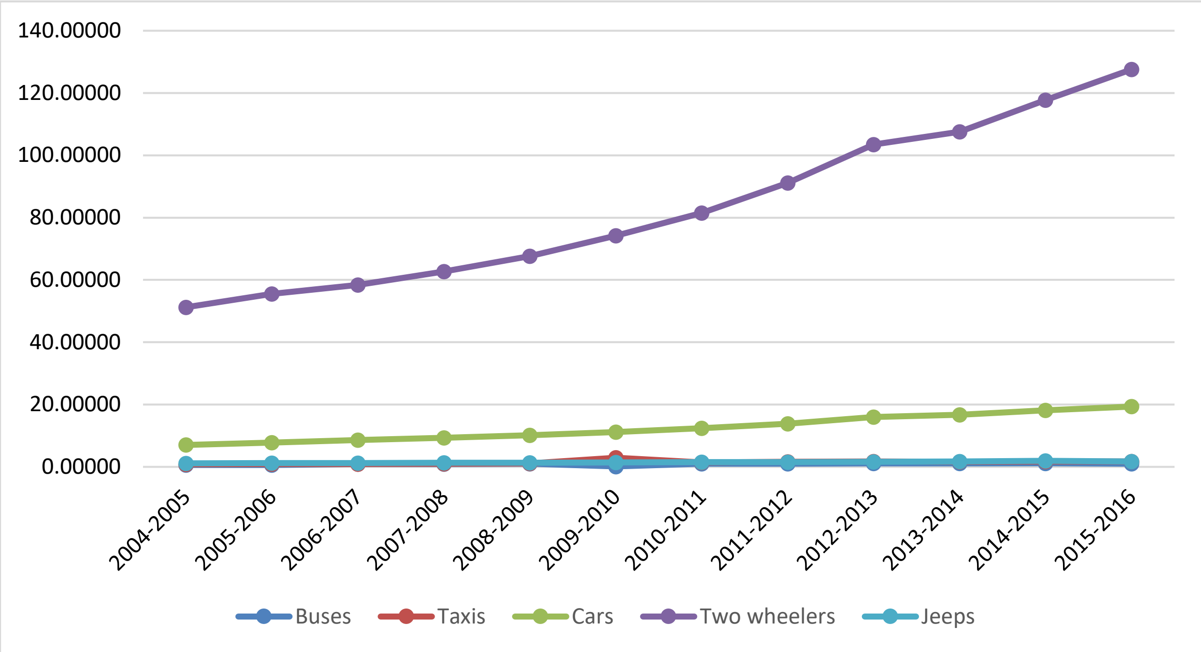
After acquiring the required dataset, time series plots (see graphs 1-7) were plotted to get a picture of the fluctuations present in the data.

GRAPH – 1: Total registered vehicles in India from 2004- 2016



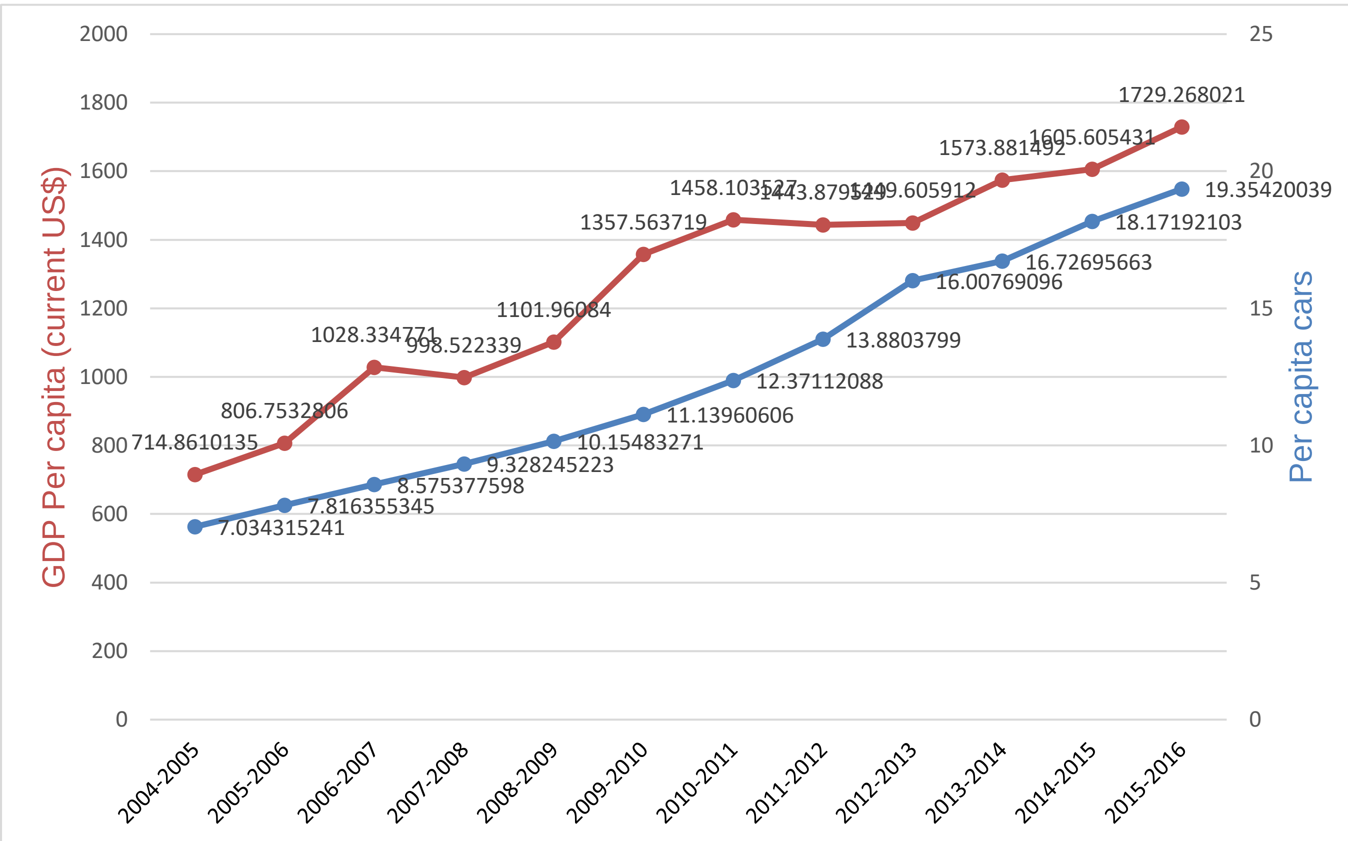
In the above graph we can clearly see that the sale of two wheelers is significantly higher than any other vehicle in India. Also, sales for all types of vehicles are increasing every year.

GRAPH–2: Vehicles per 1000 persons in India from 2004- 2016

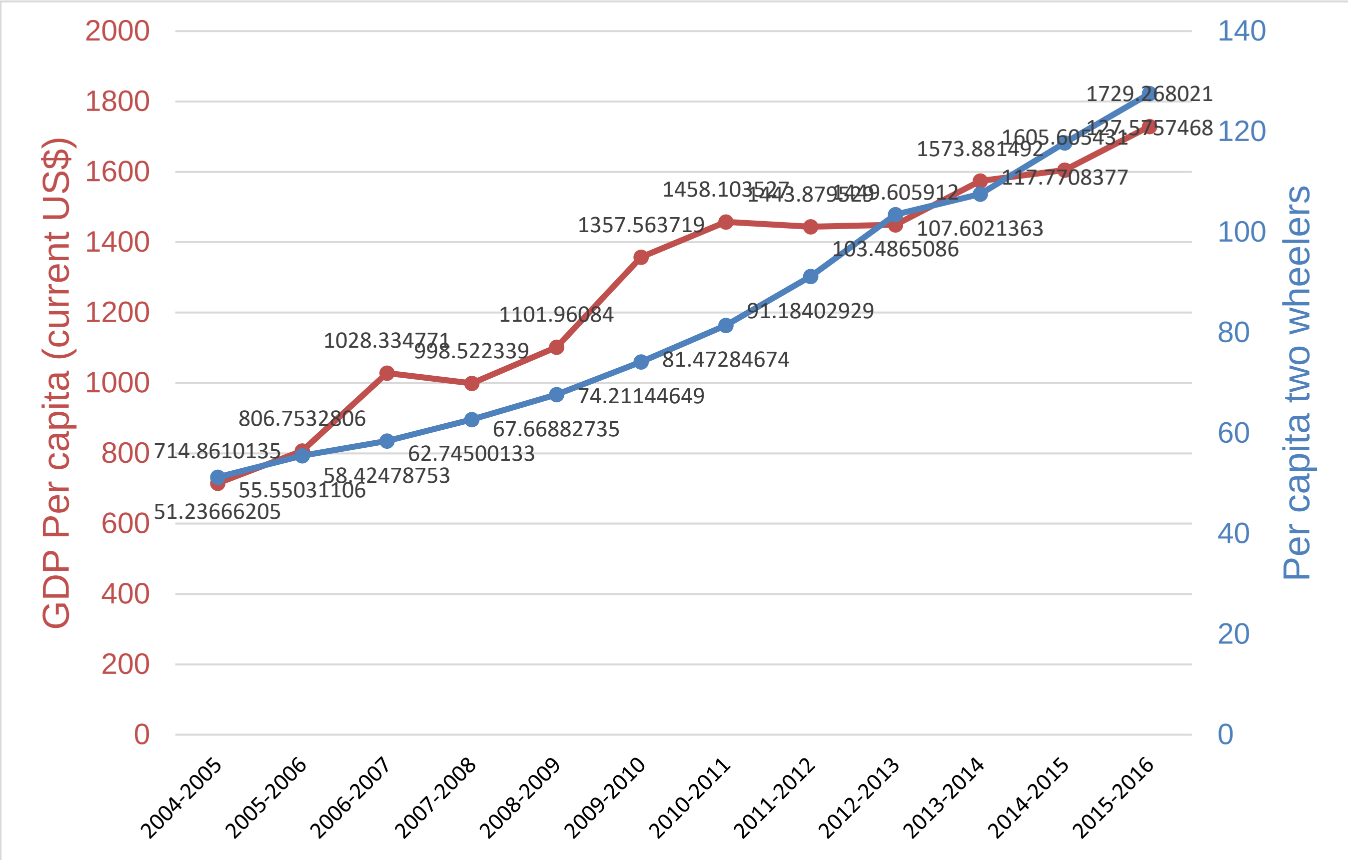


There were 127.5757468 two wheelers, 19.35420039 cars, 1.710435333 jeeps, 1.767729747 taxis and 1.045473745 bus available per 1000 persons in the year 2015-2016 (see table-2).

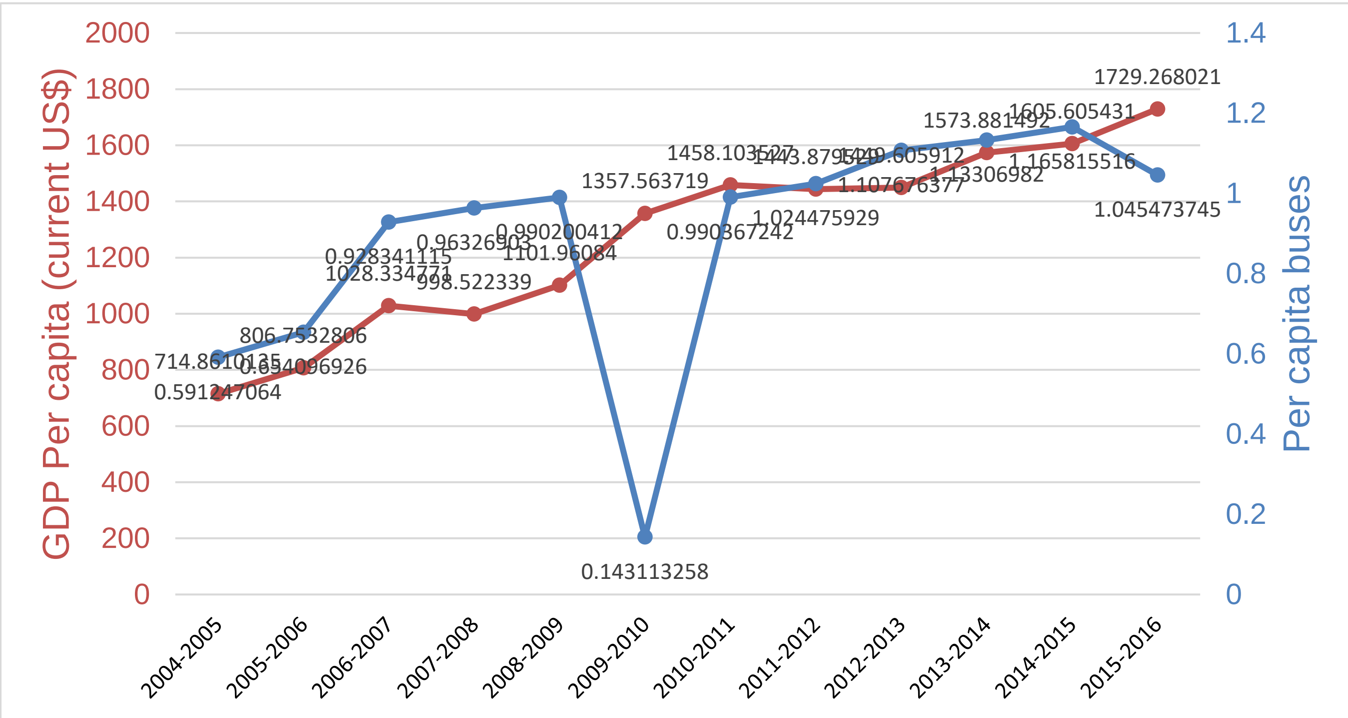
GRAPH–3: GDP Per capita income and per capita cars from 2004-2016



<u>GRAPH–4:</u> GDP Per capita income and per capita two wheelers from 2004-2016

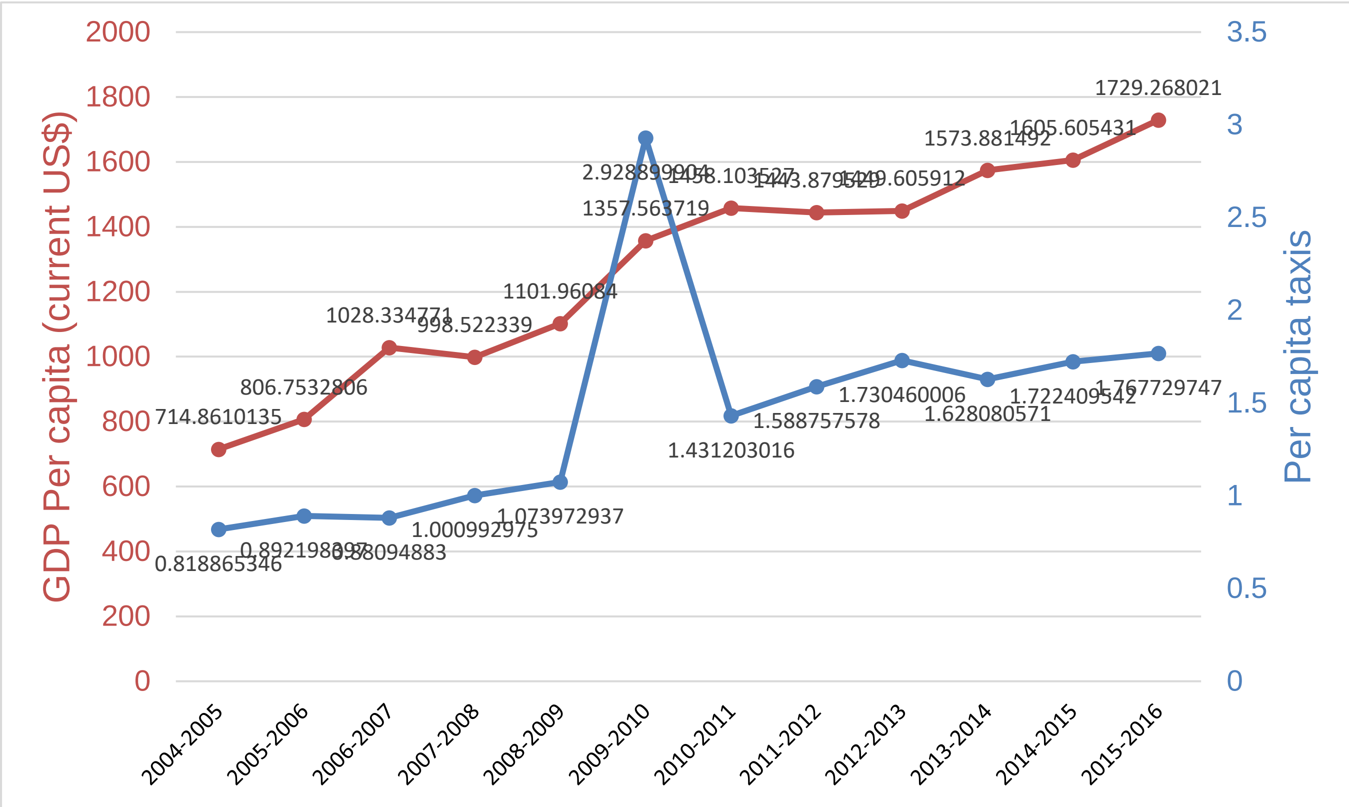


<u>GRAPH–5:</u> GDP Per capita income and per capita buses from 2004-2016



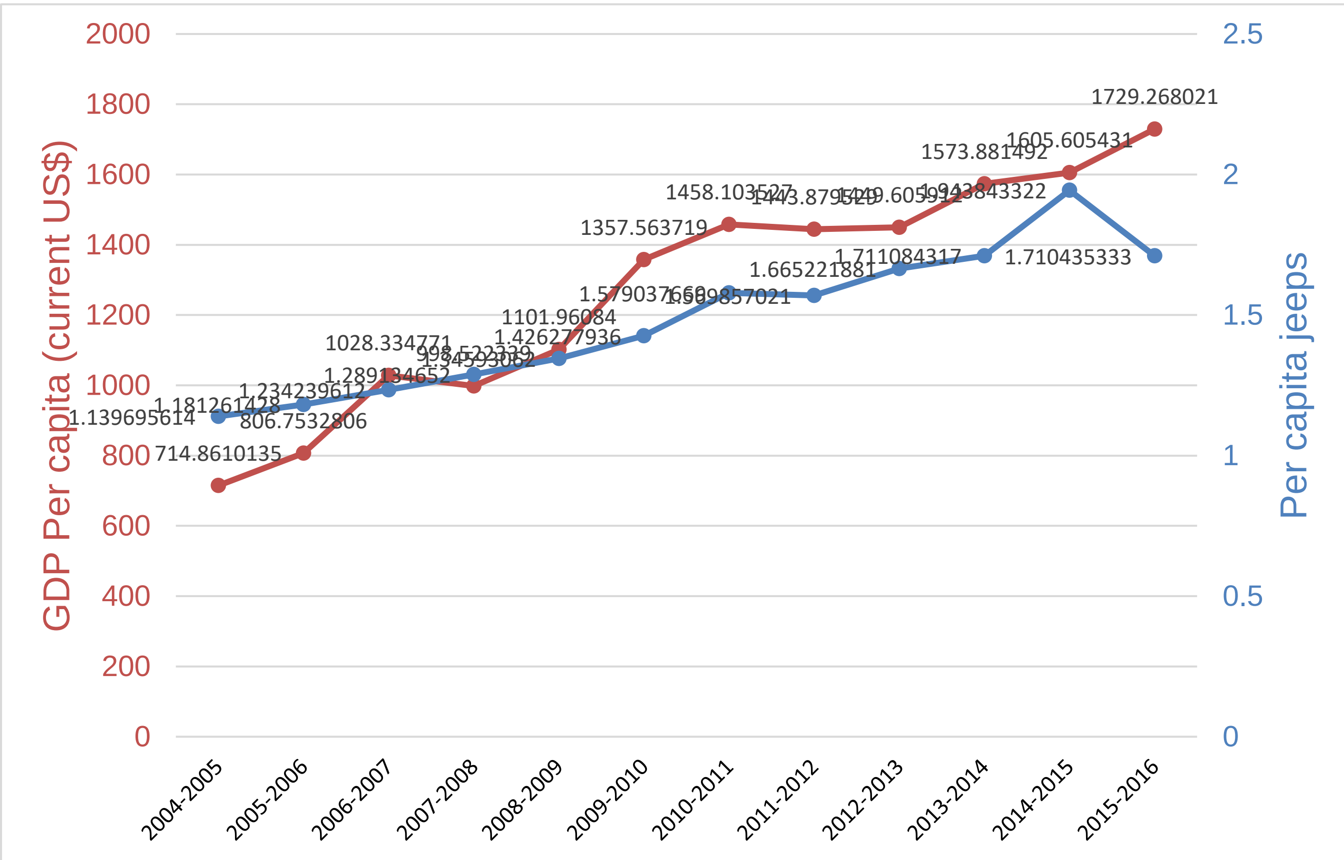
A significant drop is observed in the no. of buses available per person in the year 2009- 2010.

<u>GRAPH–6:</u> GDP Per capita income and per capita taxis from 2004-2016



A significant increase is observed in the no. of taxis available per person in the year 2009- 2010.

<u>GRAPH–7:</u> GDP Per capita income and per capita jeeps from 2004-2016



Graphs 3 – 7 give a very strong intuition that, income and vehicular growth are strongly correlated.

An increasing trend was observed from 2004 till 2016 in all the series because of which an attempt was made to find out the trend values so as to eliminate its effect in further analysis. Trend values were found under additive and multiplicative models both and then were subtracted and divided respectively to de-trend the series. Same treatment was done to eliminate seasonal and cyclic components from the series. After careful comparison between both the models, the multiplicative model was used finally for further analysis (see table – 5).

The ordinary least square and moving average method were then used to calculate trend values. All the fitted trend curves can be seen in graphs 8 - 14 and the trend values so obtained are given in table – 6.

TABLE -5: Values obtained by taking natural log on original values (see table 1, 2, 3) in order to use multiplicative model.

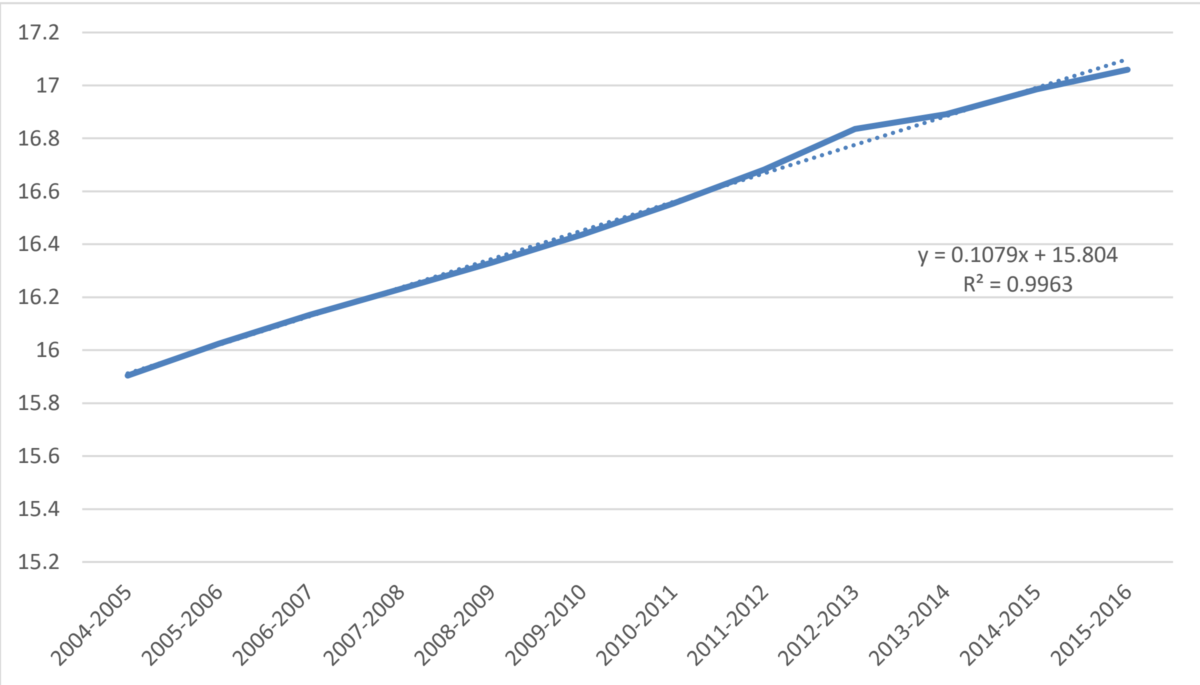
YEAR	BUSES	TAXIS	CARS	TWO WHEELER	JEEP	ROAD ACCIDENTS	GDP Per capita (current US$)
2004-2005	13.428	13.753	15.904	17.890	14.084	12.993	6.572
2005-2006	13.544	13.855	16.025	17.986	14.135	13.041	6.693
2006-2007	13.909	13.857	16.133	18.051	14.194	13.080	6.936
2007-2008	13.961	13.999	16.231	18.137	14.252	13.091	6.906
2008-2009	14.003	14.084	16.330	18.227	14.310	13.095	7.005
2009-2010	12.082	15.101	16.437	18.333	14.381	13.122	7.213
2010-2011	14.029	14.397	16.554	18.439	14.496	13.118	7.285
2011-2012	14.075	14.514	16.682	18.564	14.502	13.103	7.275
2012-2013	14.165	14.611	16.836	18.702	14.573	13.095	7.279
2013-2014	14.199	14.562	16.892	18.753	14.612	13.101	7.361
2014-2015	14.239	14.629	16.986	18.854	14.750	13.125	7.381
2015-2016	14.141	14.666	17.059	18.945	14.633	13.083	7.455

<u>TABLE -6</u>: Values obtained from the fitted trend curves (see graphs 8-14)

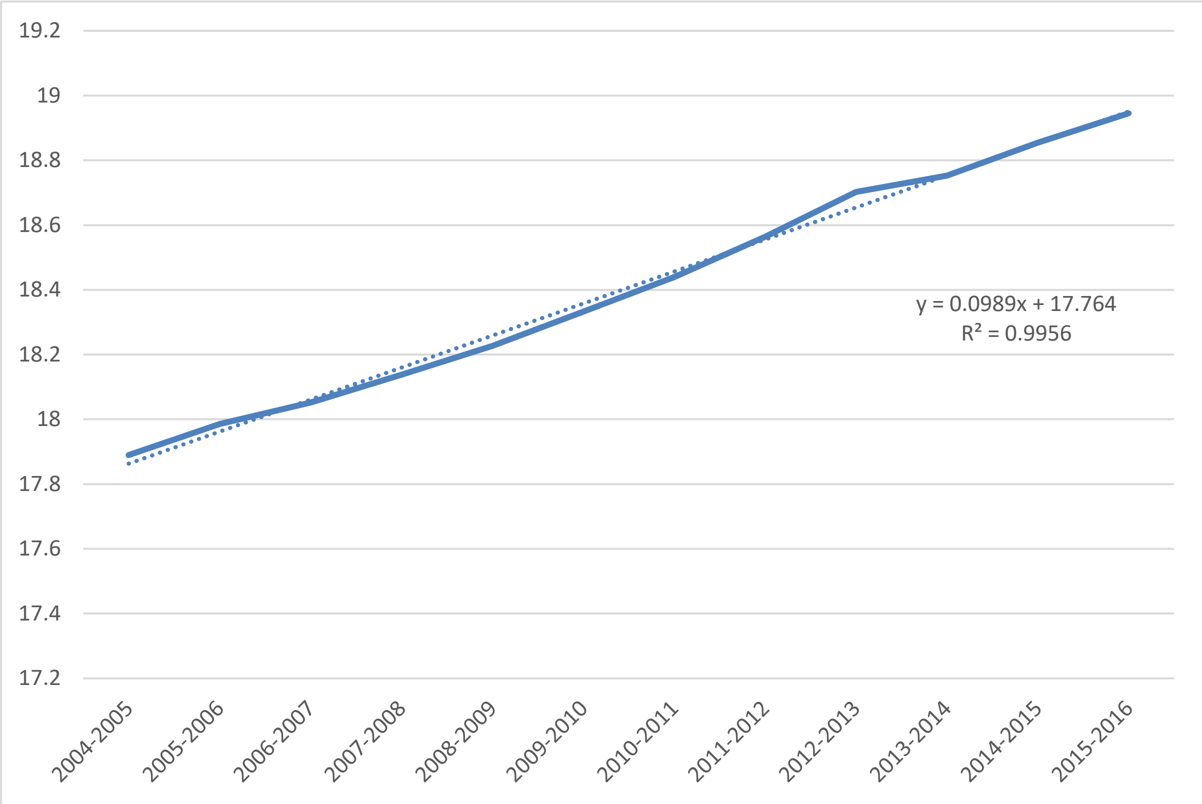
trend values							
YEAR	**BUSES**	**TAXIS**	**CARS**	**TWO WHEELER**	**JEEP**	**ROAD ACCIDENTS**	**GDP Per capita (current US$)**
2004	#N/A	#N/A	15.91218	17.86306	14.08721	13.04996	6.69745
2005	13.48591	13.80397	16.02009	17.96194	14.14593	13.05673	6.77310
2006	13.72677	13.85578	16.12801	18.06082	14.20466	13.06349	6.84875
2007	13.93518	13.92818	16.23592	18.15970	14.26339	13.07026	6.92440
2008	13.98181	14.04162	16.34383	18.25858	14.32212	13.07702	7.00006
2009	13.04226	14.59224	16.45175	18.35747	14.38084	13.08379	7.07571
2010	13.05554	14.74901	16.55966	18.45635	14.43957	13.09055	7.15136
2011	14.05229	14.45578	16.66757	18.55523	14.49830	13.09732	7.22701
2012	14.12034	14.56279	16.77548	18.65411	14.55703	13.10408	7.30266
2013	14.18236	14.58665	16.88340	18.75299	14.61576	13.11085	7.37832
2014	14.21925	14.59563	16.99131	18.85187	14.67448	13.11761	7.45397
2015	14.19005	14.64781	17.09922	18.95075	14.73321	13.12438	7.52962

We have the following graphs representing the fitted trend line;

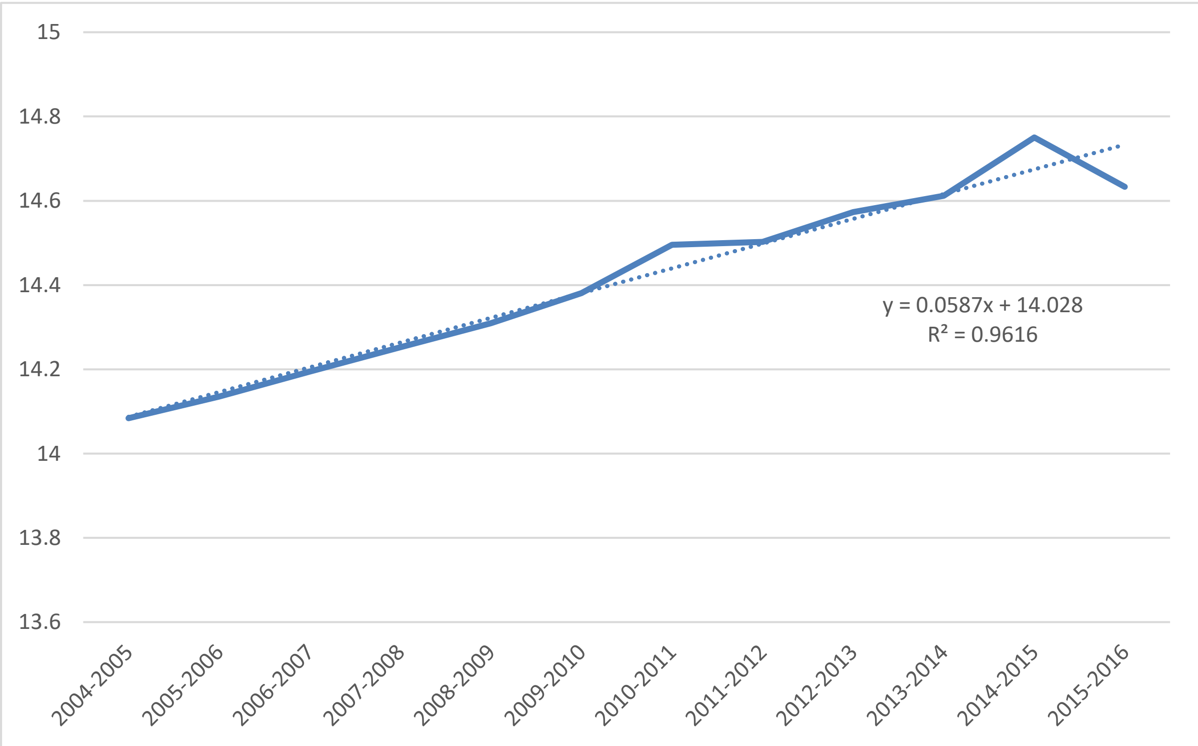
<u>GRAPH–8:</u> Fitted trend line by OLS method for computing trend values of Cars



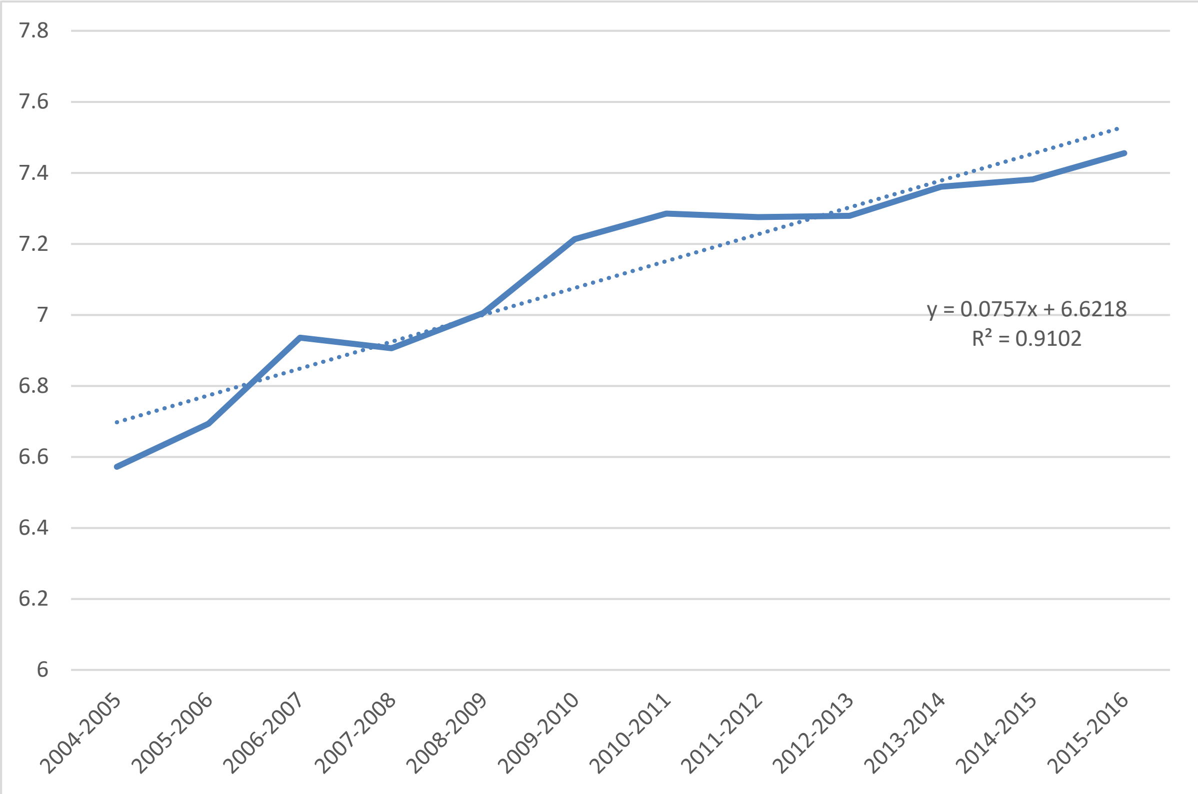
<u>GRAPH–9:</u> Fitted trend line by OLS method for computing trend values of Two wheelers



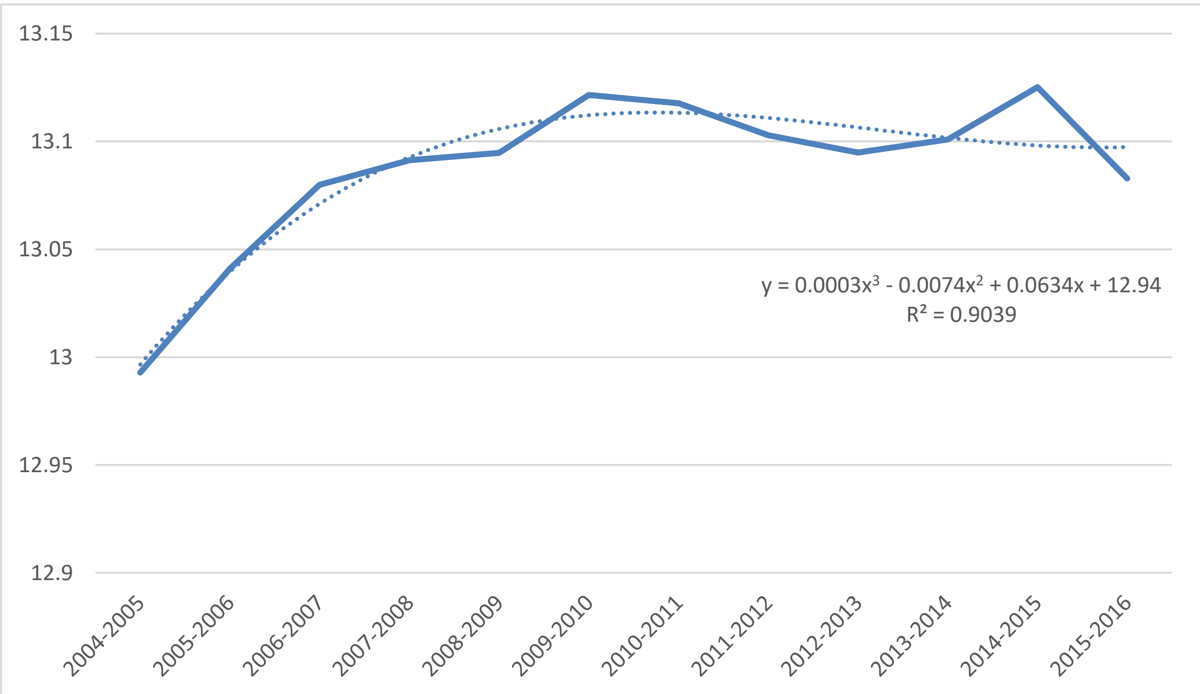
<u>GRAPH–10:</u> Fitted trend line by OLS method for computing trend values of Jeep



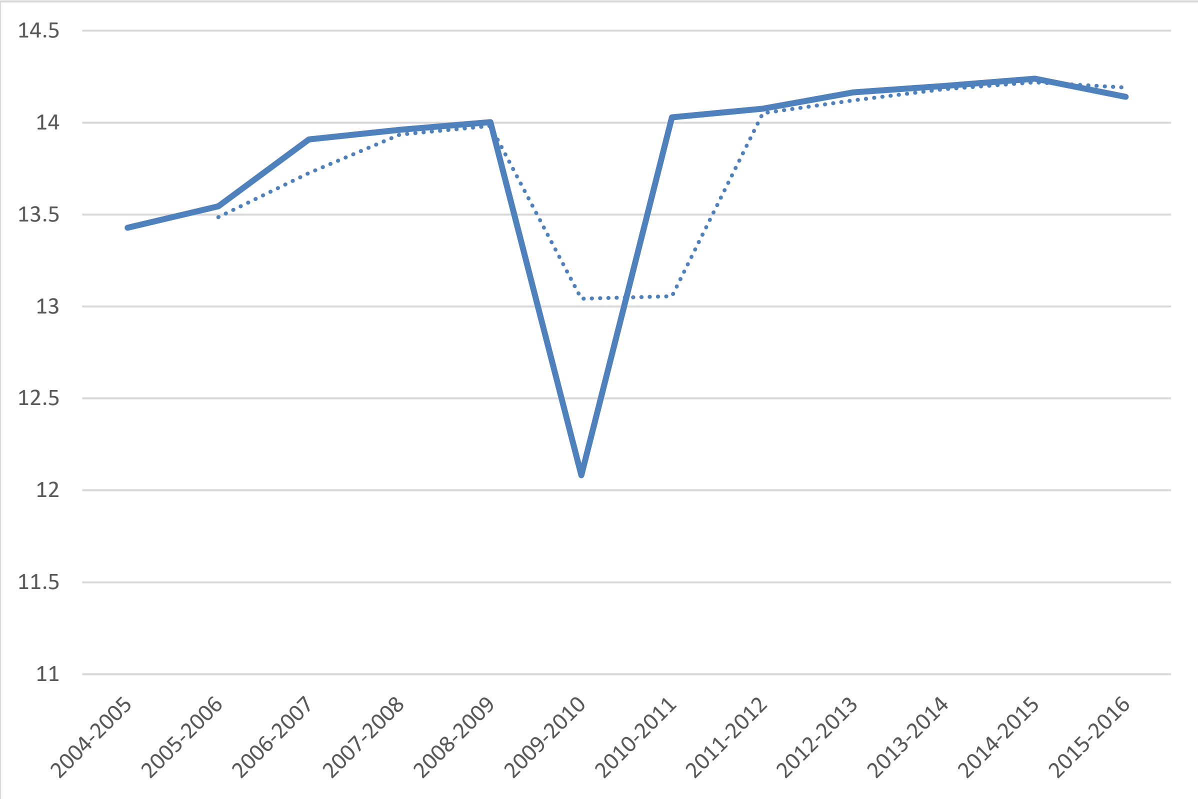
<u>GRAPH–11:</u> Fitted trend line by OLS method for computing trend values of GDP Per capita (current US$)



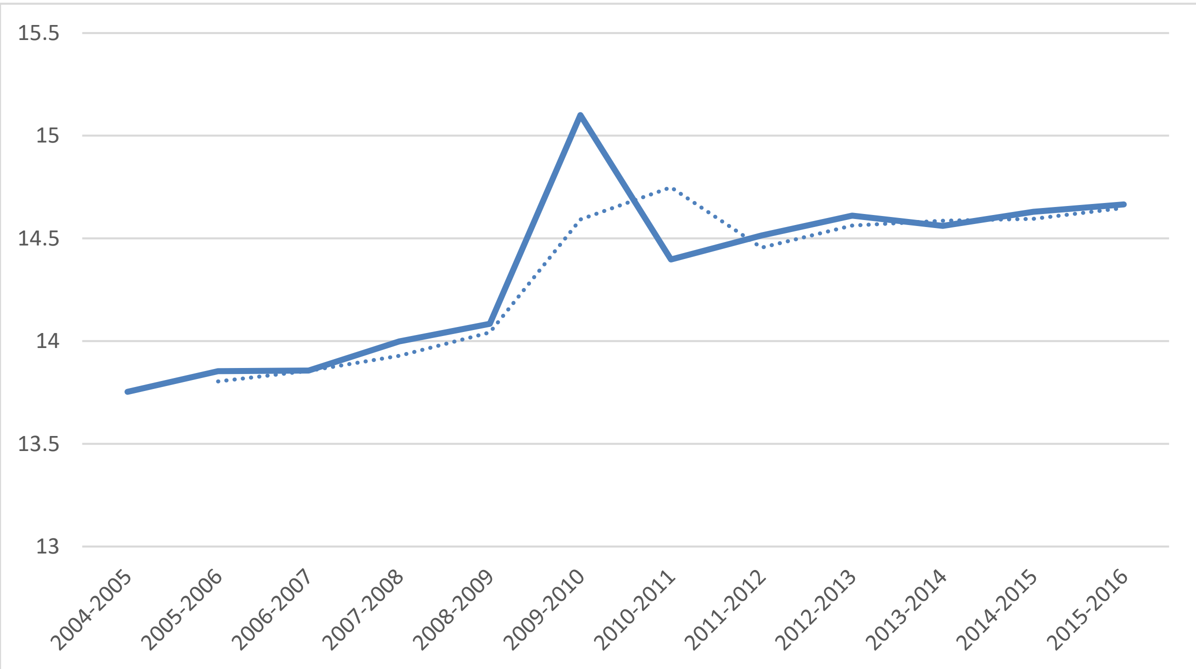
<u>GRAPH–12:</u> Fitted trend line by OLS method for computing trend values of Road accidents



GRAPH–13: Fitted trend line by two period moving average method for computing trend values of Bus



GRAPH–14: Fitted trend line by two period moving average method for computing trend values of Taxis



Now, in order to find out if the different time series involved in this study are stationary or not, we applied the Augmented Dickey- Fuller (ADF) test in SPSS and the results are given in table – 7, the hypothesis to be tested here is;

H_0: given series is not stationary
H_1: given series is stationary

TABLE -7: Results obtained after applying ADF test.

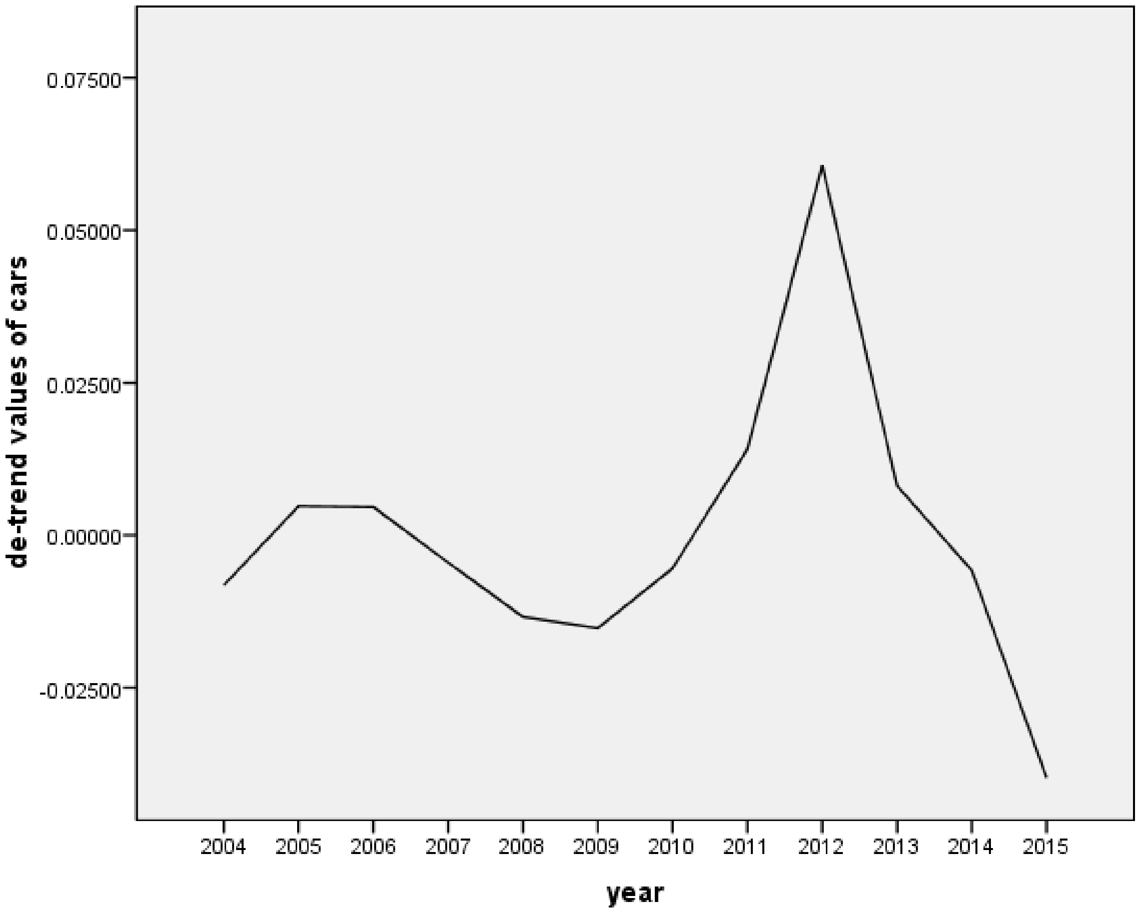
VARIABLE (Y_t)	τ COMPUTED	CRITICAL τ AT 1%	CRITICAL τ AT 5%	CONCLUSION
CAR	-1.2	-4.38	-3.60	Non stationary
TWO WHEELER	-1.853	-4.38	-3.60	Non stationary
JEEP	-1.666	-4.38	-3.60	Non stationary
BUS	-0.317	-4.38	-3.60	Non stationary
TAXI	-9.211	-4.38	-3.60	Stationary
GDP PER CAPITA	-2.481	-4.38	-3.60	Non stationary
ROAD ACCIDENTS	-1.356	-4.38	-3.60	Non stationary

As the table shows, all the variables are non-stationary except for the variable 'taxi' and therefore an attempt is made to make them stationary by model fitting de-trending method i.e. by eliminating fitted trend values from the original series so as to eliminate trend variation and find out error in prediction. A series of these error values, upon testing, turned out to be stationary with zero mean and a constant variance (see table – 8 and graphs 15 – 21). Therefore, all the series under study are now stationary.

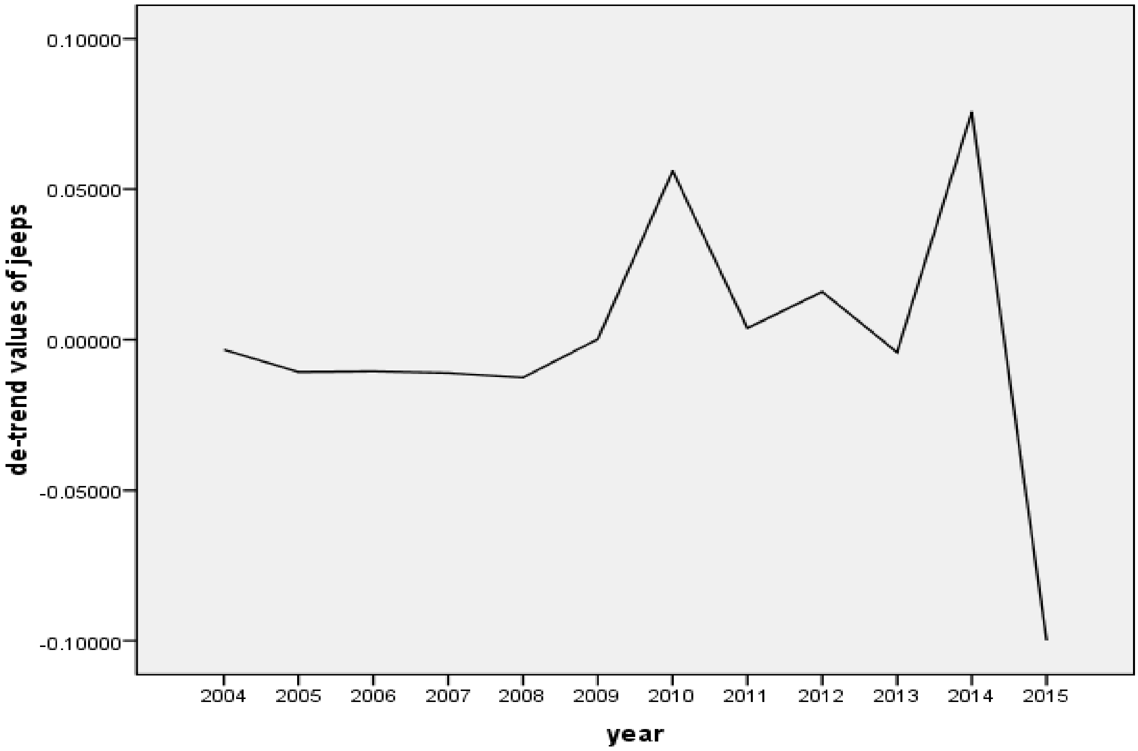
TABLE -8: Zero Mean and constant variance providing evidence that variables are now stationary.

VARIABLE (Y_t)	MEAN	VARIANCE
CAR	.00000000	.01072302
TWO WHEELER	.00000000	.01063320
JEEP	-1.22205073E-16	.01798852
BUS	-3.41790088E-15	.77440411
TAXI	-4.01068068E-15	.04406111
GDP PER CAPITA	.00000	.01622
ROAD ACCIDENTS	-7.90191325E-15	.00998810

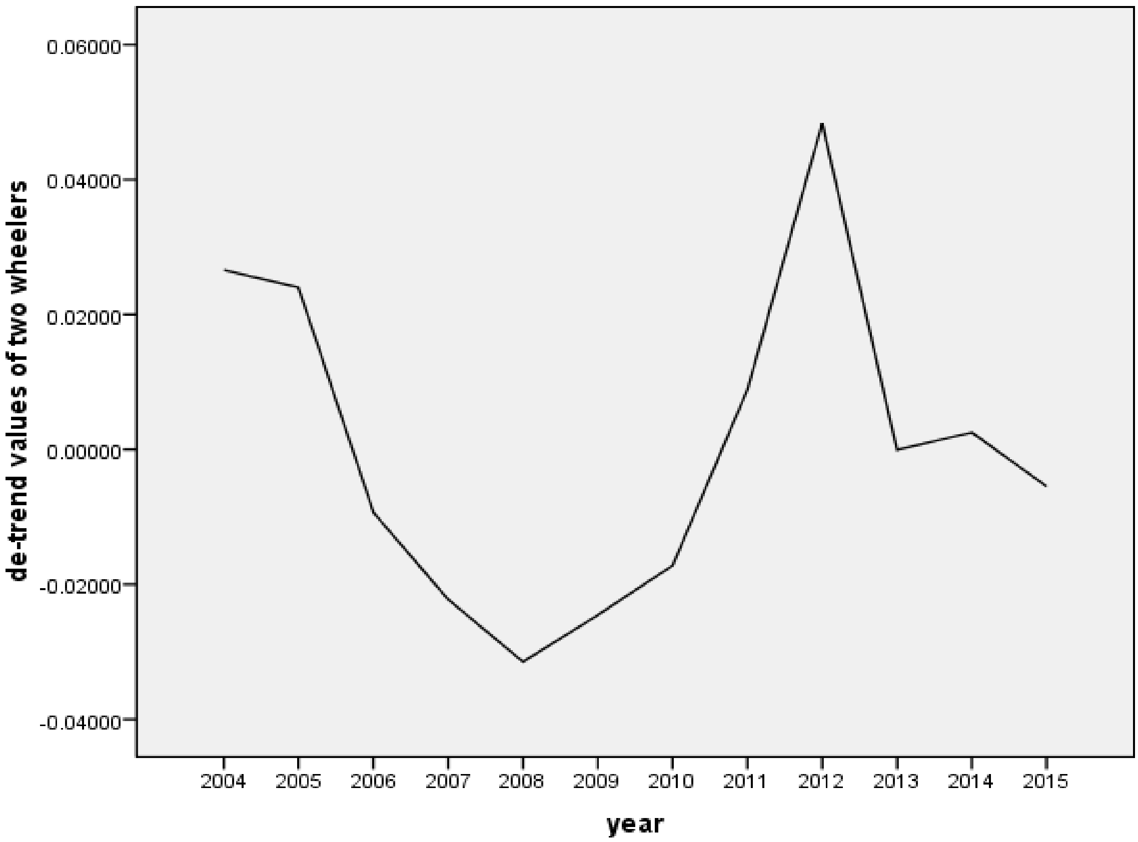
GRAPH–15: sequence chart supporting the conclusion that the de-trend values of variable cars is stationary



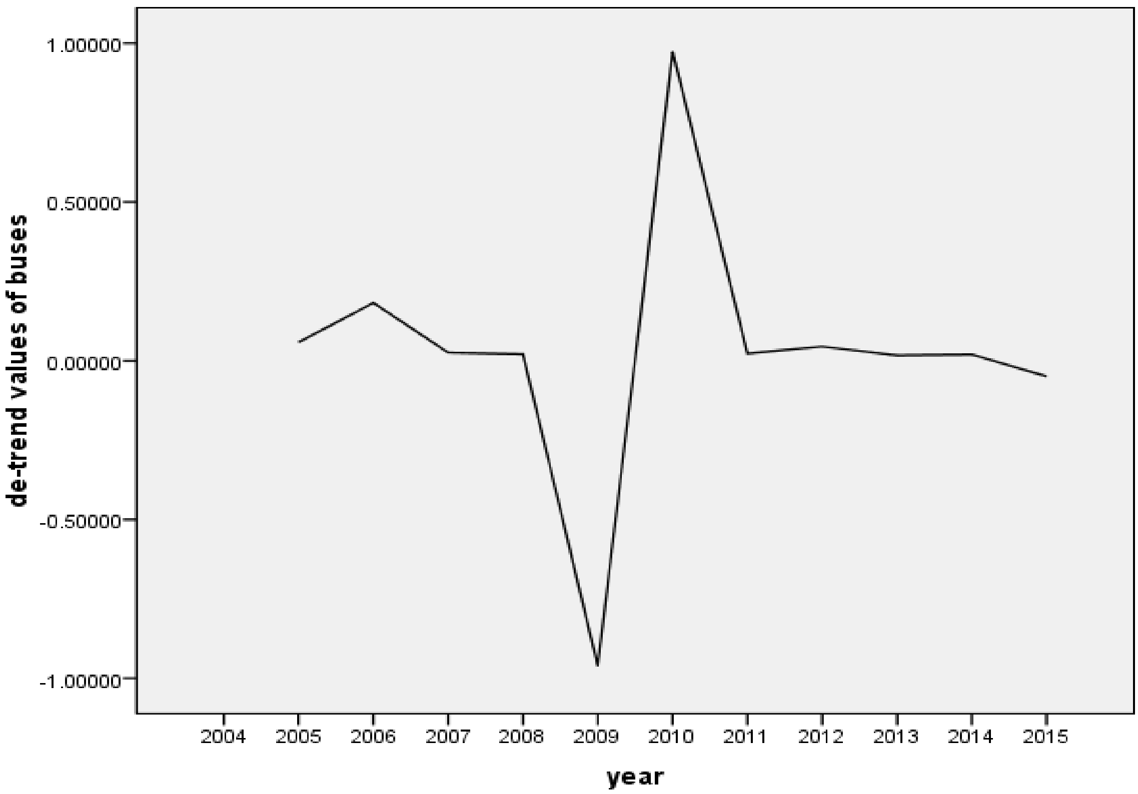
<u>GRAPH–16:</u> sequence chart supporting the conclusion that the de-trend values of variable jeeps is stationary



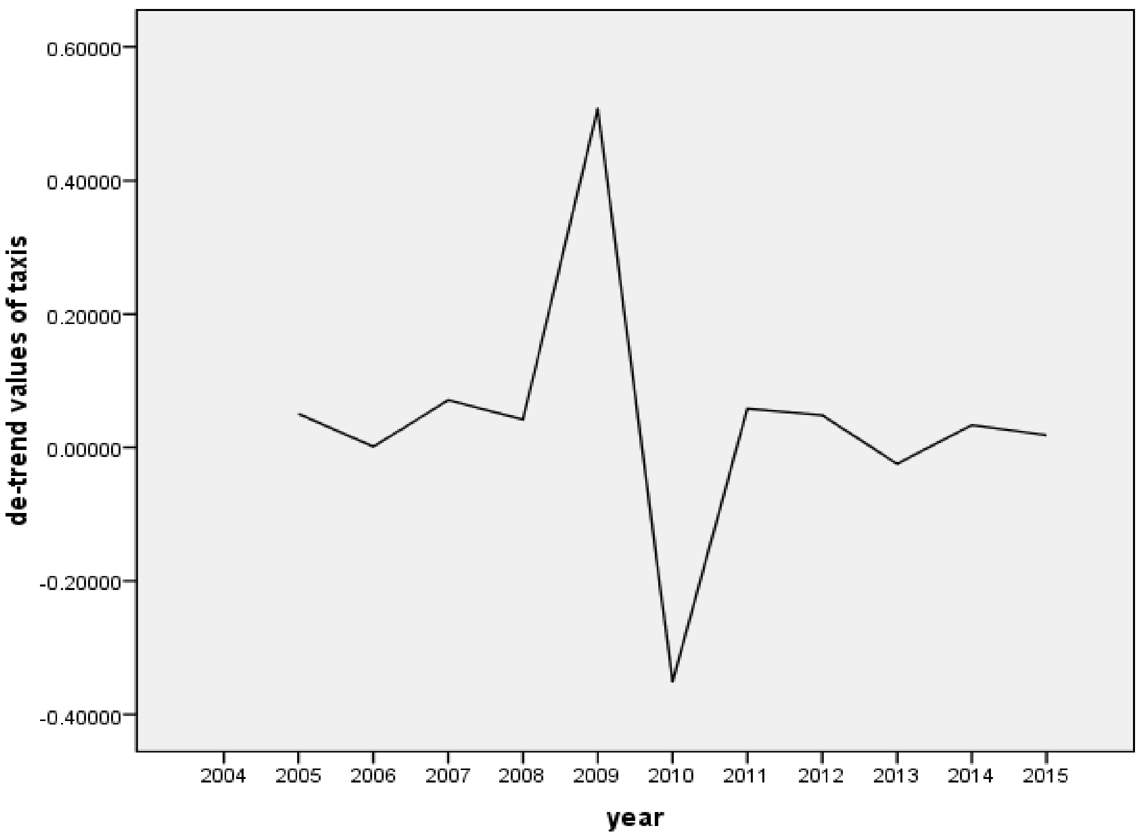
<u>GRAPH–17:</u> sequence chart supporting the conclusion that the de-trend values of variable two wheelers is stationary



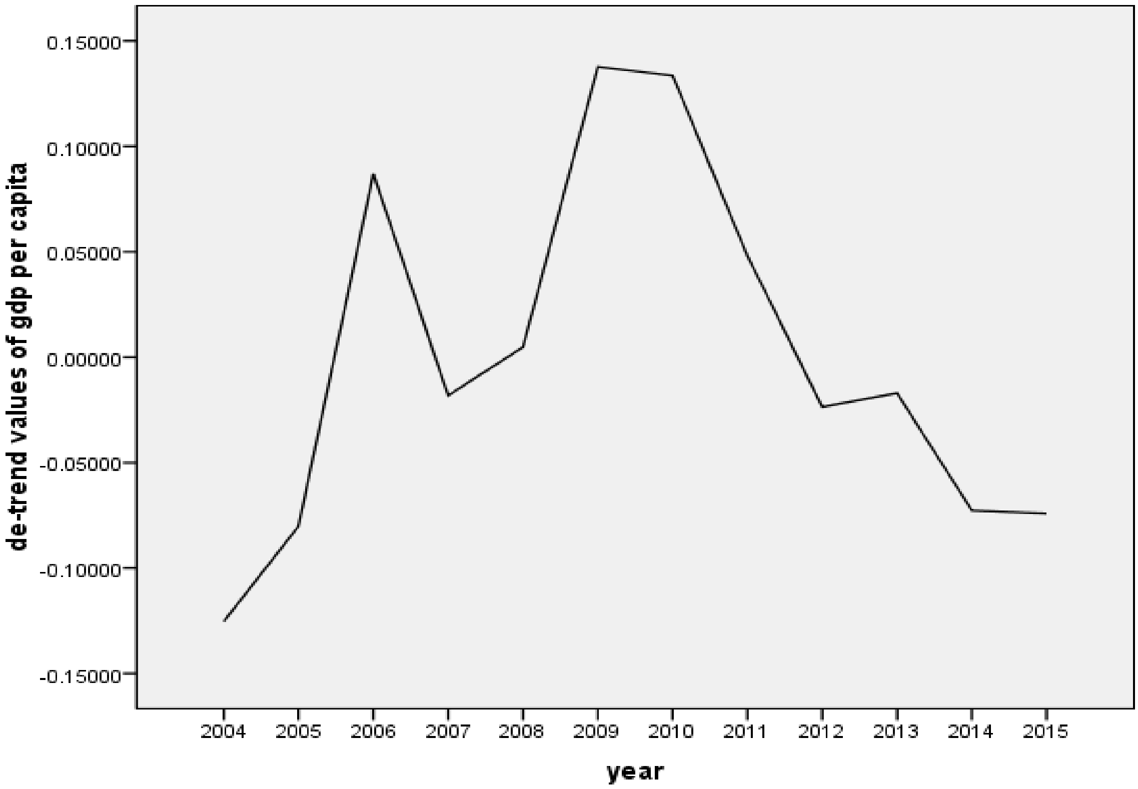
<u>GRAPH–18:</u> sequence chart supporting the conclusion that the de-trend values of variable buses is stationary



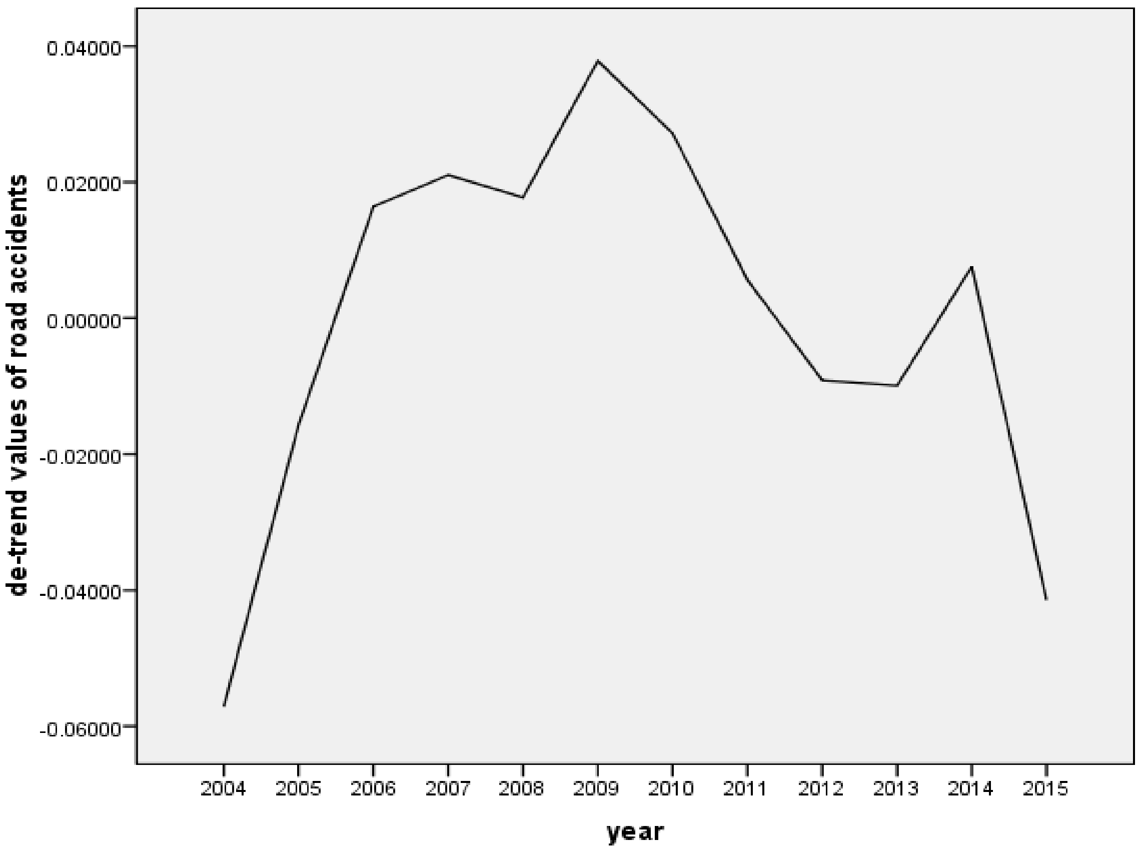
<u>GRAPH–19:</u> sequence chart supporting the conclusion that the de-trend values of variable taxis is stationary



<u>GRAPH–20:</u> sequence chart supporting the conclusion that the de-trend values of variable GDP per capita is stationary



<u>GRAPH–21:</u> sequence chart supporting the conclusion that the de-trend values of variable road accidents is stationary



Since, all of the variables were observed to be first order integrated, it was quite possible that a regression between them could lead to a spurious regression. To test the same, co-integration test was conducted in SPSS taking into consideration the following models;

1. $Ln(Y_{1t}) = \beta_{01} + \beta_{11} Ln(X_{1t}) + \beta_{21} Ln(X_{2t}) + u_{t1}$

2. $Ln(Y_{2t}) = \beta_{02} + \beta_{12} Ln(X_{1t}) + \beta_{22} Ln(X_{2t}) + u_{t2}$

3. $Ln(Y_{3t}) = \beta_{03} + \beta_{13} Ln(X_{1t}) + \beta_{23} Ln(X_{2t}) + u_{t3}$

4. $Ln(Y_{4t}) = \beta_{04} + \beta_{14} Ln(X_{1t}) + \beta_{24} Ln(X_{2t}) + u_{t4}$

5. $Ln(Y_{5t}) = \beta_{05} + \beta_{15} Ln(X_{1t}) + \beta_{25} Ln(X_{2t}) + u_{t5}$

Where; Y_{1t} = total no. of cars
Y_{2t} = total no. of buses
Y_{3t} = total no. of taxis
Y_{4t} = total no. of jeeps
Y_{5t} = total no. of two wheelers
X_{1t} = GDP per capita (current US$)
X_{2t} = total road accidents
Ln = natural log
u_{ti} = error term; i= 1 to 5
t = year (from 2004 to 2015)

1. $Ln(Y_{1t}) = \beta_{01} + \beta_{11} Ln(X_{1t}) + \beta_{21} Ln(X_{2t}) + u_{t1}$

We get the following fitted model;

$$\widehat{Ln(Y_{1t})} = 50.061 + 1.665\, Ln(X_{1t}) - 3.469\, Ln(X_{2t}) \quad \ldots (*)$$

Having,	t	=	(2.930)	(9.240)	(-2.507)
	p- value	=	(0.017)	(0.000)	(0.033)
	R^2	=	0.946		
	d	=	1.983		

Where, t is the computed t statistic, R^2 is coefficient of determination and d is the Durbin Watson d statistic.

Then we subject u_{t1} to unit root analysis to find if it is stationary, i.e. I (0), where u_{t1} is given by, $\hat{u}_{t1} = Ln(Y_{1t}) - \widehat{Ln(Y_{1t})}$

Through unit root analysis we get;

$$\widehat{\Delta\hat{u}_{t1}} = -1.001\hat{u}_{t1-1}$$

Having,	t	=	(-3.163)
	p- value	=	(0.010)
	R^2	=	0.500
	d	=	1.915

Since the p- value 0.010 < 0.05, the unit root test shows that the residuals are stationary at 5% LOS.

In this case we saw that although Ln (Y_{1t}), Ln (X_{1t}) and Ln (X_{2t}) were I(1) individually, their linear combination is I(0). Therefore the variables Ln (Y_{1t}), Ln (X_{1t}) and Ln (X_{2t}) are co-integrated.

From the regression model (*), we find that;

- The value of R^2 is fairly high depicting that the model is of a good fit.
- The Durbin Watson statistic *d* is greater than d_u = 1.579 (at 5% LOS) so we can say that there is no autocorrelation present in the model.
- Also the p-values for both the variables, Ln (X_{1t}) and Ln (X_{2t}), are less than 0.05 suggesting that the resulting coefficients are significantly different than zero.

2. Ln (Y_{2t}) = β_{02} + β_{12} Ln (X_{1t}) + β_{22} Ln (X_{2t}) + u_{t2}

We have the following fitted model;

$$\widehat{\mathbf{Ln}\,(\mathbf{Y_{2t}})} = \mathbf{86.554 + 1.258\ Ln\ (X_{1t}) - 6.242\ Ln\ (X_{2t})} \qquad ...(**)$$

Having,	t	=	(0.823)	(1.135)	(-0.733)
	p- value	=	(0.432)	(0.286)	(0.482)
	R^2	=	0.534		
	d	=	2.287		

Through unit root analysis we get;

$$\widehat{\Delta\hat{u}_{t2}} = -1.159\hat{u}_{t2-1}$$

Having,	t	=	(-3.753)
	p- value	=	(0.004)
	R^2	=	0.585
	d	=	2.095

Since the p- value 0.004 < 0.05, the unit root test shows that the residuals are stationary at 5% LOS. Therefore the variables Ln (Y_{2t}), Ln (X_{1t}) and Ln (X_{2t}) are co-integrated.

From the regression model (**), we find that;

- The value of R^2 is more than average depicting that the model is of a good fit.
- The Durbin Watson statistic *d* is greater than d_u = 1.579 (at 5% LOS) so we can say that there is no autocorrelation present in the model.
- Also the p-values for both the variables, Ln (X_{1t}) and Ln (X_{2t}), are more than 0.05 suggesting that the resulting coefficients are not significantly different than zero.

3. Ln (Y_{3t}) = β_{03} + β_{13} Ln (X_{1t}) + β_{23} Ln (X_{2t}) + u_{t3}

We have the following fitted model;

$$\widehat{\mathbf{Ln}\,(\mathbf{Y_{3t}})} = \mathbf{-9.779 + 1.102\ Ln\ (X_{1t}) + 1.244\ Ln\ (X_{2t})} \qquad \dots (***)$$

Having,	t	=	(-0.234)	(2.497)	(0.367)
	p- value	=	(0.821)	(0.034)	(0.722)
	R^2	=	0.717		
	d	=	2.498		

Through unit root analysis we get;

$$\widehat{\Delta\hat{u}_{t3}} = -1.268\hat{u}_{t3-1}$$

Having,	t	=	(-4.226)
	p- value	=	(0.002)
	R^2	=	0.641
	d	=	2.147

Since the p- value 0.002 < 0.05, the unit root test shows that the residuals are stationary at 5% LOS. Therefore the variables Ln (Y_{3t}), Ln (X_{1t}) and Ln (X_{2t}) are co-integrated.

From the regression model (***), we find that;

- The value of R^2 is fairly high depicting that the model is of a good fit.
- The Durbin Watson statistic *d* is greater than d_u = 1.579 (at 5% LOS) so we can say that there is no autocorrelation present in the model.
- Also the p-value of Ln (X_{1t}) is less than 0.05 suggesting that its coefficient is significantly different than zero and the p –value of Ln (X_{2t}) is more than 0.05 so its coefficient is not significantly different than zero.

4. $\ln(Y_{4t}) = \beta_{04} + \beta_{14} \ln(X_{1t}) + \beta_{24} \ln(X_{2t}) + u_{t4}$

We have the following fitted model;

$$\widehat{\ln(Y_{4t})} = 21.673 + 0.823 \ln(X_{1t}) - 1.002 \ln(X_{2t}) \qquad \ldots (****)$$

Having,	t	=	(1.804)	(6.500)	(-1.031)
	p- value	=	(0.105)	(0.000)	(0.330)
	R^2	=	0.913		
	d	=	2.275		

Through unit root analysis we get;

$$\widehat{\Delta \hat{u}_{t4}} = -1.206 \hat{u}_{t4-1}$$

Having,	t	=	(-3.743)
	p- value	=	(0.004)
	R^2	=	0.583
	d	=	1.960

Since the p- value 0.004 < 0.05, the unit root test shows that the residuals are stationary at 5% LOS. Therefore the variables Ln (Y_{4t}), Ln (X_{1t}) and Ln (X_{2t}) are co-integrated.

From the regression model (****), we find that;

- The value of R^2 is fairly high depicting that the model is of a good fit.
- The Durbin Watson statistic *d* is greater than $d_u = 1.579$ (at 5% LOS) so we can say that there is no autocorrelation present in the model.
- Also the p-value of Ln (X_{1t}) is less than 0.05 suggesting that its coefficient is significantly different than zero and the p-value of Ln (X_{2t}) is more than 0.05 so its coefficient is not significantly different than zero.

5. $\ln(Y_{5t}) = \beta_{05} + \beta_{15} \ln(X_{1t}) + \beta_{25} \ln(X_{2t}) + u_{t5}$

We have the following fitted model;

$$\widehat{\ln(Y_{5t})} = 57.957 + 1.586 \ln(X_{1t}) - 3.884 \ln(X_{2t}) \qquad \ldots (*****)$$

Having,	t	=	(3.590)	(9.316)	(-2.971)
	p- value	=	(0.006)	(0.000)	(0.016)
	R^2	=	0.942		
	d	=	2.097		

Through unit root analysis we get;

$$\widehat{\Delta \hat{u}_{t5}} = -1.206\hat{u}_{t5-1}$$

Having, t = (-3.346)
p- value = (0.007)
R^2 = 0.528
d = 1.902

Since the p- value 0.007 < 0.05, the unit root test shows that the residuals are stationary at 5% LOS. Therefore the variables Ln (Y_{5t}), Ln (X_{1t}) and Ln (X_{2t}) are co-integrated.

From the regression model (*****), we find that;

- The value of R^2 is fairly high depicting that the model is of a good fit.
- The Durbin Watson statistic *d* is greater than d_u = 1.579 (at 5% LOS) so we can say that there is no autocorrelation present in the model.
- Also the p-values for both the variables, Ln (X_{1t}) and Ln (X_{2t}), are less than 0.05 suggesting that the resulting coefficients are significantly different than zero.

In all the models considered, it was observed that their linear combination was I(0) i.e. stationary even though the variables themselves were first order integrated i.e. I(1). Therefore, the variables were co-integrated.

<u>TABLE -9</u>: table demonstrating the relationships found to be existing after co-integration test.

Vehicle type	Effect on sales due to GDP per capita in a year	Effect on sales due to total no. Of road accidents in a year
CARS	**1.665%** increase	**3.469%** decrease
TWO WHEELERS	**1.586%** increase	**3.884%** decrease
JEEPS	**0.823%** increase	No significant effect
TAXIS	**1.102%** increase	No significant effect
BUSES	No significant effect	No significant effect

CONCLUSION

1. This study was able to conclude empirically, that a change in personal income positively affects the overall sales of vehicles in India while the total no. of road accidents negatively affected the overall sales in some types of vehicles (see table – 9).

2. No significant effect of personal income and road accidents on the overall sales of buses seems plausible since buses are mostly used in public transportation.

3. A possible explanation for the total number of accidents negatively impacting a customer's choice only while buying cars and two wheelers could be because of the major share of two wheelers meeting with a fatal accident and issues with proper opening of airbags, break failure etc. in case of cars.

4. There are other related factors too which contribute to the huge increase in vehicular population in the country. Some of them are
   a. Increasing working age population,
   b. Rapid urbanisation,
   c. Availability of finance i.e. loans,
   d. Relaxation on import duty,
   e. More second hand purchasing of vehicles,
   f. Inadequate public transport,
   g. Innovation in automobile industry which has led to a decreased cost of operating the vehicles thereby increasing the demand for it.

   These factors were not considered under this study. However, a careful study of above mentioned factors can provide more robust estimates.

REFERENCES

1. Fundamentals of Applied statistics, S.C. Gupta and V.K. Kapoor

2. Basic Econometrics, Damodar N. Gujarati, Dawn C. Porter and Sangeetha Gunasekar

3. Population dataset, Census of India 2001 & 2011. GOI.

4. Profile of road accidents, Road Transport Year Book 2012-13. New Delhi: Transport Research Wing, Ministry of Road Transport and Highways, GOI. (https://morth.nic.in/road-transport-year-books)

5. Registered motor vehicles in India, https://data.gov.in/catalog/total-number-registered-motor-vehicles-india?filters%5Bfield_catalog_reference%5D=92126&format=json&offset=0&limit=6&sort%5Bcreated%5D=desc

6. GDP per capita (current US$), Reserve Bank of India. (https://www.rbi.org.in/Scripts/AnnualPublications.aspx?head=Handbook%20of%20Statistics%20on%20Indian%20Economy)

7. GDP of India (in US$), World bank (https://databank.worldbank.org/indicator/NY.GDP.MKTP.CD/1ff4a498/Popular-Indicators)