Poisson and Negative Binomial Regression Models Application to Model the Factors of Car Ownership in Akure, South West, Nigeria

Joseph O. Oyedepo[1] and Japheth E. Etu[2,*]

[1,2]*Department of Civil and Environmental Engineering,
Federal University of Technology, Akure, Nigeria.*

Abstract

Increase in number of cars without commensurate increase in the number of transport facilities and infrastructures has led to diverse traffic problems in many Nigerian cities like Akure. Factors which contribute to increase in the numbers of cars owned in Akure metropolis were investigated in this study. The study area was divided into three density zones namely High, Medium and Low while, data was collected using well-structured household questionnaire survey distributed amongst residents; with the survey yielding a return of 1002 questionnaire out of the 1181 distributed. Results from field findings gave the average number of cars owned per household in the study area as 0.62. Results of the Poisson Regression Model show that a change in the number of employed household members will decrease the number of cars owned in the study area by 9% while, a unit increase in the number of driver's license holders in the household, academic qualification and average monthly income of the household will increase the number of cars owned by 60%, 26% and 30% respectively. The negative binomial model indicates that a change in the number of employed household members will decrease the number of cars owned by 10% whereas a change in the number of driver's license holders in the household and monthly income will lead to an increase in the number of cars owned by 101% and 24% respectively. The test of model effects affirm that all the predictor variables are statistically significant indicating a good fit for the model predicted. Out of the two models, Poisson regression model is found to be a superior model due to a higher log likelihood ratio Chi Square and improved statistically significant variables. The findings in this research will assist government agencies to plan future transportation infrastructure development.

Keywords: Poisson, Binomial, Regression, Model, Household, Factors.

1. Introduction

This study applies the Poisson and Negative Binomial Regression models to investigate the influence of household factors on car ownership in Akure metropolis, the capital of Ondo State located in south west Nigeria. The decision to buy a car is one of the most significant decisions made by a consumer. Not only is it one of the most significant financial investments for many people, but it also represents a dramatic increase in mobility and is often viewed as a status symbol. It is also a key feature of the transportation system that to a great extent influences travel behaviour and participation in out-of-home activities. The increase in utilization patterns and high-levels of household car ownership have a direct effect on energy consumption and air quality levels in local and global scales [1].

The population growth of Nigerian cities is at an alarming rate, thereby leading to increased car ownership in the cities. However, with these high level of car ownership and reliance, there is an absence of studies on the factors influencing the numbers of vehicles per households [2]. The general problems that are associated with increase in vehicular movement in urban areas such as Akure are: Traffic congestion and parking difficulties; longer commuting; public transport inadequacy; loss of

public spaces; environmental impact and energy consumption; and land consumption to transportation facilities [2]. Increasing awareness about these problems is expected to affect dramatically mobility habits and to create opportunities for changes in the automotive industry over the next five to ten years [3].

According to Awoyemi, *et al* (2012), the study of vehicle ownership in developing countries can forecast the future level of vehicle ownership and help transportation officials in road system planning, policies and design. He also stated that while developed countries are confronting serious auto-oriented problems, motorization in most developing countries is still in its infancy. Scott & Axhausen, (2005) stated that car ownership is a key element in the study and simulation of urban systems and as such is of interest to develop models of car ownership that are capable of explaining and predicting household's choices. Understanding of how households choose the number of vehicles to own, based on where they live and other explanatory variables is of vital importance to urban planners, transportation engineers, national government and judgment makers [6]. This is why this study is being undertaken to determine the factors that influence car ownership in Akure metropolis.

2. Car ownership modelling

Many studies on car ownership models carried out by Cirillo and Liu (2013) and De Jong *et al* (2004) among others have shown that model types can be classified into aggregate and disaggregate models. De Jong (1994), however divided the disaggregate models into statistical models and indirect utility models in which the statistical model assumes that a household has a structural desired annual travel distance, which depends on attributes of the household. Only if this desired travel distance exceeds a threshold, the household will own a car. Explanatory variables for car ownership and use in this model are household income, household size, age, gender and occupation of the head of the household. Furthermore, the relationship between car ownership and car use is included in the indirect utility model in a way that is consistent with economic theory. The basic idea is that households compare combinations of car ownership and car use with each other and choose the combination that gives then the highest utility. The model also contains fixed car cost and variable car cost as explanatory variables (besides the variables that are in the statistical model).

Dargay, J. and Gately, D. (1999) confirmed the relationship between countries' gross domestic products (GDPs) and car ownership levels across a range of developed and transitional countries, where higher GDPs are strongly associated with higher car ownership levels. Furthermore, it was estimated that the ultimate common car ownership saturation values for all countries lies around 0.85 vehicles per capita, including both light and heavy vehicles. Also Mokonyama, and Venter, (2007) cited from the analysis of low-income countries' car ownership data that as low-income countries become more prosperous, there is an inevitable and rapid rise in their car ownership and use.

A study by Kalenoja (2001) calibrated car ownership model for Tampere, Finland, based on the availability of local services in a traffic zone by six types of areas. It was found that car ownership was highest in the low service level suburbs and also in sparsely populated areas but lowest in the central business district and surrounding areas. Similar spatial patterns exist in South Africa to some extent, where apartheid spatial and economic planning has left historically black townships/suburbs with low levels of retail and other services, which would tend to increase the perceived need for private transport and thus car ownership. As explained by Dargay and Vythoulkas, (1999) household size does not directly determine car ownership but rather the age structure of the household. It was confirmed in their research results based on some cohort data that household car ownership continues to increase as the offspring, still living in their parents' household, acquire cars, then declines as they leave home, taking their cars with them.

The negative binomial model has been used to evaluate count data, when the mean and variance of observed frequency significantly varies while the Poisson model is appropriate for count data when the mean and variance do not show over-dispersion. Motorcycle ownership models were formulated

using Poisson Regression and Negative binomial logistic regression models in Ifedore Local Government Area of Ondo state, Nigeria. The two models show quite similar results as they both show that the number of motorcycles owned will reduce as educational status of the rider and average monthly income of household increases while there will be an increase in the number of motorcycles owned as the age range of the rider increases [12]. The negative binomial model has been used to evaluate crash frequency data, when the mean and variance of the crash frequency significantly varies, and these results have been used to recommend appropriate safety-related countermeasures. Agbelie (2016), while carrying out a comparative empirical analysis of statistical models for evaluating highway segment crash frequency found the random-parameters negative binomial model to be a superior model due to a relatively higher log likelihood at convergence, better overall model fit, r^2, and improved statistical significance of the estimated parameters.

3. Study methodology

3.1. Description of the Study Area

Akure is a traditional Nigerian city and like other traditional Yoruba towns in the country, it existed long before the advent of British colonial rule in the country, however, history has it that the town had been firmly established by AD 1440. Akure was made Ondo state capital on February 3, 1976 by the Military Government. Akure with the provisional census figure of 387,087 people according to 2006 national population census, is located on latitude 70 20" N and longitude 50" E and is easily accessible to other major urban centres within and outside the state, such as Owo in the East, Ado-Ekiti in the North and Ondo in the South, all within 50km radius.
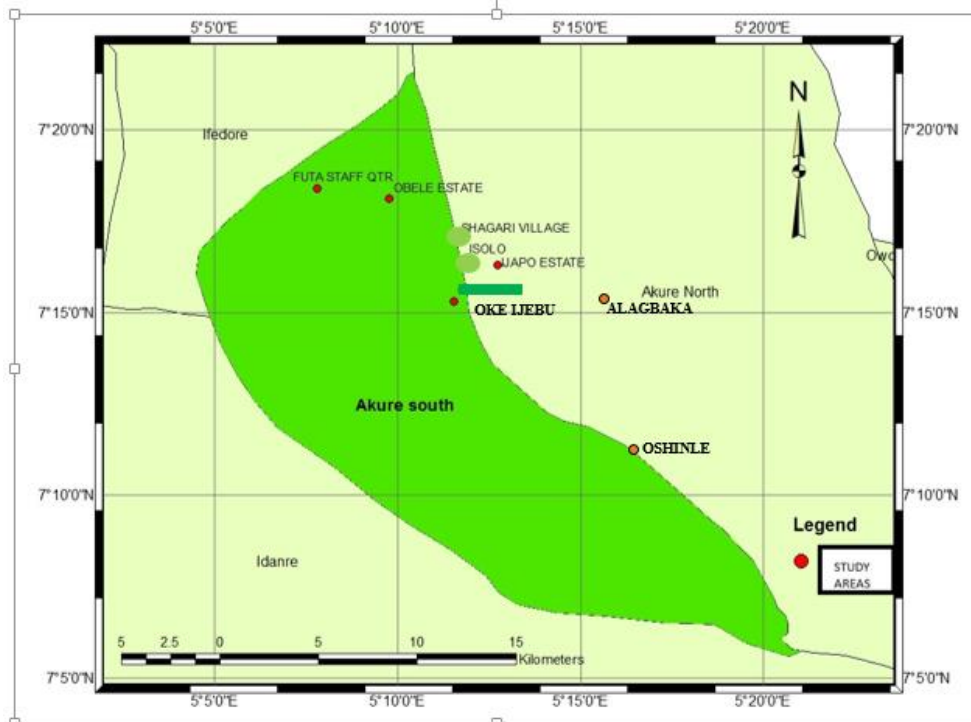

Figure 1: The Map of Akure Metropolis.

Within 100 km radius are the towns of Ikare, IkoleEkiti, ljeroEkiti to the North and Okitipupa in the South. The topography of the town is relatively flat. Akure is connected to other parts of the

country by road network system. Being an administrative centre of the state, the rate of population growth in Akure has been greater than that of any town in the state.   The natural pattern of development is linear along its main roads namely Oyemekun, Oba Adesida, Hospital, Arakale and Oda roads. The existing land use is characterized by a medium density of structure within the inner core areas. Akure is mostly residential with residential areas forming over 90% of the developed area with little or no industrial areas. The additional activities such as warehousing, manufacturing, workshops and other commercial purposes are commonly located within the residential neighbourhoods. Figure 1 is the map of Akure showing the study location.

3.2. Data acquired and source

Data for this study were sourced from questionnaire survey on households within the study area. The questionnaire focused on socio-economic and demographic situation of the households, also, information on the household residential location, vehicle attributes and other questions that will help in achieving the aim of the study.

Three different zones classified based on their population density were selected for this study. Table 1 shows a total number of 1002 as the number of questionnaire returned. It also shows the distribution of the questionnaire across the zones and the different locations. The highest number of respondents as expected is from the high density zone since this is the most populated zone of all the zones sampled.

Table 1: Questionnaire returned from the Field Survey

Density Zone	Locations	Number of questionnaire distributed	Number of questionnaire returned	Total
LOW DENSITY	ALAGBAKA	155	108	
	IJAPO ESTATE	153	107	215
MEDIUM DENSITY	ORITA OBELE	181	163	
	AULE	205	171	334
HIGH DENSITY	OSHINLE	255	242	
	OKE-IJEBU	232	211	453
	TOTAL	1181	1002	1002

3.3. Data Analysis and modelling framework

Data obtained from different zones were coded, imputed and analysed in SPSS 22 environment. SPSS 22 is a software package used for statistical analysis. SPSS was used in both univariate analysis (that is, frequency distribution and percentages) and content analysis to complement the survey findings and formulation of Predictive Models of Car Ownership using Poisson and Negative Binomial Regression Models.

The objective of the Poisson regression model is to estimate log functions of a count variable. Count variables are all positive integers and for rare events so the Poisson distribution (rather than the Normal) is more appropriate since the Poisson mean > 0.

Since observed data will almost always display pronounced overdispersion, analysts typically seek alternatives to the Poisson model, such as the negative binomial model. The negative binomial model is employed as a functional form that relaxes the equi-dispersion restriction of the Poisson model.

These models are used in this study because the dependent variable for this study is Number of Cars, N_{CAR} (which is a count variable). Table 2 shows the variables used in the model.

The logarithm of the response variable in these models is linked to a linear function of explanatory variables such that:

$$Log_e(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \tag{1}$$

So that;

$$Y = (e^{\beta_0})(e^{\beta_1 X_1})(e^{\beta_2 X_2})(e^{\beta_n n}) \tag{2}$$

Where
  Y = Dependent variable
  β_0 = CONSTANT or intercept
  β_1 = coefficients of the 1st independent variable
  X_1 = 1st independent variable
  β_2 = coefficients of the 2nd independent variable
  X_2 = 2nd independent variable
  β_n = coefficients of the nth independent variable
  X_3 = nth independent variable

In other words, the typical Poisson and Negative Binomial regression model expresses the log outcome rate as a linear function of a set of predictors.

The results of the model analysis and field surveys are presented in section 4 below.

Table 2: Variables in the Model

	S/No	Variable Name	Variable code
Dependent Variable:		Number of cars	N_{CAR}
Independent Variables:	1	Number of household members	N_H
	2	Number of license holders	N_L
	3	Number of employed	N_E
	4	Academic qualification of respondent	A_Q
	5	Average Monthly income of household	M_I

4. Results

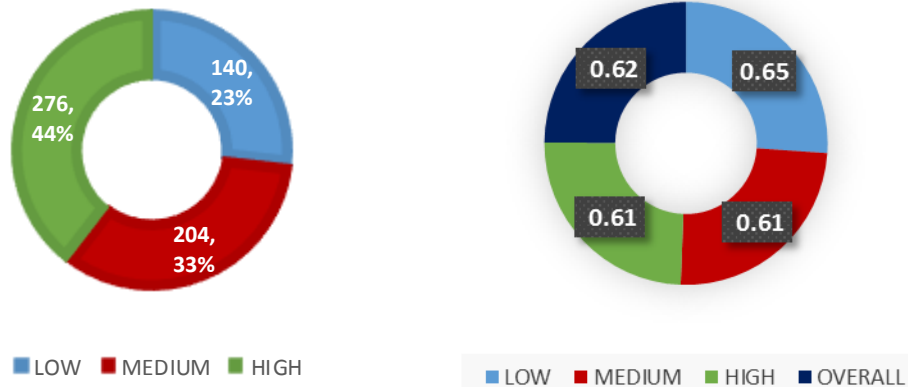## 4.1. Field Survey Results

Figure 2a: Number of cars owned in the study area
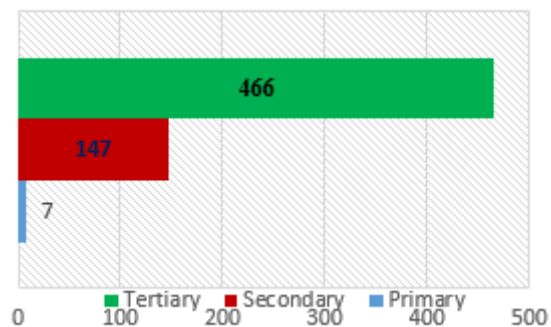
Figure 2b: Number of cars per household

Figure 3: Number of cars owned and academic qualification

The number of cars owned in the zones are shown in Figure 2a below. It is seen that highest number of cars are owned in the high density zone, followed by the medium and then the low density zone. It is believed that the high density zone has the highest number of cars because it consists of the highest number of respondents to the survey and not because of superior economic power. This is evident from figure 2b which shows the low density zone as having the highest number of cars per household. The results show 0.65, 0.61 and 0.61 cars per household for the low, medium and high density zones respectively. While the computation for the number of cars per household across the entire study area is 0.62cars.

A comparison of the average monthly income of the household with the number of cars owned across the zone showed that the number of cars owned increases as income increases. The peak income group i.e. 100,000 and above, accounted for 229 of households in which cars were owned, and this goes in a descending order to 205, 155 and 30 for the lower income categories respectively. This is as shown in figure 4 which also shows the vehicle distribution across the zones for the different

income categories. This means that a household is more likely to own a car as its average monthly income increases. This makes money a major determinant of car ownership in the study area.
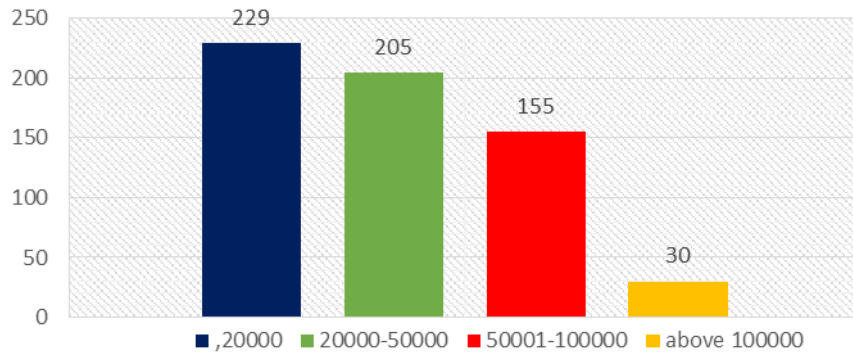

Figure 4: Number of cars owned and household income

A relative analysis of the household characteristics such as the number of household members, number of employed household members and number of driver's license owners with the number of cars owned across the zone showed that the showed that a household with fewer members tend to own more cars than those with lesser members. Households with two employed people own more cars in the study area while households with a single driver's license owner tend to own more cars in the area. Increasing number of driver's licence owners in a household reduces the number of cars owned as can be seen from figure 5 above.
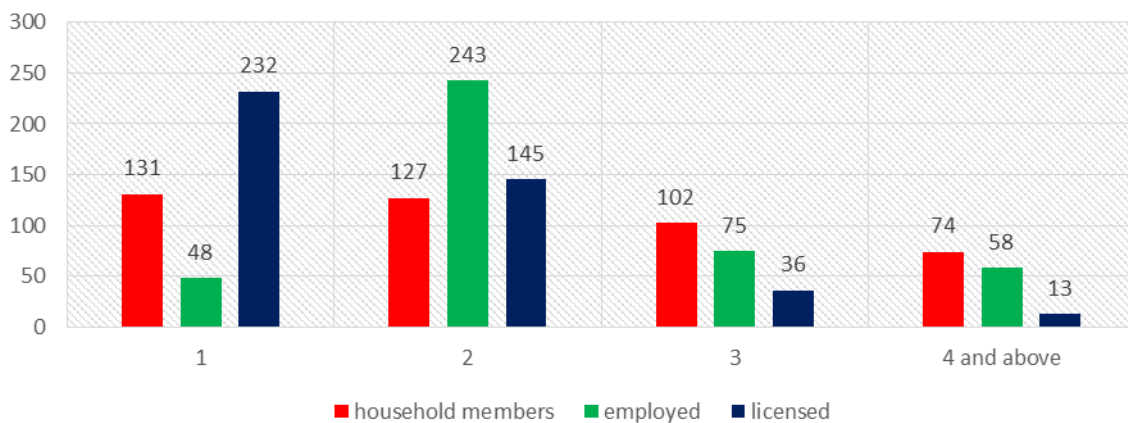

Figure 5: Household characteristics

4.2. Model results

The data across all of the zones are combined here to develop an overall model for the study area.

4.2.1 Poisson regression model for the study area

In table 3 are the parameter estimates derived for the entire study area using the Poisson model. The parameters shows that it is only the number of household members that has no significant influence on the dependent variable. All other variables bear a significant relationship at the 95% confidence level. The number of employed household members in the model has a negative significant relationship with the number of cars owned, that is to say a unit change in the number of employed household members will lead to a decrease in the number of cars owned in the study area whereas the number of license holders, the academic qualification and average monthly income positively influence the number of cars owned meaning that a unit change in any of them will in turn lead to an increase in the number of cars owned in the study area.

Table 31: Parameter Estimates of Poisson Regression Model

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)
			Lower	Upper	Wald Chi-Square	Df	Sig.	
(Intercept)	-2.149	.3565	-2.848	-1.451	36.344	1	.000	0.11
num_of_household	.023	.0186	-.013	.060	1.558	1	.212	1.26
num_of_license	.473	.0357	.403	.543	175.489	1	.000	1.60
num_of_employed	-.097	.0347	-.165	-.029	7.833	1	.005	0.91
academic_qualification	.233	.0903	.056	.410	6.666	1	.010	1.26
monthly_income	.258	.0446	.170	.345	33.379	1	.000	1.30

The model equation from table 3 considering only the significant variables is:

$$N_{CAR} = 0.11 + 1.60N_L - 0.91N_E + 1.26A_Q + 1.30M_I \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.(3)$$

Equation 3 implies that a change in the number of driver's license holding household members will lead to an increase in the number of cars owned by 60% whereas a change in the number of employed household members will decrease the number of cars owned by 9% (from Exp(B) -1 =0.91 -1 =9%). A change in the academic qualification and monthly income will lead to an increase in the number of cars owned by 26% and 30% respectively.

Table 4: Omnibus Test of Model Effects for Poisson Regression Model

Likelihood Ratio Chi Square	df	Sig.
316.968	5	.000

The test of model Effects as shown in Table 4 shows that the model is a proper fit for the variables considered as there is a general significance of 0.000 which is within the 95% confidence level. Also the Standard error values from table 3 are all below 2 thereby showing a very good prediction of the dependent variable with minimal errors. The very high chi-square value of 316.968 shows that the model has adequately explained the variables contained therein. It can therefore be concluded from the foregoing that the model has been able to show a significant relationship between the dependent variables and the independent variables and that the effects of the factors identified on the number of cars owned in the study area are not by chance.

4.2.2   Negative binomial regression model for the study area

The parameters derived for the Negative Binomial Regression Model shows that the number of household members that has no significant influence on the dependent variable as in the Poisson model, however, the variable academic qualification of household head also had no significance influence on the number of cars owned in the study area.

The number of employed household members in the model has a negative significant relationship with the number of cars owned while the number of license holders and average monthly income both have positive influence on the number of cars owned thereby displaying the same influence as in the Poisson model.

Table 52: Parameter Estimates of Negative Binomial Regression Model

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)
			Lower	Upper	Wald Chi-Square	Df	Sig.	
(Intercept)	-2.306	.4934	-3.273	-1.339	21.837	1	.000	0.100
num_of_household	.023	.0258	-.028	.073	.785	1	.376	1.023
num_of_license	.699	.0706	.561	.837	98.019	1	.000	2.012
num_of_employed	-.105	.0498	-.202	-.007	4.414	1	.036	0.900
academic_qualification	.227	.1266	-.021	.475	3.212	1	.073	1.255
monthly_income	.215	.0652	.087	.343	10.876	1	.001	1.240

From table 5 the model equation excluding the insignificant variables is derived thus;

$$N_{CAR} = 0.10 + 2.012N_L - 0.900N_E + 1.240M_I \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(4)$$

The negative binomial model equation from Equation 4 shows that a change in the number of employed household members will decrease the number of cars owned by 10% whereas a change in the number of driver's license holders in the household and monthly income will lead to an increase in the number of cars owned by 101% and 24% respectively.

Table 6: Omnibus Test of Model Effects for Negative Binomial Regression Model

Likelihood Ratio Chi Square	df	Sig.
188.129	5	.000

Again the omnibus test of model coefficients has shown a strong relationship between the dependent variable and the independent variables as can be seen in table 6.

The results of the models have showed the absence of over-dispersion in the data used since the estimated parameters from both models showed consistency in their estimates. However, due to a relatively higher likelihood ratio Chi-Square and improved number of statistically significant variables the Poisson Regression Model is found to be a superior model and is recommended for use by decision makers.

5. Conclusion

The aim of this study was to develop car ownership model for Akure metropolis by identifying factors affecting car ownership in the area. The data for the study were obtained from households in the metropolis by administering questionnaires to each household in three different zones classified based on their population density; namely High, Medium and Low density zones. Data from the survey were analysed using the Statistical Package for Social Science (SPSS 22) for both univariate analysis (that is, frequency distribution and percentages) and to generate a Poisson Regression Model of Car ownership for the study area.

Results from field findings gave the average number of cars owned per household in the study area as 0.62. However, the low density zone had the highest number of cars per household with an ownership rate of 0.65 cars per household while the high and medium density both gave 0.61 cars per household.

The Poisson regression model shows that the number of employed household members has a negative significant relationship with the number of cars owned which implies that there will be a 9% decrease in the number of cars owned for every unit increase in the number of employed household members whereas a unit increase in the number of driver's license holders in the household, academic qualification and average monthly income of the household will lead to 60%, 26% and 30% increase in the number of cars owned respectively. The negative binomial model equation shows that a change in the number of employed household members will decrease the number of cars owned by 10% whereas a change in the number of driver's license holders in the household and monthly income will lead to an increase in the number of cars owned by 101% and 24% respectively. Out of the two models, Poisson regression model is found to be a superior model due to a relatively higher log likelihood ratio Chi Square and improved statistically significant variables.

Generally, the Standard error values for the models are below 2 which is the standard for a good prediction of the dependent variable. Also, the models effectively explained the variables contained by yielding a high chi-square value. These have shown that the model is a good fit for the selected variables while the p-values which are less than 0.05 thereby validating the model also.

It is expected that government will use the findings in this research to plan future transportation infrastructure development so as to minimize the problem of congestion and pressure facing the current transportation infrastructure while relevant traffic control agencies are expected to apply the results of this studies in carrying out their functions so as to reduce the problems of road accidents and recklessness currently being experienced on the highways in the metropolis. Future studies relating to vehicle purchase price, fuel costs and operation cost are also recommended

References

[1.] Fenger, J. (1999): Urban Air Quality. *Atmospheric Environment,* Vol. 33, pp. 48-9

[2.] Rodrigue, J-P et al. (2013).The Geography of Transport Systems, Hofstra University, Department of Global Studies & Geography. Retrieved from http://www.people.hofstra.edu/geotrans.on 03/10/2013 at 12:00noon

[3.] Cirillo, C. and Liu, Y. (2013) Vehicle Ownership Modeling Framework for the State of Maryland: Analysis and Trends from 2001 and 2009 NHTS Data Transportation Model (MSTM). *American Society of Civil Engineers.*

[4.] Awoyemi, O. K., Ita, A. E., Oke, M. O., Abdulkarim, I. A. & Awotayo, G. P.(2012). An Analysis of Trip Generation and Vehicular Traffic Pattern in Akure Metropolis Ondo State, Nigeria. *Journals of Social Science and Public Policy*, Volume 4.

[5.] Scott, D.M. & Axhausen, K.W. (2005). Household Mobility Tool Ownership: Modelling Interactions between Cars and Season Tickets *Transportation (forthcoming).*

[6.] De Jong, G., Fox, J., Pieters, M.,Daly, A. and Smith, R., (2004) A comparison of car ownership models *Transport Reviews* 24(4), pp.379.

[7.] De Jong, G. C., Kitamura, R. and Klooster, J (1994): A disaggregate model of vehicle holding duration, type choice and use.

[8.] Dargay, J. and Gately, D. (1999). Income's effect on car and vehicle ownership worldwide: 1960–2015. *Transport atcion Research Part A: Policy and Practice*, 33(2):101–138.

[9.] Mokonyama, M. and Venter, C. (2007). *Journal of the South African institution of civil engineering* Vol 49 no 3, pages 2–10, paper 633.

[10.] Kalenoja, H. (2001). Car ownership effects on the availability of local service. *The Nordic Research Network on modelling transport, land use and environment*, Finland

[11.] Dargay J. and Vythoulkas P., (1999) Estimation of a Dynamic car Ownership Model; a Pseudo-panel Approach. *Journal of transport economics and policy*, September 199, vol 33, part 3, pp 287-302.

[12.] Oyedepo J. O. and Etu J. E. Statistical Model Analysis of Dependence on Motor Cycle Transport at Ifedore LGA Ondo Nigeria *Futojnls, Volume2,* Issue-1, Pp- 118 - 126

[13.] Agbelie, B.R.D.K. A comparative empirical analysis of statistical models for evaluating highway segment crash frequency. *Journal of Traffic and Transportation Engineering (English Edition) (2016)*, http://dx.doi.org/ 10.1016/j.jtte.2016.07.001