Journal of the Nigerian Association of Mathematical Physics
***Volume 48* (*Sept. & Nov., 2018 Issue*), pp189 –198**
© J. of NAMP

MATHEMATICAL MODELS OF ROAD ACCIDENT IN A DEVELOPING ECONOMY

Hamza Mansur [1], UL Okafor[2] and Bashir Yusuf [3]

[1,2]Department of Mathematics, Nigerian Defence Academy Kaduna
[3]Department of Mathematics, Federal University Dutse

Abstract

In this paper, two linear Models were developed using Cochrane Orcutt Iteration Method (CO) and Multiples Linear Regressions Method (MLR) to analyze road traffic accidents in developing countries. The factors influencing such accidents have been analyzed and corrective measure was offered using the Models. The errors of the models are taken in to account and the validity of the models was checked by normality test, linearity test, autocorrelation test, and homoscedasticity test. The two linear models were compared with the help of graph.

Key word: *Road accidents, Iteration, Linearity, Autocorrelation, Homoscedasticity*

1. Introduction and Literature review

Road accidents are becoming very common in developing countries and are robbing the nation of its valuable human resources; many people are dying every day in road accidents and many are injured. Every year more than 1.17million people die in road crashes around the world, 70% of these occur in developing countries; globally, every 10million people are crippled or injured each year, 65% of deaths involved pedestrians, 35% pedestrians are children [1]. It has been estimated that more will die and 60 million will be injured during the next 10 years in developing countries unless urgent action is taken, according to [2] who also reported that one person is dying in roadway during crashes nearly every 12 minutes, and of that number 25,136 died in roadway accident/crashes, 9,213 in intersection crashes and 4,749 in pedestrian crashes. However, World Health Organization has estimated that nearly 25% of fatal injuries worldwide are as result of road traffic crashes, with 90% of the fatalities occurring in low and middle income countries, World Report on Road Traffic Injury Prevention, [1]. Road accidents cause significant social and economic costs typically between 1 and 3 percent of Growth National Product, [3]. They also result in the use of a high proportion of medical facilities and the depletion of score foreign exchange. The research forecasted that by the year 2020 road accident would move up to third place leading causes of death and disability facing the world community. Moreover, in developing countries, growth in urbanization and in the number of vehicles has lead to increased traffic congestion in urban centers and increase in traffic accidents on road networks, which were never designed for the volumes and types of traffic that they now to carry. In Nigeria particularly, about 300,000 persons lost their lives in 1,000,000 road accidents between 1960 and 2005, while over 900,000 people suffered various degrees of injuries within the same period [4].

The main aim of this research is to develop models for capturing the number of road accidents victims using 48 months data available from 2010 - 2014. A number of accident prediction models have been developed in the last decades to estimate the expected accident frequencies on roads as well as to identify various factors associated with the occurrence of accidents. For examples; [5] used Multivariate Regression Techniques for Analyzing Auto Crash Variables in Nigeria, Accidents Statistics covering period of five years were collected (2003-2007) from Lagos State Command of the Federal Road Safety Corps.

[6] argued that Smeed's model [D / N = 0.0003 (N / P) -0.67] where D, N, P are deaths, motor vehicles and population respectively, was based on only one year data and each country has different social and economical conditions, therefore, this model is not applicable for all countries. He proposed the use of the number of motorized vehicles as dependent variables in accident models to estimate number of deaths. He obtained different B1 and B2 exponential values in the model. In which $C, B_1 and\ B_2$ are coefficients determined by regression analysis. [7], used regression method to analyzed road traffic accidents in Nigeria for the period 1975 – 2009, the statistical package used in the study was SPSS (Statistical Package for

Corresponding Author**:** Hamza M., Email: Hamzamansir@gmail.com, Tel: +2347036187203

the social science). The regression summary also shows that the improvement in safety measures as a result of the establishment of the Federal Road Safety Commission by the Federal Government in 1988 has helped to reduce motor vehicle deaths to some extent. This is shown by the negative correlation of the variable (YR_t – 1988) on motor vehicle deaths. However the model shows that the two independent variables are statistically insignificant as shown by the F = 3.289 value which is less than the table value of 3.32 at 0.05 level of significance. Gajendran et.al [8] discuss about three types of accident prediction model "System Dynamic Model, Fuzzy logic and Bayesian Method. The Complex, Dynamic and Non-linear interaction can be understood using system dynamic model. Fuzzy logic deals with occurrence of sets and elements, Bayesian refer to methods in probability and Statics which has held to model the interaction between road geometry, traffic characteristics and accident frequencies by means of linear regression model. Also, [9] developed Poisson and Negative-Binomial models based on traffic police report data within the Galle police division for years 2011, 2012 and 2013. From the results of Negative Binomial model they found that, the key variables which cause the occurrence of accidents are experience of the driver (year of driver license issue), vehicle type, light condition and time of the accident.

2. Materials and Methods

The data used in this study is secondary data collected from Federal Road safety Corps Jigawa State Headquarter (Dutse) and covered the period of five years (2010 - 2014). It should be clear that the number of variables to be included in the model depend on the nature of the phenomenon being studied and the purpose of the research.

2.1 Methods

The Cochrane – Orcutt method reports one fewer observation than Ordinary Least Squares; this reflects the fact that the first transformed observation is not used in the CO method. Asymptotically, it makes no difference whether or not the first observation is used, but many time series samples are small, so the difference can be notable in applications. However, in some application of the Cochrane – Orcutt the estimates differ in practically important ways from the OLS estimates. Typically, this has been interpreted as verification of feasible CO superiority over OLS. To see why consider the regression Model of;

$$y_t = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4 + e_t \qquad (2.1)$$

Where; $y_t = Number\ of\ accident\ at\ year\ t,\ X_1 = Number\ of\ Death$
$X_2 = Number\ of\ injured,\ X_3 = People\ involved\ in\ the\ accident$
$X_4 = Number\ of\ Vehicle\ involved\ in\ the\ traffic$, and
$a_0\ is\ the\ intercept,\ a_1, a_2, a_3\ and\ a_4$ are the coefficients of independents variable, e_t is the random error. We asserted that CO was consistent under the strict exogeneity assumption, which is more restrictive. In fact, it can be shown that the weakest assumptions that must be hold for CO to be consistent is that the sum of $x_{t-1}\ and\ x_{t+1}$ is correlated with u_t:

$$Cov[x_{t-1} + x_{t+1}, u_t] = 0 \qquad (2.2)$$

Practically speaking, consistency of CO requires u_t to be uncorrelated with $x_{t-1}, x_t\ and\ x_{t+1}$. In this case we need $E(x_t - \rho x_{t-1})(u_t - \rho ut-1=0$, where $ut-\rho ut-1$ is the error, if we expand the expectation, we get

$$E[(x_t - \rho x_{t-1})(u_t - \rho u_{t-1})] = -\rho[E(x_{t-1}u_t) + E(x_tu_{t-1})]$$

Because $E(x_tu_t) = E(x_{t-1}u_{t-1}) = 0$ by assumption. Now, under stationary, $E(x_tu_{t-1}) = E(x_{t+1}u_t)$ because we are just shifting the time index one period forward $E[(x_{t-1} + x_{t+1})u_t]$ and the last expectation is covariance in equation (2.2) because $E(u_t = 0)$. Our expectation is that, MLR and CO might give significantly different estimates because (2.2) for CO fail. In this case, MLR which is still consistent under $cov(x_t, t_u) = 0$ is preferred to CO (which is still inconsistent). Since MLR and CO are different estimation procedures, we never expect them to give the same estimates. If they provide similar estimates of a_i, then CO is preferred if there is evidence of serial correlation, because the estimator is more efficient and the CO test statistics are at least asymptotically valid. One of the most common and serious mistake is to accept a regression Model without plotting the Residual against each independent variables, those variables not included in the Model y and Residuals of the previous period.

3. Results

Based on statistical analysis of secondary data of road accidents obtained from Federal Road Safety Corps Jigawa State relationship between numbers of accidents per month versus number of death, injury, passengers and vehicle was found. Using Cochrane – Orcutt estimation Method and NCSS software package the results are summarized in table 1 to 15.

3.1 Regression Report using Cochrane Estimation method

Table 1: Run Summary section

Parameter	**Value**	**Parameter**	**Value**
Dependent Variable	Y	Rows Processed	48
Number Ind. Variables	3	Rows Filtered Out	0
Weight Variable	None	Rows with X's Missing	1
$\boldsymbol{R^2}$	0.8251	Rows with Weight Missing	0
Adj $\boldsymbol{R^2}$	0.8126	Rows with Y Missing	0
Coefficient of Variation	0.3004	Rows Used in Estimation	47
Mean Square Error	1.896333	Sum of Weights	46.000
Square Root of MSE	1.377074	Completion Status	Normal Completion
Ave Abs Pct Error	24.865	Autocorrelation (Rho)	0.1921

The above $\boldsymbol{R^2}$ of 0.8251 indicate that the model is able to explained for about 82.51% of the change in the rate of accidents with only 17.49% not been explained by the model.

Table 2 Regression Equation Section of CO model

Independent Variable	Regression Coefficient b(i)	Standard Error Sb(i)	T-Value to test H0:B(i)=0	Probability Level	Reject H0 at 5%?
Intercept	5.6805	0.6752	8.414	0.0000	Yes
X_1	-0.0694	0.0654	-1.061	0.2949	No
X_3	-0.1462	0.0205	-7.118	0.0000	Yes
X_4	0.0298	0.0022	13.568	0.0000	Yes

Estimated Model

$$\hat{Y} = 5.6805 - 0.0694X_1 - 0.1462X_3 + 0.0298X_4 \tag{3.1}$$

Note that from table 2 above the coefficient of X_3 *and* X_4 are statistically significant, as its absolute t – value of 7.118 and 13.568 is greater than the table value test of 2.018 which is used to calculate the confidence limit. And we reject the null hypothesis of slope equal to zero if the probability level is greater than the alpha (0.05) at 5% level of significance.

Table 3 Regression Coefficient Section of CO model

Independent Variable	Regression Coefficient	Standard Error	Lower 95% C.L	Upper 95% C.L.	Standardized Coefficient
Intercept	5.6805	0.6752	4.3180	7.0431	0.0000
X_1	-0.0694	0.0654	-0.2015	0.0627	-0.0692
X_3	-0.1462	0.0205	-0.1877	-0.1048	-0.4984
X_4	0.0298	0.0022	0.0254	0.0343	0.9587

Note: The T-Value used to calculate these confidence limits was 2.018.

From the standard error, the model fit the data and can be used for accidents analysis as it has less error for predictions.

Table 4 Analysis of Variance Section for CO model

Source	DF	R^2	Sum of Squares	Mean Square	F-Ratio	Prob. Level
Intercept	1		966.5437	966.5437		
Model	3	0.8251	375.7436	125.2478	66.047	0.0000
Error	42	0.1749	79.64599	1.896333		
Total (Adjusted)	45	1.0000	455.3895	10.11977		

The degree of freedom (DF) and mean square are used to calculate the f ratio for testing the significant of the regression. The probability level is used to test null hypothesis at alpha level when the probability level is less than the alpha we reject the null hypothesis otherwise we accept.

Table 5 Serial Correlation of Residuals from Corrected Model

Lag	Serial Correlation	Lag	Serial Correlation	Lag	Serial Correlation
1	-0.0222	9	0.0329	17	0.0409
2	-0.0121	10	-0.1174	18	-0.0245
3	0.0787	11	-0.0959	19	-0.1514
4	0.0495	12	0.0670	20	-0.0468
5	0.0782	13	0.0030	21	0.1790
6	-0.3212	14	-0.0708	22	-0.0136
7	-0.0662	15	-0.1286	23	-0.1802
8	-0.0903	16	0.0768	24	0.0583

Above serial correlations significant if their absolute values are greater than 0.291730, from table 5 there is no serial correlation of the residuals as their absolute value are less than 0.291730 which is the value for significant serial correlation.

Table 6 Durbin-Watson Test for Serial Correlation of Corrected Model

Parameter	**Value**	**Did the Test Reject H0: Rho(1) = 0?**
Durbin-Watson Value	1.7886	
Prob. Level: Positive Serial Correlatio	0.2657	No
Prob. Level: Negative Serial Correlatio	0.6573	No

From the Durbin – Watson test for serial correlation there is no either positive or negative serial correlations among the variables, as the test reject the null hypothesis for serial correlation as shown in table 6 above.

Table 7 Predicted Values with Prediction Limits of Individuals for Cochrane Model

(a) Row	(b) Actual Y	(c) Predicted $\hat{Y}$	(d) Residual e_t	(e) Standard Error of Predicted	(f) $(e_t - e_{t-1})^2$	(g) 95% Lower Pred. Limit of Individual	(h) 95% Upper Pred. Limit of Individual
2	0.000	-	-	-	-	-	-
3	9.000	6.047	2.953	1.487	8.720	3.047	9.047
4	4.000	4.948	-0.948	1.460	15.218	2.002	7.894
5	4.000	5.268	-1.062 1.4	1.438	0.605	2.395	8.173
6	3.000	5.062	-2.062	1.430	1	2.177	7.948
7	4.000	5.284	-1.284	1.432	0.605	2.395	8.173
8	5.000	5.426	-0.426	1.421	0.736	2.558	8.294
9	0.000	4.018	-4.018	1.424	12.902	1.144	6.891
10	4.000	5.411	-1.411	1.461	6.796	2.463	8.359
11	5.000	5.435	-0.435	1.411	0.953	2.587	8.282
12	9.000	6.799	2.201	1.408	6.948	3.957	9.641
13	7.000	5.869	1.131	1.423	1.145	2.997	8.741
14	7.000	6.070	0.930	1.404	0.040	3.237	8.903
15	8.000	6.580	1.420	1.404	0.240	3.746	9.413
16	7.000	6.126	0.874	1.406	0.298	3.287	8.964
17	7.000	6.258	0.742	1.400	0.017	3.433	9.082
18	7.000	6.320	0.680	1.399	0.004	3.497	9.143
19	8.000	6.949	1.051	1.400	0.138	4.124	9.774
20	9.000	7.568	1.432	1.408	0.145	4.727	10.410
21	8.000	6.995	1.005	1.410	0.182	4.150	9.840
22	0.000	1.909	-1.909	1.425	8.491	-0.967	4.784
23	7.000	7.119	-0.119	1.449	3.204	4.194	10.043
24	15.000	12.420	2.580	1.491	7.285	9.412	15.429
25	0.000	0.984	-0.984	1.541	12.702	-2.126	4.095
26	7.000	7.306	-0.306	1.475	0.460	4.328	10.284
27	7.000	6.883	0.117	1.405	0.179	4.048	9.718
28	8.000	7.780	0.220	1.403	0.011	4.948	10.612
29	4.000	4.344	-0.344	1.402	0.318	1.515	7.174
30	6.000	6.384	-0.384	1.401	0.002	3.556	9.212
31	6.000	6.278	-0.278	1.396	0.011	3.460	9.096
32	7.000	7.265	-0.265	1.401	0.002	4.437	10.093
33	8.000	8.242	-0.242	1.414	0.005	5.389	11.095
34	7.000	7.251	-0.251	1.410	0.008	4.405	10.097
35	10.000	10.514	-0.514	1.448	0.069	7.591	13.437
36	7.000	7.237	-0.237	1.431	0.077	4.349	10.126
37	8.000	8.611	-0.611	1.422	0.140	5.742	11.480
38	3.000	2.969	0.031	1.423	0.412	0.097	5.840
39	3.000	3.259	-0.259	1.431	0.084	0.372	6.146
40	4.000	4.395	-0.395	1.418	0.018	1.533	7.257
41	3.000	3.076	-0.076	1.423	0.102	0.205	5.947
42	7.000	8.098	-1.098	1.440	1.044	5.193	11.004
43	5.000	5.319	-0.319	1.422	0.607	2.449	8.189
44	6.000	6.772	-0.772	1.426	0.205	3.894	9.651
45	0.000	-1.315	1.315	1.502	4.356	-4.347	1.717
46	0.000	-1.045	1.045	1.538	0.073	-4.149	2.060
47	8.000	10.021	-2.021	1.514	9.400	6.965	13.076
48	0.000	-1.893	1.893	1.540	15.319	-5.000	1.214

$$\sum(e_t)^2 = 75.417 \qquad \sum(e_t - e_{t-1})^2 = 120.6768$$

The residuals (e_t) of the model in table 7 above were got by subtracting the predicted ($\hat{Y}$) from the actual (Y), the 95% lower and upper limit is the confidence interval estimate of the mean of Y at that value of X.

Plot Section for Cochrane Model

$\boldsymbol{C_1 = Y, C_2 = X_1, C_3 = X_3, C_4 = X_4}$

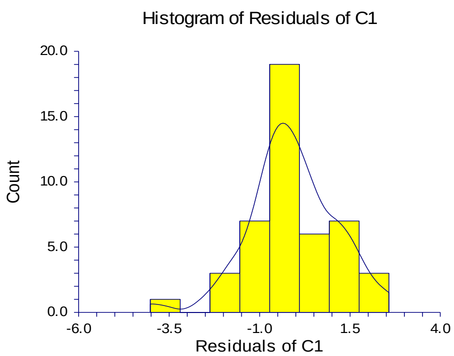


Figure 1: Histogram of Residuals

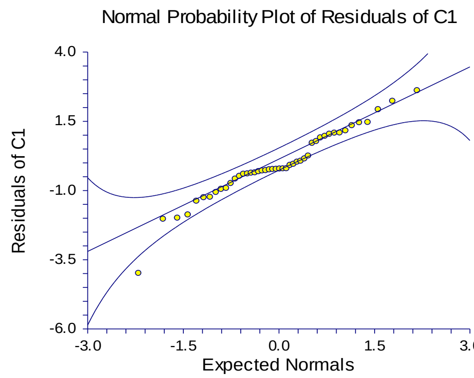


Figure 2: Probability plot of Residuals

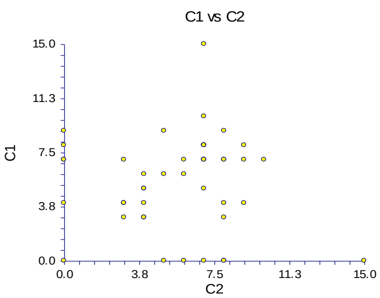


Figure 3: y vs. X_1

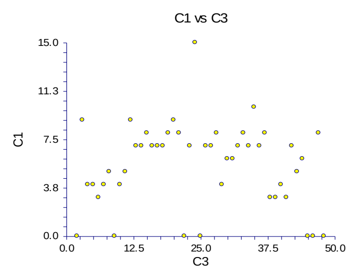


Figure 4: y vs. X_2

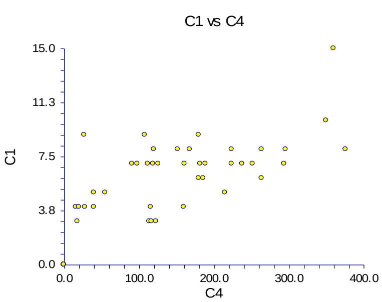


Figure 5: y vs. X_3

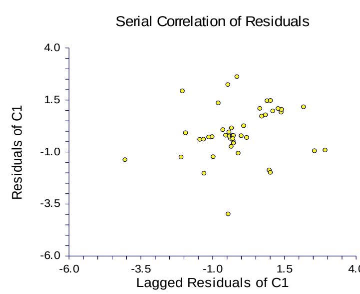


Figure 6: Lagged Residuals of y

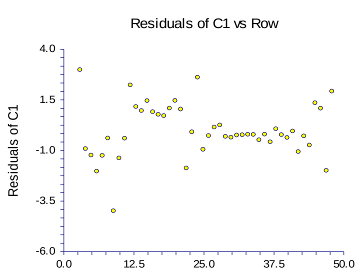


Figure 7: Residuals of y vs. Row

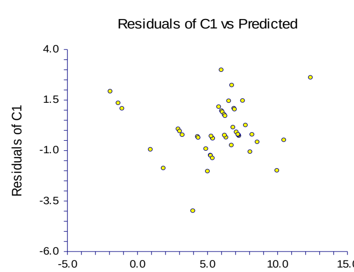


Figure 8: Residuals of y vs. Predicted

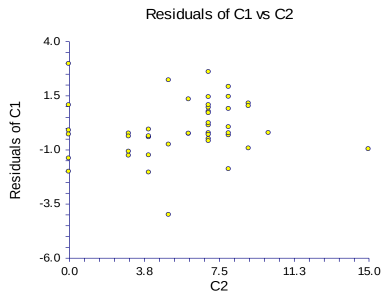


Figure 9: Residuals of y vs. X_1

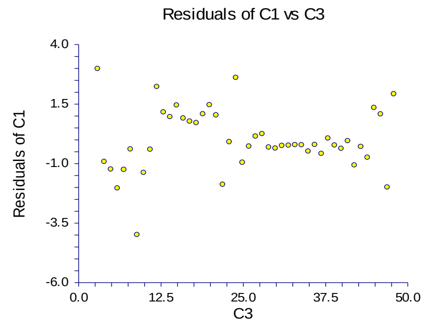


Figure 10: Residuals of y vs. X_2

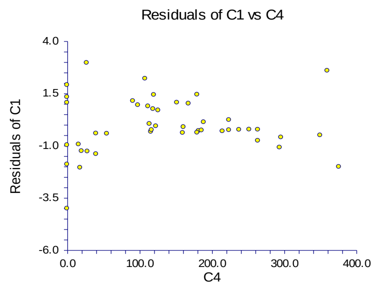


Figure 11: Residuals of y vs. X_3

From the histogram of residuals the errors are normally distributed as the curve of the plot follow range from negative values to positive, also the probability plot of the residuals follow the normal distribution as all the residuals fall within the normal curve and form a straight line about zero mean.

Table 8 Multiple Linear Regression Report including Four independent Variables

Run Summary Section Parameter	**Value**	**Parameter**	**Value**
Dependent Variable	Y	Rows Processed	48
Number Ind. Variables	4	Rows Filtered Out	0
Weight Variable	None	Rows with X's Missing	0
R^2	0.8009	Rows with Weight Missing	0
Adj. R^2	0.7819	Rows with Y Missing	0
Coefficient of Variation	0.2337	Rows Used in Estimation	48
Mean Square Error	2.261129	Sum of Weights	47.000
Square Root of MSE	1.503705	Completion Status	Normal Completion
Ave Abs Pct Error	26.685	Autocorrelation (Rho)	-0.1543

Four independents variables were used against single variable y in this method using the 48 month data collected from federal road safety corps and R – square, Adjusted R – square, Mean square error and Coefficient of variation were displayed in the table.

Table 9 Regression Equation Section

Independent Variable	**Regression Coefficient b(i)**	**Standard Error Sb(i)**	**T-Value to test H0:B(i)=0**	**Prob. Level**	**Reject H0 at 5%?**
Intercept	1.3078	0.4004	3.266	0.0022	Yes
X_1	0.0357	0.0842	0.424	0.6734	No
X_2	-0.0086	0.0364	-0.237	0.8142	No
X_3	0.0412	0.0441	0.934	0.3558	No
X_4	0.4184	0.0906	4.618	0.0000	Yes

Estimated Model

$$\hat{Y} = 1.3078 + 0.0357X_1 - 0.0086X_2 + 0.0412X_3 + 0.4184X_4 \qquad (3.2)$$

From the table above only X_4 are statistically significant as it's t – values of 4.618 is greater than the table value test of 2.018 which is used to calculate the confidence limit. And the probability level is less than the alpha at 5% level of significant.

Table 10 Regression Coefficient Section

Independent Variable	Regression Coefficient	Standard Error	Lower 95% C.L.	Upper 95% C.L.	Standardized Coefficient
Intercept	1.3078	0.4004	0.4997	2.1158	0.0000
X_1	0.0357	0.0842	-0.1342	0.2056	0.0707
X_2	-0.0086	0.0364	-0.0822	0.0649	-0.0608
X_3	0.0412	0.0441	-0.0479	0.1303	0.3581
X_4	0.4184	0.0906	0.2356	0.6013	0.5753

Note: The T-Value used to calculate these confidence limits was 2.018.
The error of the model can be accepted since all the standard errors of the model are less than the alpha at 5% level of significance.

Table 11 Analysis of Variance Section

Source	DF	R^2	Sum of Squares	Mean Square	F-Ratio	Prob. Level
Intercept	1		1945.175	1945.175	-	-
Model	4	0.8009	381.961	95.49024	42.231	0.0000
Error	42	0.1991	94.96741	2.261129	-	-
Total (Adjusted)	46	1.0000	476.9284	10.36801	-	-

From the table above the probability level of the model is 0.0000 which shows the goodness of the model for road accidents analysis in developing state.

Table 12 Durbin-Watson Test for Serial Correlation of Uncorrected Model

Parameter	Value	Did the Test Reject H0: Rho(1) = 0?
Durbin-Watson Value	2.1365	
Prob. Level: Positive Serial Correlation	0.6303	No
Prob. Level: Negative Serial Correlation	0.3232	No

Table 13 Durbin-Watson Test for Serial Correlation of Corrected Model

Parameter	Value	Did the Test Reject H0: Rho(1) = 0?
Durbin-Watson Value	1.9658	
Prob. Level: Positive Serial Correlation	0.4125	No
Prob. Level: Negative Serial Correlation	0.4848	No

From the Durbin - Watson test for serial correlation the variables are serially correlated as the null hypothesis for no serial correlation has been rejected.

Table 14 Pearson Correlations Section (Row-Wise Deletion)

	Y	X_1	X_2	X_3	X_4
Y	1.000000	0.684480	0.746567	0.811396	0.874853
X_1	0.684480	1.000000	0.805243	0.881402	0.653405
X_2	0.746567	0.805243	1.000000	0.957655	0.744169
X_3	0.811396	0.881402	0.957655	1.000000	0.813199
X_4	0.874853	0.653405	0.744169	0.813199	1.000000

There is a high correlation between the coefficients when all the variables were used to obtain a model.

Table 15 Predicted Values with Prediction Limits of Individuals for full Variable Model

(a) Row	(b) Actual Y	(c) Predicted $\hat{Y}$	(d) Residua e_t	(e) Standard Error of Predicted	(f) $(e_t - e_{t-1})^2$	(g) 95% Lower Pred. Limit of Individual	(h) 95% Upper Pred. Limit of Individual
1	7.000	7.334	-0.334	-	0.112	-	-
2	0.000	1.308	-1.308	1.553	0.949	-1.826	4.441
3	9.000	8.060	0.940	1.674	5.054	4.681	11.438
4	4.000	5.065	1.065	1.552	4.020	1.933	8.197
5	4.000	3.547	0.453	1.538	2.304	0.444	6.650
6	3.000	3.932	-0.932	1.538	1.918	0.829	7.035
7	4.000	4.547	-0.547	1.564	0.148	1.391	7.704
8	5.000	4.425	0.575	1.529	1.259	1.340	7.510
9	0.000	1.308	-1.308	1.564	3.546	-1.848	4.464
10	4.000	3.638	0.362	1.539	2.789	0.533	6.744
11	5.000	4.778	0.222	1.528	0.020	1.693	7.863
12	9.000	6.489	2.511	1.532	5.240	3.397	9.580
13	7.000	6.196	0.804	1.536	2.914	3.096	9.295
14	7.000	6.689	0.311	1.634	0.243	3.391	9.988
15	8.000	6.257	1.743	1.598	2.051	3.032	9.481
16	7.000	7.004	-0.004	1.578	3.052	3.820	10.187
17	7.000	6.202	0.798	1.553	0.643	3.067	9.337
18	7.000	5.847	1.153	1.557	0.126	2.705	8.989
19	8.000	8.169	-0.169	1.608	1.748	4.924	11.415
20	9.000	5.904	3.096	1.530	10.660	2.817	8.992
21	8.000	8.562	-0.562	1.736	13.381	5.059	12.066
22	0.000	1.308	-1.308	1.560	0.557	-1.841	4.457
23	7.000	4.842	2.158	1.555	12.013	1.705	7.979
24	15.000	13.250	1.750	1.751	0.166	9.716	16.784
25	0.000	1.308	-1.308	1.551	9.351	-1.823	4.438
26	7.000	5.095	1.905	1.551	10.323	1.964	8.225
27	7.000	7.062	-0.062	1.549	3.869	3.936	10.188
28	8.000	10.273	-2.273	1.600	4.889	7.045	13.501
29	4.000	4.792	-0.792	1.528	2.193	1.707	7.876
30	6.000	5.034	0.966	1.550	3.091	1.905	8.162
31	6.000	6.041	-0.041	1.542	1.014	2.929	9.154
32	7.000	5.766	1.234	1.544	1.626	2.650	8.882
33	8.000	9.031	-1.031	1.628	5.130	5.745	12.316
34	7.000	7.511	-0.511	1.604	0.270	4.275	10.747
35	10.000	8.476	1.524	1.629	4.141	5.189	11.764
36	7.000	12.872	-5.872	1.694	54.701	9.453	16.290
37	8.000	7.971	0.029	1.582	34.822	4.779	11.163
38	3.000	2.856	0.144	1.532	0.013	-0.236	5.949
39	3.000	2.939	0.061	1.542	0.007	-0.174	6.051
40	4.000	3.664	0.336	1.542	0.076	0.553	6.776
41	3.000	3.185	-0.185	1.537	0.271	0.083	6.287
42	7.000	6.329	0.671	1.541	0.733	3.220	9.439
43	5.000	5.924	-0.924	1.718	2.544	2.458	9.391
44	6.000	5.952	0.048	1.599	0.945	2.725	9.179
45	0.000	1.308	-1.308	1.557	1.839	-1.834	4.450
46	0.000	1.308	-1.308	1.573	0.000	-1.867	4.482
47	8.000	7.800	0.200	1.731	2.274	4.306	11.294
48	0.000	1.308	-1.308	1.554	2.274	-1.828	4.443

$\sum(e_t)^2 = 97.539$ $\sum(e_t - e_{t-1})^2 = 221.309$

Plots Section for Full Variable Model

$C_2 = Y, C_3 = X_1, C_4 = X_2, C_5 = X_3, C_6 = X_4$

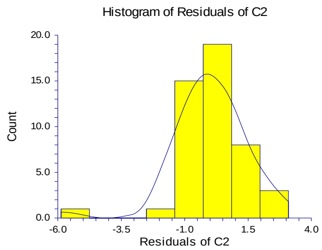


Figure 12: Histogram of Residuals of y

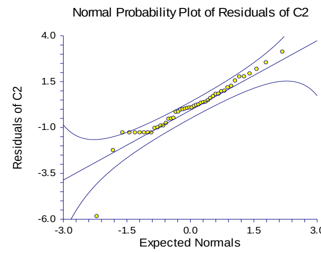


Figure 13: Probability plot of Residuals of y

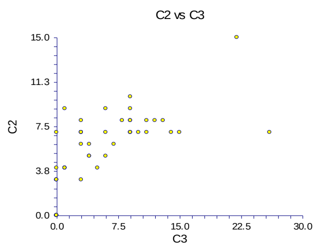


Figure 14: y vs. X_1

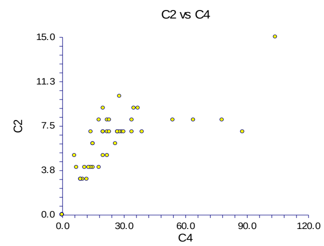


Figure 15: y vs. X_2

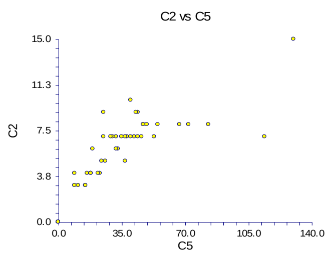


Figure 16: y vs. X_3

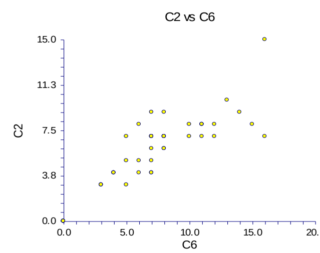


Figure 17: y vs. X_4

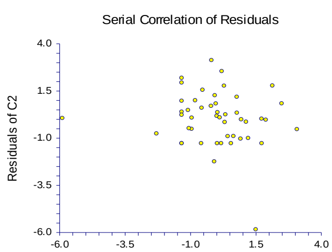


Figure 18: Lagged Residuals of y

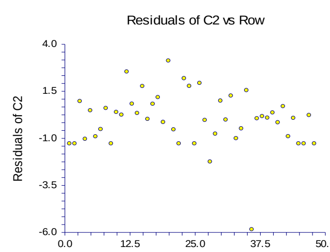


Figure 19: Residuals of y vs. row

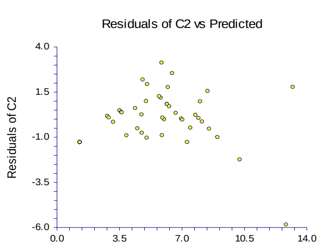


Figure 20: Residuals of y vs. predicted

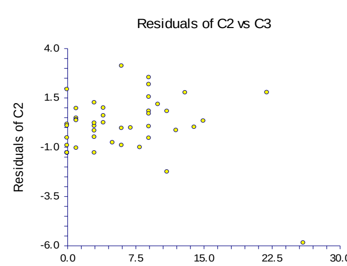
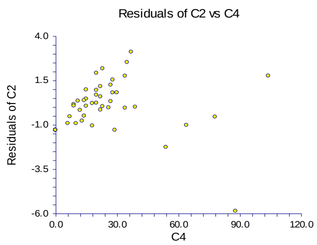
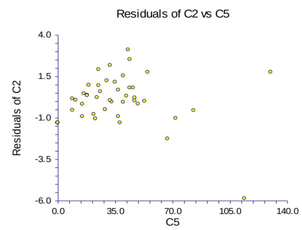


Figure 21: Residuals of y vs. X_1





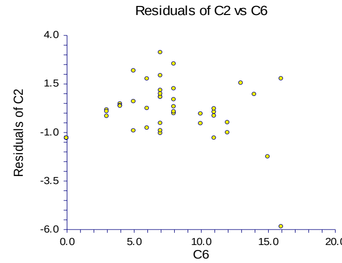


Figure 22: Residuals of y vs. X_2 **Figure 23: Residuals of y vs. X_3** Figure 24: Residuals of y **vs. X_4**

Note that, from the histogram of figure 12 the errors are normally distributed and the normality plot of figure 13 shown that all the variables are within the expected normal curve with only one variable is outside the expected normal. Furthermore, from the residuals plots there is high correlation among the independents variables.

4. Discussion

Two methods were employed in this work (Cochrane and Multiple linear Regression method) and two different Models were developed (3.1 and 3.2). The Cochrane method has the highest coefficient of determination (R^2) of 0.8251 and from the Durbin Watson test of serial correlation (table 6) there is no evidence of serial correlation of the variables, also in (table 2) X_3 and X_4 are statistically significant in the model as its absolute values are greater than the test value of 2.018. The multiple linear regressions method has the lowest coefficient of determination of 0.8009, from the test conducted (table 12 and 13) there is no serial correlation of the residuals, but only the coefficient of X_4 is statistically significant as its t – value of 4.818 is greater than 2.018 which is the value used to test the significant of variable in the model.

5. Conclusion and Recommendations

Two models were developed and compared in this paper, using Cochrane – Orcutt iteration method and Multiples linear regression method with full variables. In both methods X_4 (number of vehicles on traffic) is statistically significant in the models, which cause the rate of accidents (y) to be high. Therefore, increase in the number of vehicles on traffic (X_4) contribute significantly to the rate of accidents. The result obtained for the two methods were compared with help of graphs and the probability plot of the residuals indicates that the individual probabilities of the residuals are normally distributed since the points are near or closer to one another with the plotted points forming an approximately straight line. Based on the research finding the number of vehicles on traffic cause the rate of accidents (y) to be high, so followings recommendations are suggested;

1. Police traffic should be provided in all road junctions to control the traffic in the state
2. All hawkers should be withdrawn from the roads in order to reduce high traffic intensive in the state
3. Fly over should be constructed in four road junctions to minimize traffic congestion in the state.
4. High traffic volume can be reduce by constructing more roads in the state
5. The existing roads should be improve to two lane and three lane in order to reduce the traffic congestion on the roads.

References

[1] World Health Organization (2004), World report on road traffic injury prevention. World Health Organization, Geneva Switzerland. Pp 125 - 131

[2] Donatus L, Velmajasiniene O. (2011), Analysis of accidents prediction feasibility on the road of Lithuania; *International Conference on Road Safety and Simulation*, Indianapolis, USA pp (1-17).

[3] Ross, A. S. Lundebye and R. Barrett (1991) Road Safety Awareness and Commitment in Developing Countries. Infrastructure Note RD-6, World Bank, Washington DC, USA.

[4] FRSC and Balogun A, (2006), Research and Development Opinions, Federal Road Maintenance Agency; Pp 13-16.

[5] Olawale O. and Idowu A. (2011), Multivariate Regression Techniques for Analyzing Auto Crash Variables in Nigeria; *Journal of Natural Sciences Research* ISSN 2225-0921 (Online) Vol.1, No.1, 2011 233 - 241

[6] Andreassen D. (1991), Population and registered vehicle data and road deaths, accident anal Prev. 23(5): *Journal of Transportation and Statistics*, 4(1): 1 – 26.

[7] Adekunle J. Aderamo (2012), Assessing the trends in road traffic accidents casualties. *Journal of social science* 31(3) 19 – 25

[8] Gajendran C, Serin VK, Seenu SG, Swati P (2015), Different methods of accident forecast based on real data. *Journal of civil environmental engineering* 5(4) 43 – 49

[9] Liyanage T and Terrance M (2015), Traffic accident analysis and development of accident prediction model, *Annual sessions of IESL* pp 105 – 109.