Journal of Mathematical Problems, Equations and Statistics

E-ISSN: 2709-9407
P-ISSN: 2709-9393
JMPES 2022; 3(2): 134-137

www.mathematicaljournal.com
Received: 17-05-2022
Accepted: 10-07-2022

Ajibode IA
Department of Mathematics Statistics, The Federal Polytechnic, Ilaro, Nigeria

Agbolade Olumuyiwa O
Department of Mathematics Statistics, The Federal Polytechnic, Ilaro, Nigeria

Corresponding Author:
Ajibode IA
Department of Mathematics Statistics, The Federal Polytechnic, Ilaro, Nigeria

A comparison of ordinary least squares and weighted regression for road accident causality models

Ajibode IA and Agbolade Olumuyiwa O

Abstract
The examination of data on road accident fatalities was used in this study to determine which of the two estimators, the Weighted Least Square Estimator (WLS) and the Ordinary Least Square Estimator (OLS), is most effective. The Federal Road Safety Corps' official website provided statistics for this study's principal data source for a 50-year span, from 1972 to 2021, making it secondary in nature. In order to estimate the relevant parameters, the Ordinary Least Squares (OLS) and Weighted Least Squares (WLS) estimators were utilised. Numerous assessment measures were used to compare these estimators, including R-squared, Adjusted R-squared, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Mean Squared Error (MSE). The evidence indicates that, when these particular qualities are taken into account, WLS performs better than OLS. The conclusion that the weighted least squares (WLS) estimator is better than the ordinary least squares (OLS) estimate in the setting of traffic data is therefore logical. So, before putting the best model into practise, it is crucial to identify it.

Keywords: AIC, BIC, OLS, WLS, Accident, casualties

Introduction

The transport industry contributes significantly to the economy, influences human development and well-being, and is essential to the growth of every country (Rodrigue and Notteboom, 2009) [11]. The World Health Organization (WHO, 2009) claims that road travel is essential for enabling the movement of people and products, resulting in improved accessibility to jobs, economic markets, education, leisure, and healthcare. As a result, this benefits both the population's health directly and indirectly. Despite the numerous advantages, traffic accidents have a negative impact on each country's sociological, political, and economic development.

The World Health Organisation (WHO, 2010) [14] reported that automobile accidents were responsible for almost 1.3 million deaths. This number might be extended to mean that more than 3,000 people die per day. In addition, accidents occur in a significant number of minor injuries each year, between twenty and fifty million. These injuries contribute significantly to the total burden of impairments on a global level. Additionally, 80% of road accident deaths, or a considerable fraction thereof, occur in low- and middle-income nations, according to data done by the World Health Organisation (WHO). Given that just 33% of all registered cars in the world come from these countries, this is notable. One of the top causes of death for people between the ages of 5 and 44 is traffic-related injuries.

According to Nantulya and Reich (2002) [10], a significant proportion of road fatalities in low-income countries and areas such as Africa, Asia, the Caribbean, and Latin America are attributed to various road users including walkers, passengers, cyclists, motorcyclists, and occupants of buses and minibuses. Nevertheless, when considering the relative casualty rates, which account for fatalities in relation to any measure of exposure, the aforementioned geographical differences are not apparent. It is well acknowledged that those classified as vulnerable road users, including pedestrians, cyclists, and motorcyclists, are at a greater risk of fatality in vehicular accidents compared to passengers of motor vehicles (ETSC, 2003).

In addition to the loss of life and physical harm, road-related accidents give rise to significant levels of anguish and distress, with substantial financial burdens in terms of medical expenses and diminished productivity. The significant impact of highway safety on society has led to the implementation of expensive measures aimed at ensuring road safety, the establishment of legislation governing road use, and the development of various policies pertaining to the production of automobiles that meet safety standards. The estimated fraction of the worldwide economic effect resulting from motor vehicle collisions ranges from 1% to 3% of the Gross

National Product (GNP) of each nation, with an average total sum exceeding $500 billion. The reduction of road accidents and deaths has the potential to alleviate human suffering, foster economic progress, and reallocate resources for more productive purposes.

In the context of road traffic fatality estimates, the fundamental idea of the classic least square technique involves estimating the parameter vectors β by the minimization of the sum of squared errors. This process aims to identify the optimal function that aligns with the available data. When using the conventional least squares technique for regression fitting, it is common practise to exclude data points with substantial effect on the regression estimate (assuming the original data is accurate). However, this approach may result in the loss of valuable information when working with small sample sizes. The regression line will tend to shift in the direction of larger data points if this requirement is not satisfied. The regression model, according to Kong, Li, and Zhang (2010), has a poor degree of fitting accuracy, making it impossible to get the desired results.

Research studies often analyse the frequency of accidents, taking into account variables like traffic and weather conditions as well as the synergistic effects brought on by the combination of traffic and meteorological variables. Ahmed, Abdel-Aty, and Yu (2012b) [4] looked at the impact of geometrical, traffic, and climatic conditions on the incidence of accidents on highways. It has been shown that variables including decreased visibility, higher precipitation, and fluctuations in speed enhance the likelihood of accidents during the winter. Interestingly, the combination of low average speeds and less visibility during the dry season increases the risk of an accident.

To estimate the chance of accidents under varied traffic circumstances, Xu *et al*. (2012) [15] created accident risk models. The research found that different traffic situations and variables affecting traffic flow had different effects on safety. In this research, it was shown that in two traffic scenarios-congested traffic and the change from free flow to congested flow-the average downstream occupancy had a mitigating influence on accident probability. It was also highlighted that the general model predicted an increase in the chance of accidents.

In a research on the relationship between road traffic accidents in Nigeria, Kupolusi, Adeleke, Akinyemi, and Oguntuase (2015) [7] compared the analyses of least square regression and fixed effect panel data regression. The fixed effect panel data regression model, a particular sort of regression approach that is intended to provide reliable standard error estimates in the presence of heteroscedasticity, was described by the authors. The authors said that in their opinion, an estimate is resilient if it shows minimal sensitivity to changes in the underlying assumptions it depends on. The use of fixed effect techniques requires two essential types of data. First and foremost, it's crucial to evaluate the dependent variable for each participant at least twice. The measurements must be directly comparable, which requires that they have the same significance and unit of measurement. The relevant predictor variables (Regressors) must also show differences in value for a sizable portion of the sample on each of the two occasions. According to the findings of their investigation, the fixed effect panel data regression technique employing heteroscedasticity variance-covariance tools consistently yields appropriate estimates of the regression parameters among the many estimation methodologies studied.

According to Moutari *et al*. (2005) [9], the increasing expansion of traffic density is a substantial challenge for road traffic engineers since it calls for the resolution of both ongoing traffic congestion and safety issues. To simulate traffic accidents, the researchers employed macroscopic second-order models of traffic flow. Eulerian coordinates were used in the formulation of the model under study. The problem with fixed interfaces may be solved by using a Lagrangian coordinate representation of the model. The Lagrangian frame's incorporation of a coupling mechanism, however, requires the identification of the precise class of drivers who are more likely to be engaged in traffic accidents. Further research is needed in the fields of examining suitable coupling conditions and establishing well-posedness in the Lagrangian frame.

The Driver-Vehicle-Road-Environment (DVRE) system was used by Kurakina, Evtiukov, and Rajczyk (2018) [8] to perform research on the prediction of traffic accidents. The researchers created a model to evaluate any risks connected to auto accidents. The findings of the study showed that a visual approach to conflict situations enabled the gathering of information on the interactions between subsystem state variables. The approaches for evaluating potential risks that were utilized made it feasible to collect data on both actual and predicted factors related to the likelihood that fatalities and accidents would occur on the particular road stretch under investigation.

Real-time traffic and weather data integration with highways has recently been identified as a highly promising method. The incorporation of both accident and non-accident data in the analysis is a commonly used method in accident prediction. This same methodology has been employed in several other studies (Abdel-Aty and Pande, 2005; Abdel-Aty, 2007; Ahmed and Abdel-Aty, 2012; Yu and Abdel-Aty, 2013) [2, 1, 3, 16]. According to the research done by Ahmed *et al*. (2012) [3], it was shown that a greater chance of rear-end accidents may be caused by an increase in speed variation within a particular accident section combined with a decrease in average speed in the corresponding downstream segment. According to Ahmed and Abdel-Aty's (2012) [3] study, there is a positive correlation between the likelihood of an accident and the size of speed fluctuations, but a negative correlation between the likelihood of an accident and the average speed at the particular section where the accident occurs, specifically during the time period of 5 to 10 minutes prior to the accident occurrence. However, it is essential to remember that velocity fluctuation does not always represent a danger factor. Kockelman and Ma (2007) [6] claim that there is no evidence linking 30-second speed variations to an increased risk of collisions.

Research studies often analyse the frequency of accidents, taking into account factors like traffic and weather conditions as well as the synergistic effects brought on by the combination of these two elements. In a study done in 2012 by Ahmed, Abdel-Aty, and Yu, the researchers looked at how geometry, traffic, and weather conditions affected the incidence of accidents on highways. It has been shown that variables including decreased visibility, increasing precipitation levels, and fluctuations in speed enhance the likelihood of accidents during the winter. Interestingly, the combination of low average speeds and less visibility during the dry season increases the risk of an accident.

Theofilatos, Yannis, Kopelias, and Papadimitriou (2016) [12] provide forecasts regarding traffic accidents using a rare events modelling approach. The results of the modelling research revealed a negative association between the frequency of accidents and the natural logarithm of speed at the accident location. They also showed a reasonable statistical fit. Through the application of cutting-edge models, such as the rare-events logit, which is being used for the first time in the safety evaluation of motorways, this study intends to improve understanding of accident occurrence on highways. In light of this, King and Zeng (2001) [5] claim that the problem of underestimating event probability is successfully resolved.

Since premature deaths result in a decline in population, the frequency of road accidents is worrisome. However, it is crucial to create a reliable statistical model that can accurately predict the likelihood of traffic accidents involving cars and motorcycles. In a similar vein, there is a dearth of existing research that examines and compares various models pertaining to road accident casualties. Therefore, the goal of this study is to compare the effectiveness of weighted least squares (WLS) and ordinary least squares (OLS) regression models to see which is better at forecasting traffic accidents.

2. Material and Methods

The research mostly sourced its materials from secondary sources. The data was obtained from the annual reports of the Federal Road Safety Corps (FRSC) spanning the years 1972 to 2021. The dataset comprises information about the overall count of casualties resulting from road accidents, as well as the specific counts of road accidents attributed to cars and motorcycles.

2.1 The OLS Estimator

Ordinary least square regression is a statistical technique for predicting dependent variable Y based on measurements of independent variable X. It has a model of the form

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + e_i \qquad (1)$$

Where the X's are the explanatory variables and Y is the response variable. In this case,

Y = Road accident casualty

X_{1i}= Number of road accidents due to car

X_{2i} = Number of road accidents due to motorcycle

β_0 = Autonomous road accident casualty estimate when car and motorcycle accidents are held constant

β_1= Coefficient of road accidents due to car

β_2= Coefficient of road accidents due to car with weights attached

β_3 = Coefficient of road accidents due to motorcycle with weights attached

e_i = Error term

In addition, the data were subjected to the assumptions of the linear regression model. Hence, the model is estimated using:

$$\hat{B} = (X'X)^{-1} X'Y \qquad (2)$$

2.2 The WLS Estimator

The weighted least-squares method is used when the variance of errors is not constant, that is, when the following hypothesis of the least-squares method is violated: the variance of errors is constant (equal to the unknown value σ^2) for any observation i (that is, whatever the value of the concerned x_{ij}). Weighted Least Squares is an extension of Ordinary Least Squares regression. Non-negative constants (weights) are attached to data points. It is used when the data violates the assumption of homoscedasticity.

For simple linear regression model, the weighted least squares function is given as

$$\sum w_i e_i^2 = \sum w_i (y_i - \alpha - \beta x_i)^2 \qquad (3)$$

In this case, weights are added to equation 3

3. Result and Discussion of the Analysis

Table 1: Descriptive Statistics

Statistic	Casualty	Car	Motorcycle
Mean	31027.00	2663.580	2715.500
Median	32827.00	2336.000	3004.500
Maximum	47219.00	5775.000	4418.000
Minimum	16047.00	1153.000	1027.000
Std. Dev.	7826.177	1385.351	999.4903
Skewness	-0.014526	1.133130	-0.247265
Kurtosis	2.277444	3.102873	1.686154
Jarque-Bera	1.089440	10.72191	4.105732
Probability	0.580004	0.104696	0.128366
Observations	50	50	50

Source: Extracted from e-views Output, Version 10

The variables of measurements were jointly described as provided in Table 1. The results obtained from the Jarque-Bera test statistics provide confirmation that all of the variables exhibit a normal distribution. The relevant values for the variables, including the mean, median, minimum, and maximum, are effectively shown.

Table 2: Comparison between OLS and WLS Estimators for Road Accident Casualty

Variable	OLS Estimator	WLS Estimator
Road Accident Due to Car	3.258954* (0.694999) [0.0000]	3.764308* (0.699627) [0.0000]
Road Accident Due to Motorcycle	0.527890 (0.963308) [0.5863]	0.669004 (0.809955) [0.4163]
Constant	20913.03* (2822.372) [0.0000]	25466.09* (1895.156) [0.0000]
	R^2 = 0.3624 Adj. R^2 = 0.3353 F-statistic = 13.356 p-value = 0.0000 DW = 0.5099 AIC = 20.5322 BIC = 20.4618 MSE = 6380.852	R^2 = 0.57632 Adj. R^2 = 0.54373 F-statistic = 17.683 p-value = 0.0000 DW = 1.95376 AIC = 19.3573 BIC = 19.4987 MSE = 6380.852

Source: Extracted from E-views Output, Version 10, The bracket contains the standard error while the braces contains the p-value.

The comparative examination of the OLS and WLS metrics, as shown in table 2, suggests that among all the metrics included in both models, the weighted least square regression method was seen to be more effective in comparison to the ordinary least square regression strategy. The enhancement in the R-square and adjusted R-square of the weighted model provides evidence of this phenomenon. Furthermore, it was observed that the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) exhibited lower values in comparison to the Ordinary Least Squares (OLS) estimator. Additionally, there was an enhancement in the F-statistic and its corresponding p-value. The Durbin-Watson statistic of

1.9537 indicates that there is no significant serial correlation present in the residuals of the weighted regression model. This value is close to 2, further supporting the absence of serial correlation. The weighted least squares regression model had a lower mean squared error (MSE) of 3680.696, in contrast to the ordinary least squares (OLS) model which had an MSE of 6380.852. This result further substantiates the effectiveness of the model in accurately predicting road accident casualties. Therefore, the accurate representation of the impact of car accidents and motorcycle accidents on the casualties of road accidents in Nigeria is really in accordance with the anticipated outcome.

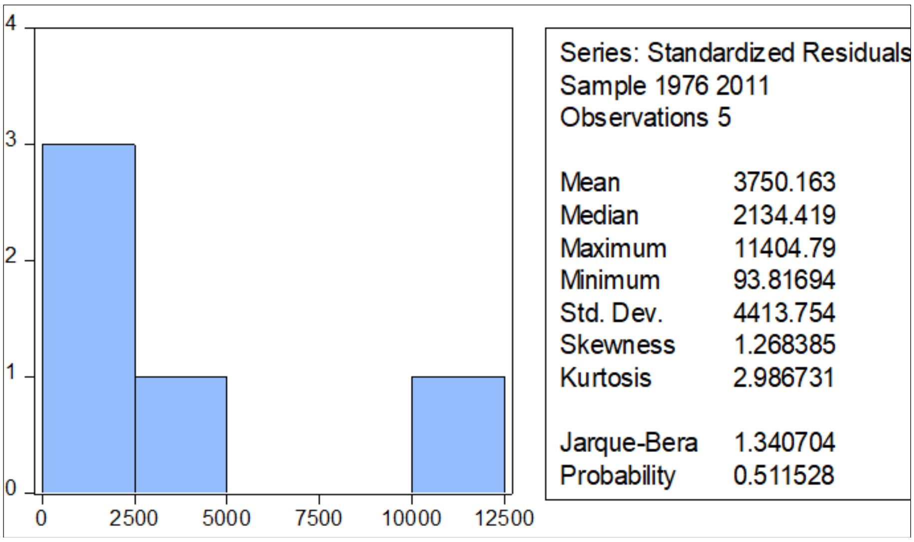

Fig 1: Residual Diagnostic of the weighted least square Model

4. Conclusion and Recommendation

The primary objective of this research is to conduct a comparative analysis between the Ordinary Least Squares (OLS) and Weighted Least Squares (WLS) estimators. The study focuses on examining the relationship between road accident casualties in Nigeria and the occurrence of road accidents involving cars and motorcycles. The aim is to establish the best suitable model for estimating this relationship between the two aforementioned estimators. The findings confirmed that the WLS estimate outperformed the OLS estimator in relation to the variables under investigation.

Therefore, it is recommended in this work that the use of an adequate estimator is necessary while conducting a study that focuses on models for predicting future occurrences.

5. References

1. Abdel-Aty M. Modeling left-turn crash occurrence at signalized intersections using various statistical models. Accident Analysis & Prevention. 2007;39(6):1120-1129.
2. Abdel-Aty M, Pande A. Exploring the safety effects of road features and traffic characteristics on rural two-lane highways. Accident Analysis & Prevention. 2005;37(5):913-924.
3. Ahmed MM, Abdel-Aty MA, Yu R. Investigating the effects of reduced visibility distance due to fog on crash injury severity. Journal of Safety Research. 2012;43(6):447-452.
4. Ahmed MM, Abdel-Aty MA, Yu R. Modeling traffic crash flow by time of day. Accident Analysis & Prevention. 2012b;45:55-63.
5. King G, Zeng L. Logistic regression in rare events data. Political Analysis. 2001;9(2):137-163.
6. Kockelman KM, Ma J. The impact of weather on freeway traffic flow and breakdown: A case study of San Antonio. Transportation Research Part A: Policy and Practice. 2007;41(3):294-309.
7. Kupolusi J, Adeleke A, Akinyemi O, Oguntuase O. Estimating road traffic accidents in Nigeria: Comparing least squares and fixed effect regression models. International Journal of Transportation Engineering. 2015;2(4):269-278.
8. Kurakina E, Evtiukov S, Rajczyk M. Predicting traffic accidents using the Driver-Vehicle-Road-Environment system. Proceedings of the Institution of Civil Engineers – Transport. 2018;171(4):173-183.
9. Moutari S, Zhang G, Yerra B. Traffic accident and road safety models using cellular automata. Transportation Research Part C: Emerging Technologies. 2005;13(3):211-235.
10. Nantulya VM, Reich MR. The neglected epidemic: Road traffic injuries in developing countries. BMJ. 2002;324(7346):1139-1141.
11. Rodrigue JP, Notteboom T. Transport modes and technologies. In The Geography of Transport Systems; c2009 Retrieved from https://transportgeography.org/contents/3-transport-modal-solutions/transport-modes-and-technologies/
12. Theofilatos A, Yannis G, Kopelias P, Papadimitriou E. Predicting motorway accident frequency using a rare events modeling approach. Accident Analysis & Prevention. 2016;92:108-116.
13. World Health Organisation (WHO). Global status report on road safety: Time for action; c2009. Retrieved from https://www.who.int/violence_injury_prevention/road_safety_status/2009/en/
14. World Health Organisation (WHO). Global status report on road safety: Time for action; c2010. Retriev from https://www.who.int/violence_injury_prevention/road_safety_status/2010/en/
15. Xu C, Li Z, Zhang HM. Investigation of factors affecting freeway accident occurrence using various analysis methods. Accident Analysis & Prevention. 2012;45:387-394.
16. Yu R, Abdel-Aty M. Utilizing support vector machine in real-time crash risk evaluation. Accident Analysis & Prevention. 2013 Mar 1;51:252-9.