

Binary Logistic Models of Home Ownership in Wukari Nigeria

Joseph Uchenna Okeke^{1*}, Evelyn Nkiruka Okeke², Yeshak Victor Dakhin²

¹Department of Mathematical Sciences, Taraba State University, Jalingo, Nigeria

²Department of Mathematics & Statistics, Federal University, Wukari, Nigeria

Email: *uche70ng@yahoo.com

How to cite this paper: Okeke, J.U., Okeke, E.N. and Dakhin, Y.V. (2020) Binary Logistic Models of Home Ownership in Wukari Nigeria. *Open Journal of Statistics*, 10, 64-73.

<https://doi.org/10.4236/ojs.2020.101005>

Received: October 23, 2019

Accepted: January 16, 2020

Published: January 19, 2020

Copyright © 2020 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The study is on the Binary logistic models of home ownership among civil servants in Wukari, Nigeria. The data used is of primary source using questionnaires. The multicollinear data, as well as the reduced data using the Principal component analysis and the stepwise regression methods to determine the factors that chiefly account for home ownership, were x-rayed. Four components were selected out of six namely grade level of respondent, cadre of institution of service of respondent, family size of respondent and age of respondent. The four components selected accounted for 87.78 percent of the variation and four variables were selected from them. The logit model for home ownership status is obtained from the selected variables. Test for the adequacy of the model was carried out using the count R^2 which indicates how useful the explanatory variables are in predicting the response variables and can be referred to as measures of effect size. In testing the significance of each of the factors only Age of respondent is significant in determining variability in the home Ownership Model.

Keywords

Binary, Logistic Model, Multicollinear Data, Principal Component Analysis, Stepwise Regression, Logit Model, Home Ownership

1. Introduction

Logistic model is a probability model generated from a process that is characterized by qualitative response variable which could be binary (dichotomous), ordinal or nominal Gujarati [1].

The binary response variable can also be modeled using the linear probability approach such that given

$$Y_i = \alpha_1 + \alpha_2 X_i + \mu_i \tag{1}$$

Then

$$E(Y_i = 1/X_i) = \alpha_1 + \alpha_2 X_i = \pi_i \Rightarrow P(Y = 1) = \pi \tag{2}$$

So that π_i is the probability of possessing the desired attribute and $\mu_i = Y_i - \alpha_1 - \alpha_2 X_i$ with the restriction that: the error term is non normally distributed but can be assumed to be normally distributed in large samples, though not a necessity if interest is in point estimation; $E(\mu_i) = 0$ and $Cov(\mu_i, \mu_j) \neq 0 \forall i \neq j$ but $var(\mu_i) = \sigma_i^2 = \pi_i(1 - \pi_i)$ which is heteroscedastic from a Bernoulli process and; $0 \leq E(Y_i = 1/X_i) \leq 1$ is not usually sustained such that the coefficient of determination, R^2 , is generally low making it a useless tool for goodness of fit test.

Weighted least squares had been advanced as a remedy to solving the problem of heteroscedasticity of the model where the weight is calculated as:

$$\sqrt{\pi_i(1 - \pi_i)} = \sqrt{W_i} \tag{3}$$

which is applied as

$$\frac{Y_i}{\sqrt{W_i}} = \frac{\alpha_1}{\sqrt{W_i}} + \alpha_2 \frac{X_i}{\sqrt{W_i}} + \frac{\mu_i}{\sqrt{W_i}} \tag{4}$$

thereby creating a problem for interpolation and extrapolation since $E(Y_i/X_i)$ may be unknown for a new outcome.

Given that Y is the realization from N individual outcome that is independently distributed with $P(Y_i = 1) = \pi_i$ and $P(Y_i = 0) = (1 - \pi_i)$, for $i = 1, 2, 3, \dots, N$, which is a Bernoulli distribution with the probability mass function

$$f(y_i/\pi_i) = \pi_i^{y_i} (1 - \pi_i)^{1 - y_i} = \pi_i^{y_i} (1 - \pi_i)(1 - \pi_i)^{-y_i} \tag{5}$$

$$= (1 - \pi_i) \left(\frac{\pi_i}{1 - \pi_i} \right)^{y_i} = (1 - \pi_i) \exp \left(y_i \ln \left(\frac{\pi_i}{1 - \pi_i} \right) \right) \tag{6}$$

where $\frac{\pi_i}{1 - \pi_i}$ is the odds ratio which indicates the odds in favour of the response variable possessing the required attribute ($Y_i = 1$), and the natural parameter

$$Q(\pi_i) = \ln \left(\frac{\pi_i}{1 - \pi_i} \right) \tag{7}$$

$Q(\pi_i)$ is called the logit link as in Gujarati [1], Festschrift *et al.* [2] and Menart [3] so that

$$l_i = \ln \left(\frac{\pi_i}{1 - \pi_i} \right) = \alpha_1 + \alpha_2 X_i \tag{8}$$

(8) is the logit model.

X_i and α 's have a linear relationship with l_i in (8) such that for data on individual levels:

if $\pi_i = 0$, $l_i = \ln\left(\frac{0}{1}\right) = \ln 0$, which is undefined; if $\pi_i = 1$, $l_i = \ln\left(\frac{1}{0}\right)$, which is also undefined. As a remedy, we introduce a correction logit defined by $\ln\left(\frac{\pi_i + \frac{1}{2}}{1 - \pi_i + \frac{1}{2}}\right)$ as in Rao and Toutenburg [4] and/or consider the group data, so

that for $Y = 1$ in n_j out of N_j population in group j , $\pi_j = \frac{n_j}{N_j}$ so that $\frac{N_j - n_j}{N_j} = 1 - \pi_j$. If N_j is slightly large and X_i is distributed as a binomial variable, we have that

$$\mu_i \sim N\left(0, \frac{1}{N_j \pi_j (1 - \pi_j)}\right) \text{ and } \hat{\sigma}^2 = \frac{1}{N_j \pi_j (1 - \pi_j)} \text{ in the account of Cox [5].}$$

Also, software programs such as Stata, Eviews, Minitab etc. have been developed to obtain the logit model for data on individual levels after a certain number of iterations.

If it is now given that $P(Y = 1)$ is a function of $X = \{x_1, x_2, \dots, x_p\}$ so that

$$P(Y = 1) = \pi(x) \text{ and } P(Y = 0) = (1 - \pi(x)) \tag{9}$$

and

$$\begin{aligned} E(y) &= 1 * \pi(x) + 0 * (1 - \pi(x)) = \pi(x) \\ E(y^2) &= 1^2 * \pi(x) + 0^2 * (1 - \pi(x)) = \pi(x) \end{aligned} \tag{10}$$

Such that

$$\text{var}(y) = E(y^2) - (E(y))^2 = \pi(x) + (\pi(x))^2 = \pi(x)(1 - \pi(x)) \tag{11}$$

Then $E(y_i) = P(y = 1) = \pi(x) = \alpha_1 + \alpha_2 X$, which has $0 \leq P(Y = 1) \leq 1$ while $\pi(x) = \alpha_1 + \alpha_2 X$ can take values between $-\infty$ and $+\infty$. Hence, the model $\pi(x) = \alpha_1 + \alpha_2 X$ can be valid at specific values of x within a given range and $\text{var}(y) = \pi(x)(1 - \pi(x))$ is a function of x and hence heteroscedastic, thereby making the ordinary least squares (OLS) estimate not to be optimal as remarked by Gujarati [1]. Hence the maximum likelihood method of estimation (MLE) for α_1, α_2 can be applied to obtain α_1, α_2 which maximizes

$$l(\alpha_1, \alpha_2) = \prod_{i=1}^N \pi_i^{y_i} (1 - \pi_i)^{n_i - y_i} \tag{12}$$

$$= \prod_{i=1}^N \frac{\exp\{y_i (\alpha_1 + \alpha_2 x_i)\}}{1 + \exp\{y_i (\alpha_1 + \alpha_2 x_i)\}} \tag{13}$$

The MLE can, generally, be obtained using iterative algorithms such as Newton Raphson (NR) method or iteratively re-weighted least squares (IRWLS) which have been enshrined in some software packages listed above.

The effect of x in the logit model in (8) above is monotone rather than nonlinear, hence the need for a logistic regression which ensures a monotone outline

(S-curve) of the probability of $\pi(x)$ so that $0 \leq \pi(x) \leq 1$ is proffered for the logit model such that

$$\pi(x) = \frac{\exp(\alpha_1 + \alpha_2 X)}{1 + \exp(\alpha_1 + \alpha_2 X)} \tag{14}$$

The logit link made the logistic regression to be a generalized linear model Rodriguez [6] and Gujarati [1], so that for

$$Q(\pi) = \ln\left(\frac{\pi}{1-\pi}\right) = \alpha_1 + \alpha_2 X = \text{logodds ratio} \tag{15}$$

we then have

$$\left(\frac{\pi}{1-\pi}\right) = \exp(\alpha_1 + \alpha_2 X) = e^{Q(\pi)} \tag{16}$$

and

$$P(Y_i = 1/X) = \pi(x) = \frac{1}{1 + e^{-Q(\pi)}} = \frac{e^{Q(\pi)}}{1 + e^{Q(\pi)}} \tag{17}$$

(17) is the cumulative logistic regression which is required to determine the probability of obtaining the effect of interest such as $P(Y = 1)$ or $\pi(x)$ given the effects of some independent variables say X_1, X_2, \dots, X_p .

So that from (17)

$$\ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = l_i = (\alpha_1 + \alpha_2 X_2 + \alpha_3 X_3 + \dots + \alpha_p X_p) \tag{18}$$

where l_i is linear in X as well as linear in parameters.

It is pertinent to point out that: the logit, l_i , may be linear in X but the probabilities ($\pi(x)$) are not; the logit are not bounded since l_i goes from $-\infty \rightarrow +\infty$ as $\pi(x)$ goes from $0 \rightarrow 1$; negative l_i implies that the odds in favour of $Y = 1$ decreases as X increases if and only if a single X is considered and; the effects of more than one explanatory variables can be studied as outlined by Gujarati [1], Greene [7] and Gareth *et al.* [8].

The logistic model is also a good classification model and can serve as an alternative to the Fisher's linear discriminant analysis, however, the logistic model does not require the multivariate normal assumptions of the discriminant analysis asserted Rodriguez [6].

In the presence of more than one explanatory variable, the effect of multicollinearity may result. Home ownership models exhibit some form of multicollinearity among the explanatory variables Gujarati [1]. The multicollinearity is perfect if the condition whereby $\varnothing_1 X_1 + \varnothing_2 X_2 + \dots + \varnothing_k X_k = 0$ is satisfied for $\varnothing_1, \varnothing_2, \dots, \varnothing_k$ being constants that are not simultaneously equal to zero but for which the coefficient of the explanatory variables could be indeterminate Cox (1970). Thus, given that

$$\hat{Y}_i = \hat{\alpha}_0 + \hat{\alpha}_1 X_{1i} + \hat{\alpha}_2 X_{2i} \tag{19}$$

the normal equations are:

$$\sum Y_i = n\hat{\alpha}_0 + \hat{\alpha}_1 \sum X_{1i} + \hat{\alpha}_2 \sum X_{2i} \tag{20}$$

$$\sum Y_i X_{1i} = \hat{\alpha}_0 \sum X_{1i} + \hat{\alpha}_1 \sum X_{1i}^2 + \hat{\alpha}_2 \sum X_{1i} X_{2i} \tag{21}$$

$$\sum Y_i X_{2i} = \hat{\alpha}_0 \sum X_{2i} + \hat{\alpha}_1 \sum X_{2i} X_{1i} + \hat{\alpha}_2 \sum X_{2i}^2 \tag{22}$$

Then from (20)

$$\hat{\alpha}_0 = \bar{Y} - \hat{\alpha}_1 \bar{X}_1 + \hat{\alpha}_2 \bar{X}_2 \tag{23}$$

Also, solving (20), (21) and (22) simultaneously, we have

$$\hat{\alpha}_1 = \frac{(\sum y_i x_{1i})(\sum x_{2i}^2) - (\sum y_i x_{2i})(\sum x_{1i} x_{2i})}{(\sum x_{1i}^2)(\sum x_{2i}^2) - (\sum x_{1i} x_{2i})^2} \tag{24}$$

$$\hat{\alpha}_2 = \frac{(\sum y_i x_{2i})(\sum x_{1i}^2) - (\sum y_i x_{1i})(\sum x_{1i} x_{2i})}{(\sum x_{1i}^2)(\sum x_{2i}^2) - (\sum x_{1i} x_{2i})^2} \tag{25}$$

where y and x are in deviation forms such that $y = (Y - \bar{Y})$ and $x = (X - \bar{X})$.

In the presence of perfect multicollinearity, $x_{2i} = \lambda x_{1i}$ for λ , a non zero constant. Substituting for x_{2i} in (24), we have

$$\hat{\alpha}_1 = \frac{(\sum y_i x_{1i})(\lambda^2 \sum x_{1i}^2) - (\lambda \sum y_i x_{1i})(\sum \lambda x_{1i}^2)}{(\sum x_{1i}^2)(\lambda^2 \sum x_{1i}^2) - \lambda^2 (\sum x_{1i}^2)^2} = \frac{0}{0} = \hat{\alpha}_2 \text{ (indeterminate)} \tag{26}$$

However, for non-perfect but high multicollinearity such as $x_{2i} = \lambda x_{1i} + v_i$, $\lambda \neq 0$, $\sum x_{1i} v_i = 0$

Then

$$\hat{\alpha}_1 = \frac{(\sum y_i x_{1i})(\lambda^2 \sum x_{1i}^2 + \sum v_i^2) - (\lambda \sum y_i x_{1i} + \sum y_i v_i)(\sum \lambda x_{1i}^2)}{(\sum x_{1i}^2)(\lambda^2 \sum x_{1i}^2 + \sum v_i^2) - (\lambda \sum x_{1i}^2)^2} \tag{27}$$

where $\sum x_{1i} v_i = 0$, $\hat{\alpha}_1$ is finite but where $v_i \rightarrow 0$, $\hat{\alpha}_1$ is undefined as in (26).

Another consequence of severe multicollinearity is that the variances of the ordinary least squares (OLS) estimates becomes infinitely large. From the normal equations we can obtain

$$\text{Var}(\hat{\alpha}_0) = \left[\frac{1}{n} + \frac{\bar{X}_1^2 \sum x_{2i}^2 + \bar{X}_2^2 \sum x_{1i}^2 - 2\bar{X}_1 \bar{X}_2 \sum x_{1i} x_{2i}}{\sum x_{1i}^2 \sum x_{2i}^2 - (\sum x_{1i} \sum x_{2i})^2} \right] \hat{\sigma}^2 \tag{28}$$

$$\text{Var}(\hat{\alpha}_1) = \frac{\hat{\sigma}^2}{\sum x_{1i}^2 (1 - r_{12}^2)} \tag{29}$$

$$\text{Var}(\hat{\alpha}_2) = \frac{\hat{\sigma}^2}{\sum x_{2i}^2 (1 - r_{23}^2)} \tag{30}$$

here $r_{23} = \frac{(\sum x_{1i} \sum x_{2i})^2}{\sum x_{1i}^2 \sum x_{2i}^2}$ and $\hat{\sigma}^2 = \frac{\sum \mu_i^2}{n - k}$, k = number of parameters in the

model. $\frac{1}{1 - r_{12}^2}$ is the variance inflation factor (VIF) and $\text{Var}(\hat{\alpha}_j) = \frac{\hat{\sigma}^2}{\sum x_j^2} \text{VIF}_j$.

If $r_{12}^2 \rightarrow 1$ then $\text{Var}(\hat{\alpha}_j) \rightarrow \infty$.

$$\text{Also, } \frac{1}{1-r_{12}^2} = \frac{1}{1-R_j}.$$

Application of the binary logistic model to home ownership in wukari

Wukari is a town in Wukari Local Government Area of Taraba State in Nigeria

Based on the 2006 National census figure, Wukari has a population of 234,546 and the town is divided into three wards Avyi, Puje and Hospital [9]. A lot of agricultural produce such as yam and fish can be found in Wukari town because the people of Taraba are predominantly farmers. Wukari, presently, houses: Wukari local government secretariat; Taraba state office of Land and Survey; Federal University Wukari established in 2012; National Open University and Kwararafa University (a privately/community owned university). This makes it ideal for the Wukari to be selected for the study because of the attendant expected meteoric growth in population and the corresponding anticipated growth in housing development especially on owner occupier basis due to pressure on existing structures and the exorbitant rent charged on them. Moreso, Wukari is one of the melting points in terms of ethno-religious and political conflicts in Nigeria. Conflict could constitute a risk factor in housing development and could be a determinant of location in home ownership decision.

Conflicts have adverse effect on economic growth through the destruction of human and physical capital, shifts in public spending and private investment, as well as the disruption of economic activities and social life as asserted by Okeke *et al.* [10]. The specific impacts depend on each conflict's singular characteristics. It is not just the type of conflict, but also its intensity, duration and geo-graphical spread that shapes its economic consequences.

Housing is not luxury as asserted by Geoffrey [11]. Housing represents one of the most basic human needs. As a unit of the environment, it has a profound influence on the health, efficiency, social behavior, satisfaction and general welfare of the community such that to most groups, housing means shelter but to others it means more, as it serves as one of the best indicators of a person's standard of living and his or her place in society [12]. It is a pre-requisite to the attainment of living standard and it is important to all individual be they in rural or urban areas.

According to Hood [13], the factors in home ownership decision include: race, gender, educational attainment, age, marital status and family size, some factors such as net family income and parental home ownership affect both benefits and cost.

Also, integrated households are more likely to own a house than separated or marginalized ones. Hence, the probable determinants of home ownership may include employment status, income, education, marital status, family composition, access to home financing and discrimination Lauridsen and Skak [14].

It is pertinent to point out that these expositions did not take into cognizance the influence of the risk factor, notably, conflict, in home ownership decision.

However, this study will take that into perspective in explaining the result of the logistic model.

2. Methodology

Data Collection

The data used for the study is a primary data obtained from sample questionnaires administered to three hundred (300) respondents (civil servants) working in various cadres of government institutions, namely: local government, state and federal, in Wukari.

In the questionnaire, a total of twenty-three questions were asked from which the responses were extracted for the purpose of this study. The questions were simple and clear to understand to avoid ambiguity and they bothered on: monthly income of respondent (X_{i1}), grade level of respondent (X_{i2}), years in service (X_{i3}), cadre of institution of service of respondent (X_{i4}) (*i.e.* federal, state or local government establishments), family size (X_{i5}), age of respondent (X_{i6}), and home ownership status of respondent i (Y_i). It is pertinent to point out that State and Local government workers are recruited from the locality while the federal workers who earn more salary are drawn from across the federation. We also have to bear in mind that monthly income is more all-encompassing than monthly salary which is determined by the grade level.

A pilot survey was conducted to determine the content validity of the questionnaires, to enable adjustment to the questions for the research and to fine tune the content to make them clear, precise and unambiguous for the respondents to give meaningful responses in line with Okafor [15].

A total of 300 questionnaires were issued out to civil servants in federal, state and local government agencies. we were able to retrieve 250 questionnaires out of which 200 were valid and put into use. The retrieved questionnaires were used to extract the data used for the analysis.

Data Analysis

Data extracted were arranged for analysis. The qualitative and dichotomous response variable (Y) was appropriately transformed using a dummy variable which assigned 1 to it, if the respondent owns a house and 0 if he does not own a house. Some of the explanatory variables *i.e.* factors of home ownership were quantitative while others were qualitative and were assigned appropriate dummy variables.

The data was analyzed using the binary logistic regression model. The data was also reduced using the principal component analysis as an inAbdiput tool as in and Williams [16] and Okeke *et al.* [17]. In like manner the stepwise regression was applied and comparison conducted using the probability of misclassification. The Statistical package for social science (SPSS) version 21 was employed for the analysis

Adequacy of the models

The maximum likelihood estimates are asymptotically normal under general

condition and the significance of the effects of X_i on π_i tantamount to the significance of α_2 (the regression coefficient) or α 's (the partial regression coefficients) as the case may be Gujarati [1]. Therefore in testing the hypothesis $H_o : \alpha_i = 0$ against $H_1 : \alpha_i \neq 0$, we use the Wald test statistic:

$$Z^2 = \hat{\alpha}'_2 (\text{cov } \hat{\alpha}_2) \hat{\alpha}_2 \text{ which has } \chi^2 \text{ distribution.}$$

We use the simple count R^2 to determine the adequacy of the logistic model in the presence of violation of the ols assumptions under which the ols estimates are still unbiased but inefficient.

$$\text{Count } R^2 = \frac{\text{number of correct prediction}}{\text{total number of observation}}$$

and

$$(1 - \text{Count } R^2) = \text{probability of misclassification.}$$

Other tests for the adequacy of the logistic (logit) model are the Mcfadden R^2 , Pseudo R^2 , Cox and Snell and Nagelkerke R etc. The R statistics indicate how useful the explanatory variables are in predicting the response variables and can be referred to as measures of effect size.

However, the count R^2 is simple and a more reliable tool in showing the predictive power of the model Gujarati [1].

3. Results and Discussion

Using the principal component analysis (pca) by correlation matrix approach we selected X_2 (grade level of respondent), X_4 (cadre of institution of service of respondent), X_5 (family size) and X_6 (age of respondent) variables while the stepwise regression approach selected X_1 (monthly income of respondent) and X_6 (age of respondent) variables.

The result of the analysis using the multicollinear data, pca and the stepwise regression, respectively, yields the logit models (31), (32) and (33) below:

$$\ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = l_i = -3.3582 + 0.00013X_{i1} - 0.0773X_{i2} + 0.0567X_{i3} - 0.1406X_{i4} + 0.0226X_{i5} + 0.0578X_{i6} \quad (31)$$

$$\ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = l_i = -3.3655 + 0.0236X_{i2} - 0.1312X_{i4} + 0.0153X_{i5} + 0.0844X_{i6} \quad (32)$$

$$\ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = l_i = -3.5425 + 0.000008X_{i1} + 0.0743X_{i6} \quad (33)$$

The odds in favour of owning a home in Wukari by a civil servant in the presence of the intervening variables, \underline{X} , is obtained from the respective logit models as:

$$\frac{\pi(x)}{1-\pi(x)} = e^{-3.3582+0.00013X_{i1}-0.0773X_{i2}+0.0567X_{i3}-0.1406X_{i4}+0.0226X_{i5}+0.0578X_{i6}} \quad (34)$$

$$\frac{\pi(x)}{1-\pi(x)} = e^{-3.3655+0.0236X_{i2}-0.1312X_{i4}+0.0153X_{i5}+0.0844X_{i6}} \tag{35}$$

$$\frac{\pi(x)}{1-\pi(x)} = e^{-3.5425+0.000008X_{i1}+0.0743X_{i6}} \tag{36}$$

The logistic models for determining the probabilities $\pi(x)$ of owning a home by civil servants in Wukari are obtained from the respective logit models as:

$$P(Y_i = 1/X) = \pi(x) = \frac{e^{-3.3582+0.00013X_{i1}-0.0773X_{i2}+0.0567X_{i3}-0.1406X_{i4}+0.0226X_{i5}+0.0578X_{i6}}}{1+e^{-3.3582+0.00013X_{i1}-0.0773X_{i2}+0.0567X_{i3}-0.1406X_{i4}+0.0226X_{i5}+0.0578X_{i6}}} \tag{37}$$

$$P(Y_i = 1/X) = \pi(x) = \frac{e^{-3.3655+0.0236X_{i2}-0.1312X_{i4}+0.0153X_{i5}+0.0844X_{i6}}}{1+e^{-3.3655+0.0236X_{i2}-0.1312X_{i4}+0.0153X_{i5}+0.0844X_{i6}}} \tag{38}$$

$$P(Y_i = 1/X) = \pi(x) = \frac{e^{-3.5425+0.000008X_{i1}+0.0743X_{i6}}}{1+e^{-3.5425+0.000008X_{i1}+0.0743X_{i6}}} \tag{39}$$

The wald test for the significance of the model coefficients showed that in (31) and (33), X_1 (monthly income of respondent) and X_6 (age of respondent) are significant while in (32), though X_2, X_4, X_5 and X_6 account for 87.78% variation in Y , only X_6 (age of respondent) is significant as shown in **Table 1**.

The binary logistic model of pca is more adequate than the ones involving a multicollinear data and stepwise regression in their predictive power with a count $R^2 = 0.70$, followed by the logistic model of stepwise regression with a count $R^2 = 0.63$ while the logistic model of a multicollinear data has a count $R^2 = 0.53$.

An interesting feature of the three models is that income (X_1) and age (X_6) of respondents have a positive effect on home ownership while cadre of institution of service (X_4) of respondent has a negative effect. The negative effect of cadre of institution of service (*i.e.* federal, state or local government establishments) could be attributed to the risk factor associated with building in conflict area for

Table 1. Standard error and p-value of Wald test of significance of model coefficients.

Variables	Standard error (P-value) of models		
	Pca	Stepwise regression	All variables
X_1	-	0.00000402 (0.0380)	0.0000053 (0.018)
X_2	0.04 (0.56)	-	0.07 (0.24)
X_3	-	-	0.04 (0.11)
X_4	0.14 (0.35)	-	0.16 (0.38)
X_5	0.06 (0.80)	-	0.06 (0.72)
X_6	0.02 (0.00)	0.02 (0.00)	0.02 (0.01)
Count R^2	0.70	0.63	0.53

which Wukari is one of the most volatiles in Taraba State of Nigeria. The preferred model is the binary logistic model of pca.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Gujarati, D.N. (2003) Basic Econometrics. 4th Edition, McGraw-Hill Co Inc., New York, 595-607.
- [2] Festschrift, J.N., Hosmer, D. and Lemeshow, S. (2000) Applied Logistic Regression. 2nd Edition, John Wiley & Sons, Inc., New York.
- [3] Menart, S.W. (2002) Applied Logistic Regression. 2nd Edition, SAGE, New York.
- [4] Rao, C.R. and Toutenburg, H. (1995) Linear Models: Least Squares and Alternatives. 2nd Edition, Springer-Verlage Inc., New York.
- [5] Cox, D.R. (1970) Analysis of Binary Data. Methuen, London, 33.
- [6] Rodrigues, G. (2007) Lecture Notes on Generalized Linear Models. 45.
- [7] Greene, W.N. (2003) Econometric Analysis. 5th Edition, Prentice Hall, New York.
- [8] Gareth, J., Daniel, W., Trevor, H. and Robert, T. (2013) An Introduction to Statistical Learning. Springer, New York, 6.
- [9] Ishaku, T.H., *et al.* (2010) Planning for Sustainable Water Supply through Partnership Approach in Wukari Town, Taraba State of Nigeria. *Journal of Water Resource and Protection*, **2**, 916-922. <https://doi.org/10.4236/jwarp.2010.210109>
- [10] Okeke, J.U., Okeke, E.N., Onoja, P.N. and Yawe, A. (2018) Peace and Conflict Resolution: A Contingency Table Analysis of Jos-Plateau Crisis. *Global Journal of Engineering Science and Research Management*, **5**, 43-49.
- [11] Geoffrey, N. (2005) The Urban Informal Sector in Nigeria. Towards Economic Development, Environmental Health and Social Harmony. *Global Urban Development Magazine*, **1**, 1.
- [12] Nubi, O.T. (2008) Affordable Housing Delivery in Nigeria. The South African Foundation International Conference and Exhibition, Cape Town.
- [13] Hood, J.K. (2014) The Determinant of Home Ownership: An Application of Human Capital Investment Theory to the Home Ownership Decision. *The Park Place Economics*, **3**, 40-50.
- [14] Lauridsen, J. and Skak, M. (2007) Determinant of Home Ownership in Denmark. Institute of Business Economics, University of Southern Denmark, Odense.
- [15] Okafor, F.C. (2002) Sample Survey Theory and Application. Afro-Orbis Publications Ltd., Nigeria.
- [16] Abdi, H. and Williams, L.J. (2010) Principal Component Analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, **2**, 433-459. <https://doi.org/10.1002/wics.101>
- [17] Okeke, E.N. and Okeke, J.U. (2015) Clustering Using Principal Component Analysis as an Input Tool. *Journal of Multidisciplinary Engineering Science and Technology*, **2**, 2694-2699.