Modelling Incidents of Building Collapse Data in Nigeria, Using Poisson and Negative Binomial Regression Models

N.L Nwakwasi¹, B.B Henshaw² and H.U Nwoke¹

¹Department of Civil Engineering SEET, Federal University of Technology Owerri, P.M.B 1526, Imo State, Nigeria.
²Department of Statistics, University of Ibadan, Ibadan, Nigeria

Corresponding Author: N.L Nwakwasi

Abstract: The rates of building collapse in Nigeria in the past twenty years have assumed a very worrisome dimension as many lives have been lost and huge investments wasted. Several causes have been attributed to natural or man-made phenomena. Survey carried out have shown that the use of substandard building materials, poor workmanship, non-enforcement of building codes or construction regulations, corruption in the building industry etc. have contributed immensely in most of the recorded cases of building failures. This paper is set to examine the application of Poisson and negative binomial regression models to identify the major factors that contributes to building collapse in Nigeria and to ascertain the better model suitable for prediction of building collapse or failure in Nigeria, using a secondary data which span from (1995 – 2016). Maximum likelihood estimation procedure was used to estimate the parameter of the selected model. With the number of incidents of building collapse data as the response variable (Y) and 7-explanatory variable (X’s), applying the forward selection criteria to the 7-explanatory variables, model 7 is best suitable for forecasting the subject under study. The result of the Poisson regression model showed that there was over dispersion in the incident of building collapse data since the dispersion parameter (3.224) was greater than 1. Hence underestimating the standard error and over estimating the coefficient of the explanatory variable, consequently giving misleading inferences. The result of the assessment criteria for Poisson regression model and Negative binomial regression model revealed that the Negative binomial regression models incident of building collapse data better in Nigeria as considered in this study. Use of quacks (UOQ), Change of Use (COU) and Corruption/greed (C_G) are the major contributors to incident of building collapse in Nigeria, with corruption/greed ranking first. A model suitable for prediction of the incident of building collapse in Nigeria has been developed.

Keywords: Building collapse, building materials, Poisson Regression Model, Negative Binomial Regression Model, Akaike Information Criteria (AIC).

I. INTRODUCTION

Building collapse, though a common phenomenon all over the world is more rampant and devastating in the developing countries. However, most of the failures were not caused by an earthquake or a terrorist attack, but rather by the poor construction and lack of over-sight Alamu and Gana (2014) added that the menace has cast a slur on the competence of the nation building community of professionals responsible for designing and monitoring construction works at building sites. Experts have also suggested that professional should not bear the blame alone. This is because it has been proved that owners should not bear the blame alone. Owner of building under construction derail from the approved plans relying more on imagination and fantasy. The approving authorities are also known to fail to monitor compliances with approved plans. Some building owners also shun professionals in order to cut cost. Also, the high cost of building materials has led greedy contractors with eyes on profits to patronize substandard material these short-cut measures has contributed immensely to the occurrence of failed buildings in the country.

Fademiro (2002) opined that the concentration of population in cities due to migration created by specific problems of providing adequate building structure. The resultant effect of this is the upsurge of various building projects from various contracting firms of doubtful competencies. Nnadozie (2005) also added that the development control unit (DCU) is most states play a limited role as they lack the requisite number of staff with expertise to supervise the many developments in progress concurrently across their areas of jurisdiction.

Ede (2010) carried out analysis and evaluation of the death rate involved in the 47 reported cases of building collapse verified between 2000 and 2010 and over 300 deaths were recorded for Lagos, Abuja and Port Harcourt which are the three major areas with high rate of casualties.
According to Obiedina (2005), the term failure can be manifested in three forms in the structure, namely: Ultimate limit state, which results in collapse; Serviceability limit state, which has to do with deflections and cracks and finally Durability which has to do with deterioration of the elements. The design of any structure must guard against the structure attaining any of this state. Furthermore, failure can occur in the sub or superstructures or both and can result from: poor design, use of substandard materials and poor-quality control and supervision.

Oloyede, Omoogun and Akinjare (2010) pointed out that different soil types pose varying problems for built foundations and the structural integrity of an entire building. Therefore, to design a foundation it is necessary to calculate the loads on the foundation and determine the nature of the sub-soil its bearing capacity, likely behavior under seasonal and ground water level changes and the possibility of ground movement.

Similarly, the components of the structure should be able to resist deformation under loading conditions. Deformation implies a change a change in size and shape when a body is subjected to stress. This means that the component should possess adequate stiffness. Thus, the stiffness of a beam or column is a measure of its resistance to bending or buckling.

2. A brief reviews/summary of the Causes of building failures or Collapses.

Generally, causes of building collapse can be classified under the following major factors, namely: types and quality of building materials used; mode/method of operation in the building industry and problems emanating from personal/human factor.

(i) Types and quality of building materials and Equipment
- The use of substandard building material of different qualities ca be obtained by using its constituent namely
  - Poor Concrete mix ratio. Cement water time and coarse aggregates, and mineral additive in different and right proportion (Gambir, 2005). A mix ratio that does not meet approved standard will have negative effect on the structure.
  - Poor construction equipment. Construction plants ranging from small hand plant such as mechanical excavators and tower cranes can be considered for use to maintain the high standards required in any building project. Sad enough, the use of manual labour is still a common practice in most of the of the construction sites in Nigeria.

(ii) Operation Problem
- Illegal conversion or change of use buildingowners in order to make more money from their property often change the use for which the building was initially approved by the town planning authority, adequate provisions are not often made for subsequent conversion/modification. Structures approved as office complex are illegally turned to a place of worships. The resultant effect is that the structure is eventually overloaded.
  - No structural drawing/design available
  - Non-enforcement of building codes or construction regulations.
  - Hasty construction: Tyler (1981) noted that the growth of strength with age is an important factor in concrete works.

Testing is usually done at 7days and 28days and concrete is regarded as more or less mature at 28 days. Sometimes, the project owner is ill advised on the consequence of loading premature concrete especially when he is in hast to get his project completed.

Before the records time.

(iii) Personal Problem
- Poor workmanship and ineffective supervision: Supervision involves the intricate knowledge of workmanship and to ensure adherences to contract documents, especially the drawings. This aspect is often lacking in most construction sites resulting in poor workmanship and subsequent collapse.
- The use of quacks instead of professionals: the activities of unqualified persons in the building industry have done a lot of harm. Building owners in most cases will engage the services of quacks in other to save cost of engaging professionals. A professional in practice is basically responsible for proving solution to problems in the field in the context of scope, functionality, quality and cost, stability, time of production, safety and durability.
- Corruption and greed in the building industry.
One of the major objectives of this study is to provide a model which will be able to predict the major contributors of building collapse in Nigeria, so as to make appropriate recommendation to the various units/departments/constructors so as to limit or reduce the number of deaths recorded from building collapse.

II. METHODS

2.1 Poisson Regression Model

By definition, \( Y \) (dependent variable) follows a Poisson distribution with parameter \( \lambda > 0 \) iff the probability distribution function is given by

\[
P(Y = k) = \frac{e^{-\lambda} \lambda^k}{k!}
\]

For \( K = 0, 1, 2, \ldots \) such that \( E(Y) = \lambda \) and \( Var(Y) = \lambda \)

For \( n \) independent random variables \( Y_1, Y_2, \ldots, Y_n \)

\( Y_i \sim P(\mu_i) \), and suppose we want to let the mean and the variance depend on the explanatory variables \( x_i \)’s

\[
g(\mu_i) = \mu_i = x_i \beta = \eta_i
\]

We can consider a generalized linear model with log link as: \( \log(\mu_i) = x_i \beta \).

Where \( X_i \) denotes the vector of explanatory variables and \( \beta \) denotes the vector of regression parameters.

2.2 Model Specification

The model for the Poisson regression is given as:

\[
\lambda_i = E(Y_i) = \exp \left( \alpha + \sum_{i=1}^k x_i \beta_j \right)
\]

Where \( \alpha \) is the intercept, Since the mean is equal to the variance the usual assumption of Homoscedasticity would not be appropriate for a Poisson data.

2.3 Parameter Estimation-Iterative Reweighted Least Squares

Estimation of parameters in Poisson regression relies on maximum likelihood estimation (MLE) method.

The log likelihood function is given as:

\[
\log L(\beta) = \sum y_i \log(\mu_i) - \mu_i
\]

To obtain the maximum likelihood for the parameter \( \beta_j \) we employ the chain rule

\[
\frac{\partial}{\partial \beta_j} = U_j = \sum_{i=1}^n \frac{\partial}{\partial \eta_i} = \sum_{i=1}^n \frac{\partial}{\partial \mu_i} \cdot \frac{\partial}{\partial \beta_j} \cdot \frac{\partial}{\partial \eta_i}
\]

Hence the iterative equation would be written as:

\[
b^{(m)} = (X^T W X)^{-1} X^T W
\]

Where \( w_{ii} = \frac{1}{\text{var}(Y_i)} \) and \( z_i = y_i \)

2.4 Negative Binominal Regression (NBR) Model
NBR is a popular generalization of Poisson regression because it loosens the highly restrictive assumption that the variance is equal to the mean made by the Poisson model. The traditional negative binomial regression model, commonly known as NB2, is based on the Poisson-gamma mixture distribution. This model is popular because it models the Poisson heterogeneity with a gamma distribution. This is given as:

\[ \lambda = E(y_i) = \exp(\alpha + \sum \beta X_i + \sum y_D) \]

(8)

\[ P(Y_i = y_i) = \frac{\exp(-\mu_i) \mu_i^{y_i}}{y_i!}, \quad y_i = 0,1,2,\ldots \]

(9)

\[ Y_i \sim NB(\lambda = \exp(x^T \beta), \psi) \]

(10)

Compounding Poisson distribution above and a gamma distribution would give a Negative Binomial distribution

\[ \text{Pr}(Y_i = y_i|\mu_i, \alpha) = \frac{\Gamma(Y_i+\alpha^{-1})}{\Gamma(Y_i+1)\Gamma(\alpha^{-1})} \left(\frac{1}{1+\alpha \mu_i}\right)^{\alpha^{-1}} \left(\frac{\alpha \mu_i}{1+\alpha \mu_i}\right)^{y_i} \]

(11)

Where \( E(Y_i) = \mu_i \) while the variance \( \text{Var}(Y_i) = \mu_i(1 + \alpha^{-1} \mu_i) \) where \( \alpha = \psi^{-1} \)

(12)

• The regression coefficients are estimated using the method of maximum likelihood.

• The significant regression parameters from (11) will be subjected to a test of hypothesis.

• The AIC would reveal the better model between equation (4) and equation (11) then the rank test would be employed to ascertain the levels of contribution of the various factors under consideration to Oil spill in the Niger Delta.

2.5 Model Evaluation

In this research work, we would be applying the techniques below for evaluating the models to also ascertain the better model suitable for modelling Oil spill in the Niger Delta from 1976 – 1996:

Deviance: This is a goodness of fit statistic for a model that is often used for statistical hypothesis testing and often used to compare two different models. Larger deviance indicates low performance of the model and less deviance indicates better performance of the model.

\[ D = 2(l(y, \phi; y) - l(\hat{y}, \hat{\phi}; y)) \]

(13)

Where \( l(y, \phi; y) \) is the log-likelihood of the full model and \( l(\hat{y}, \hat{\phi}; y) \) is the log-likelihood of the current model.

AIC: The Akaike would be used in selecting the best model that fit our data. The model with the smaller AIC is the best.

\[ \text{AIC} = -2(\text{likelihood} - K) \]

\[ \text{BIC} = -2\ln(\text{likelihood}) - K\ln(n) \]

(15)

(16)

Where \( n \) is sample size and \( K \) is number of predictors;

Pearson Residual: It is used to check for model fit of each variable (explanatory variables). It is the discrepancy between our observed and fitted values for each observation.

III. RESULTS AND DISCUSSION

Table 1: Descriptive statistics of the response variable and the explanatory variables

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>No of Casualty</td>
<td>22</td>
<td>19.18</td>
<td>26.010</td>
<td>676.537</td>
</tr>
<tr>
<td>Substandard Materials</td>
<td>22</td>
<td>13.50</td>
<td>18.278</td>
<td>334.071</td>
</tr>
<tr>
<td>Poor Concrete Mixture</td>
<td>22</td>
<td>8.73</td>
<td>11.752</td>
<td>138.113</td>
</tr>
<tr>
<td>Wrong Construction Method</td>
<td>22</td>
<td>7.73</td>
<td>10.430</td>
<td>108.779</td>
</tr>
<tr>
<td>Change of Use</td>
<td>22</td>
<td>4.95</td>
<td>6.492</td>
<td>42.141</td>
</tr>
<tr>
<td>Use of Quacks</td>
<td>22</td>
<td>11.45</td>
<td>15.586</td>
<td>242.926</td>
</tr>
<tr>
<td>Hasty Construction</td>
<td>22</td>
<td>9.82</td>
<td>13.114</td>
<td>171.965</td>
</tr>
<tr>
<td>Corruption/greed</td>
<td>22</td>
<td>6.82</td>
<td>9.069</td>
<td>82.251</td>
</tr>
</tbody>
</table>

Inmodelling the mortality rate data of collapse building in Nigeria, R statistical software version 3.3.3 was used. The Generalized Linear Model (GLM) with Poisson as the fundamental distribution for modelling a count data using the Log link function and the Negative Binomial distribution was latter employed to correct the error of over dispersion in the count data in situation where the result of the Poisson regression model shows over dispersion.
3.2 Modelling Incident of Building Collapse Data in Nigeria

Table 2: The Poisson Regression Models for Building Collapse in Nigeria 1995 – 2016 with their AIC’s

<table>
<thead>
<tr>
<th>Models</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>log(mean of No Killed STM) = ( \alpha ) + ( \beta )STM</td>
</tr>
<tr>
<td>2.</td>
<td>log(mean of No Killed STM) = ( \alpha ) + ( \beta )STM + ( \beta )PCM</td>
</tr>
<tr>
<td>3.</td>
<td>log(mean of No Killed STM) = ( \alpha ) + ( \beta )STM + ( \beta )PCM + ( \beta )WCM</td>
</tr>
<tr>
<td>4.</td>
<td>log(mean of No Killed STM) = ( \alpha ) + ( \beta )STM + ( \beta )PCM + ( \beta )WCM + ( \beta )COU</td>
</tr>
<tr>
<td>5.</td>
<td>log(mean of No Killed STM) = ( \alpha ) + ( \beta )STM + ( \beta )PCM + ( \beta )WCM + ( \beta )COU + ( \beta )UOQ</td>
</tr>
<tr>
<td>6.</td>
<td>log(mean of No Killed STM) = ( \alpha ) + ( \beta )STM + ( \beta )PCM + ( \beta )WCM + ( \beta )COU + ( \beta )UOQ + ( \beta )HC</td>
</tr>
<tr>
<td>7.</td>
<td>log(mean of No Killed STM) = ( \alpha ) + ( \beta )STM + ( \beta )PCM + ( \beta )WCM + ( \beta )COU + ( \beta )UOQ + ( \beta )HC + ( \beta )CG</td>
</tr>
</tbody>
</table>

STM=Substandard Materials, PCM=Poor Concrete Mixture, WCM=Wrong Construction Method, COU=Change of Use, UOQ=Use of Quacks, HC=Hasty Construction, CG=Corruption/greed.

From table 2 above, we see that the best model which fit the Building Collapse data in Nigeria 1995 – 2016 is model 7 because it has the smallest Akaike Information Criterion (AIC). The AIC of model 7 was found to be 153.534 with a null deviance of 504.667 on 21 degree of freedom and a residual deviance of 45.13 on 14 degree of freedom following the chi-square distribution (\( \chi^2 \)).

The parameters and their estimates, the standard errors, the Chi-squares with their associated probability values are presented in table 3 below.

Table 3: The Parameter Estimates of the Selected Poisson Regression Model

| Parameter               | Estimate | Std. Error | Z values | Pr(>|Z|) |
|-------------------------|----------|------------|----------|----------|
| (Intercept)             | 2.238    | 0.1199     | 18.669   | <2e-16   |
| Substandard Materials   | -1.161   | 0.2817     | -4.120   | 3.78e-05 |
| Poor Concrete Mixture   | -.440    | 0.2996     | -1.467   | 0.142389 |
| Wrong Construction Method | 1.264   | 0.4331     | 2.919    | 0.003512 |
| Change of Use           | -1.546   | 0.3208     | -4.820   | 1.44e-06 |
| Use of Quacks           | 1.618    | 0.4837     | 3.345    | 0.000824 |
| Hasty Construction       | -.924    | 0.4213     | -2.194   | 0.028214 |
| Corruption/greed        | 1.203    | 0.2836     | 4.240    | 2.23e-05 |
The table 3 presents the parameter estimates of the selected model for the incident of building collapse in Nigeria over 21 years ie from 1995 - 2016. The Akaike information criterion (AIC) of this model was 153,534 with a null deviances of 504.667 on 21 degree of freedom and a residual deviance of 145.134 on 14 degrees of freedom following the chi-square distribution ($\chi^2$) on one degree of freedom. The dispersion parameter was found to be 3.224 (i.e residual deviance/degree of freedom as seen in table 5) and the Omnibus test flag a P-value equal 0.000 which implies that the model is significant at 5% $\alpha$ -level. However, the assumption of equality of mean and variances in Poisson distribution has been violated since the dispersion parameter is not approximately equal to 1. The dispersion parameter of the above model is 3.224 which is greater than 1, a clear indication of over dispersion in the building collapse data. This further implies that the parameters of the stated model have been over-estimated and the corresponding standard errors have been underestimated consequently giving a misleading inference about the regression parameters. To address this issue, Negative Binomial regression was used to modify the model to nullify the effect of over dispersion in the data and the result is shown in table 3 below.

| Parameter                  | Estimate | Std. Error | Z-values | Pr(>|Z|) |
|----------------------------|----------|------------|----------|---------|
| (Intercept)                | 2.1353   | 0.1949     | 10.957   | <2e-16  |
| Substandard Materials      | -0.9910  | 0.4291     | -2.309   | 0.0209  |
| Poor Concrete Mixture      | -0.4279  | 0.4378     | -0.977   | 0.3284  |
| Wrong Construction Method  | 1.1982   | 0.6637     | 1.805    | 0.019   |
| Change of Use              | -1.2643  | 0.5157     | -2.452   | 0.0142  |
| Use of Quacks              | 1.4096   | 0.7410     | 1.902    | 0.010   |
| Hasty Construction         | -0.8899  | 0.6351     | -1.401   | 0.0473  |
| Corruption/greed           | 1.0346   | 0.4397     | 2.353    | 0.0022  |

Interpretation of Coefficients

From table 4 it can be seen that Substandard Materials, Wrong Construction Method, Change of Use, Use of Quacks, Hasty Construction and Corruption/greed were all statistically significant since (p-value < 0.05) while Poor Concrete Mixture was not statistically significant since (p-value < 0.05).

The variables Substandard Materials (STM), Wrong Construction Method (WCM), Change of Use (COU), Use of Quacks (UOQ), Hasty Construction (HC) and Corruption/greed (C_G) where all statistically significant at 5% $\alpha$ level. This implies that Building Collapse in Nigeria is not as a result of witchcraft, but from this study we see all the above listed factors contributing significantly to building collapse which in turn leads to high number of death recorded. From this study, it can be established that Substandard Materials, Wrong Construction Method, Change of Use, Use of Quacks, Hasty Construction and Corruption/greed are the major contributors to building collapse in Nigeria.

From Table 4 it is observed that the parameter estimates have been reduces and the standard errors have also been increased. The parametric analysis for the comparison between the Poisson and Negative Binomial Regression for goodness of fit of the model is shown in table 5.

<table>
<thead>
<tr>
<th>Assessment Parameter</th>
<th>Poisson Regression Model</th>
<th>Negative Binomial Regression Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null Deviances</td>
<td>504.667</td>
<td>174.196</td>
</tr>
<tr>
<td>Degree of Freedom</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>Residual Deviance</td>
<td>45.134</td>
<td>23.977</td>
</tr>
<tr>
<td>Degree of Freedom</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-137.534</td>
<td>-133.63</td>
</tr>
<tr>
<td>Dispersion Parameter</td>
<td>3.224</td>
<td>1.715</td>
</tr>
<tr>
<td>Akaike’s Information Criterion (AIC)</td>
<td>153.534</td>
<td>151.63</td>
</tr>
</tbody>
</table>

From the result presented in table 5, it is clear that the negative binomial regression model is the better model which fits the Building Collapse Data in Nigeria since the dispersion parameter has reduced from 3.224 which was given by the Poisson model to 1.715 in the Negative Binomial model. The Akaike Information Criterion (AIC) of the Poisson Regression model also reduced from 153.534 to 151.63 in the negative binomial model, we can see more supporting claims from the Null Deviances; Residual Deviance;
Modelling Incidents of Building Collapse Data in Nigeria, Using Poisson and Negative Binomial

Log Likelihood of all which also corroborated our findings that the Negative Binomial Model is the better model to modelling the number of death as a result of building collapse.

3.3 The Model for Predicting Casualty from Building Collapse in Nigeria

For negative binomial regression, the model for the Casualty from Building Collapse data is obtained as:

\[ \log(\text{mean of No of Killed}) = 2.1353 - 0.9910(STM) - 0.4279(PCM) + 1.1982(WCM) - 1.2643(COU) + 1.4096(UOQ) - 0.8899(HC) + 1.0346(C \_ G) + e \]

<table>
<thead>
<tr>
<th>s/n</th>
<th>Parameter</th>
<th>P-values</th>
<th>Ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Substandard Materials</td>
<td>0.0209</td>
<td>5th</td>
</tr>
<tr>
<td>2</td>
<td>Wrong Construction Method</td>
<td>0.0190</td>
<td>4th</td>
</tr>
<tr>
<td>3</td>
<td>Change of Use</td>
<td>0.0142</td>
<td>3rd</td>
</tr>
<tr>
<td>4</td>
<td>Use of Quacks</td>
<td>0.0100</td>
<td>2nd</td>
</tr>
<tr>
<td>5</td>
<td>Hasty Construction</td>
<td>0.0473</td>
<td>6th</td>
</tr>
<tr>
<td>6</td>
<td>Corruption/greed</td>
<td>0.0022</td>
<td>1st</td>
</tr>
</tbody>
</table>

From table 6 which ranks the P-values of the better model (Negative binomial Regression) to identify which parameter was more significant also which factor contributed more than the other. Corruption/greed was rank in the first position. Implication is that Corruption/greed factor is the most statistically significant factor that contributing to high number of death as a result of building collapse in Nigeria followed by Use of quacks, change of use, wrong construction method, substandard materials and hasty construction while poor concrete was not a significant contributor in the subject matter in this study.

IV. CONCLUSION

Negative binomial regression model is the better model suitable for modeling incident of building collapse in Nigeria as considered in this study. Substandard Materials (STM), Wrong Construction Method (WCM), Change of Use (COU), Use of Quacks (UOQ), Hasty Construction (HC) and Corruption/greed (C\_G) are major factors that contribute to the high number of death as a result of building collapse in Nigeria. The model which can be used to forecast future number of casualty from building collapse keeping the factors considered in this study constant is shown below:

\[ \log(\text{mean of No of Killed}) = 2.1353 - 0.9910(STM) - 0.4279(PCM) + 1.1982(WCM) - 1.2643(COU) + 1.4096(UOQ) - 0.8899(HC) + 1.0346(C \_ G) + e \]

REFERENCES