Factors Influencing Research Productivity at Njala University: A Count Regression Approach

Regina Baby Sesay*[1], Sheku Seppeh[2], Mohamed Kpangay[3]

[1,2,3]*Department of Mathematics and Statistics, School of Technology, Njala University, Njala, Sierra Leone*

***Abstract*— Research promotes professional excellence as it helps academics to be more innovative thereby enhancing outstanding student education. Like most universities, Njala University (NU) academic staff are required to teach, research, and carry out community outreach. Publishing in a peer-reviewed journal is evidence of the effort made by academic staff to fulfill the obligation of one of the job descriptions of Njala University. The University like almost all other academic institutions promotes it, academic staff, purely based on the strength of their research work. The research strength of each academic staff is measured by the number of original research papers published in peer-reviewed journals. As such, the publication of original academic papers in recognized peer-reviewed journals has become the dream of each NU academic staff. However, despite the huge desire for publication, some unavoidable factors are infringing on the research activities of most academic staff. This research paper, therefore, used a statistical modeling technique for count data, to identify the main factors influencing the research productivity of NU academic staff. A stratified random sampling method was employed to select 113 respondents proportionately from each school. Data were collected from the selected respondents using structured questionnaires. The Poisson regression model was used as the baseline model. Due to the evidence of over-dispersion and excess zeros in the response variable, three additional count regression models were used in the analysis. Based on statistical tests, the zero-inflated hurdle model significantly outperformed the Poisson and Negative binomial regression models. However, the difference in performance between the zero-inflated poison and the zero-inflated hurdle model was not statistically significant. Initially, several factors were considered as possible determinants of research productivity of NU academics. However, the empirical analysis showed that academic qualification; teaching experience and hours spent on research are the main (significant) factors influencing the research productivity of Njala University academic staff. Increase in the number of hours spent on research can increase the number of research publications. Academics with more teaching experience tend to publish more than those with little or no teaching experience. The higher the academic degree attained by the academic, the higher the possibility of publishing more research papers.**

***Keywords*—Count Regression, Njala University, Research Strength, Academics, Non nested models**

I. INTRODUCTION

Research is a step by step search for knowledge. It involves fact discovery, and revision of accepted theories in the light of newly discovered facts, and the applicability of such discoveries to contribute to the development of the environment in which we live. According to [5], research is an objective and systematic analysis and recording of controlled observations that may lead to generalizations, principles or theories, resulting in predictions and possible control of events. This clearly points out the importance of research in the universities as it leads to innovations and invention of new knowledge. Many concerns have been raised with regards to factors influencing the research performance of universities, Colleges and academics. Amongst these are the work of [31] and [32].

The university can be considered as a"knowledge production system ([14], [15]). One of the major outputs of educational organizations is to produce knowledge through publication and citation in high-indexed journals [38]. The primary aim of almost every university is to reach a world-class university level. Rate of publication and increase enrolment rate are the major pathways to achieve this aim.

In Sierra Leone today, like most other countries all over the word, research and publications are amongst the main determinants for university success in the battles for supremacy and rank for excellence amongst other competing universities. Reference [37], even stated that productivity in the academe is consensually regarded as an indicator of research activity conducted by individuals, institutions, countries, and regions as a whole.

Njala University once ranked as the number one university among other universities in Sierra Leone has continuously emphasized the importance of publication for its faculty members. The University promotes its academic staff purely based on the number of publications made in reputable peer-reviewed journals. To increase research productivity, Njala University has tied publication output to promotion and recognition. This is evidence of a clear manifestation of the phrase 'publish or perish' in the promotion criteria for Njala University academic staff. Presently, at Njala University, the common saying is 'no publication no promotion This means, It does not matter the number of years of service to the University, number of years in an administrative position, number of University committees attended or even the teaching workload for each semester, as long as an academic does not publish up to the required number of publications in a peer-reviewed journal, that academic staff will not be promoted. For example, academic staff with five or more publications; very few courses to teach per semester; no extra service to the university (not a member of any university committee) and no administrative experience is prefered to be promoted over a dedicated academic staff with only two publications, many courses to teach per semester, member of more than 8 university committees, and have spent many years in university administrative position. This is sometimes not easily welcome by most hard-working academics of the University.

Above all, despite the outstanding emphasis placed on publication, most Njala University, academic staff are faced with several constraints in their desire to publish high-quality papers in reputable journals. Some of these constraints include teaching workload, University services (a member of many university committees), administrative responsibility, type of discipline, gender, experience (number of years spent teaching), internet facility, and residence (on or out of university campus),

This research work, therefore, used a regression modeling technique for count data to identify the main factors influencing the number of publications made in peer-reviewed journals by each NU academic staff and to measure the effect of each factor on the number of research publications.

Ц. MATERIALS: THEORETICAL FRAMEWORKS

This section focused on the review of some regression models used in this research work. In particular, it tried to review some widely used count data regression models like: the poison regression model, negative binomial regression model and the zero-inflated regression models. In addition, this section tried to review some statistical test used to access the goodness of fit for each of these models.

- *Regression Model for Count Data*

Count based data contains events that occur at a certain rate. The rate of occurrence may change over time or from one observation to next. Regression Models for Count Data allow for regression-type analyses when the variable of interest called the dependent variable is a numerical count. They are regression models in which the dependent variable takes only the nonnegative integers. The dependent variable for count data regression model share certain properties in common: they are mostly positively skewed; they are not negative and their lowest possible value is zero

For count regression model, the conditional mean E(yi\xi) of the dependent variable, yi, is assumed to be a function of a vector of covariates, x_i.

A. The Poisson regression

The Poisson regression is an example of a broad class of regression models known as generalized linear models (GLM). It is the basic regression model from which a variety of count regression models emanate. The Poisson regression is derived from the Poisson probability mass function, given as:

$$f(y_i;\theta_i) = \frac{e^{-t_i\theta_i}(t_i\theta_i)^{y_i}}{y_i!}, \qquad y_i = 0,1,2,\dots;\ \theta_i > 0 \quad , \qquad ,$$

where y_i is the response,which is the number of publications made by an academic staff at NU; θ_i, the predicted count, which is the predicted number of publications made by the academic; (θ_i is also known as the rate parameter); t_i, the area or time in which count enter the model.

Although the normality assumption is not applicable to count data as they are not continuous, there are other conditions that must be fulfilled for the result of the Poisson regression model to be valid. These conditions include:

- The dependent variable must be a count per unit of time or space, described by a Poisson distribution.
- The observations must be independent of one another.
- The mean of the Poisson random variable must be equal to its variance.
- The logarithm of the expected value must be modeled by a linear combination of the unknown parameters

The Poisson Regression model can thus be written as:

$$Log(count) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \cdots,$$

In the case of rate data , the Poisson regression can be witten as:

$$Log(\mu/T) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \cdots,$$

Hence

$$Log(\mu) - Log(T) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \cdots,$$

Which gives

$$Log(\mu) = Log(T) + \beta_0 + \beta_1x_1 + \beta_2x_2 + \ \dots,$$

Where: μ = Expected number of event

T= Index of exposure e.g. people at risk, hours, days, sq. miles, etc.

Log(T) = offset variable and is the log of the measure of exposure.

Although the Poisson regression is the choice model when it comes to count data analysis, very often, the poisson regression displays over-dispersion which may lead to an underestimation of the standard errors and inflation of the significance of the regression parameters. Because of these back draws, other regression models have been suggested to accommodate overdispersion (and/or under-dispersion) in count data

Therefore, in this research work, three additional regression models for count data were considered. These were, the negative binomial regression, zero-inflated Poisson regression and the zero-inflated hurdle regression models. Each of these models is design for a different type of count variable distribution.

B. Negative Binomial Regression

The Poisson regression model stands out as the first regression model to consider when the outcome variable is a count of the number of occurrence of an event. However, due to the restrictive assumption (variance equal to the mean) attached to the Poisson regression model and because of the different ways count variables can be distributed, the Negative binomial regression which is a generalization of the poisson regression is sometimes considered as an alternative regression model when the poison assumption is violated. In particular, It is an alternative regression model when there is overdispersion , i.e. when $var(Y) > E(Y)$

Like the Poisson regression model, the dependent variable for the negative binomial regression is a count variable. A convenient parameterization of the negative binomial distribution is given by [21].

The fundamental negative binomial regression model for an observation i is written as:

$$\Pr(Y = y_i \backslash \mu_i, \alpha) = \frac{\Gamma(y_i+\alpha^{-1})}{\Gamma(\alpha^{-1})\Gamma(y_i+1)}\left(\frac{1}{1+\alpha\mu_i}\right)^{\alpha^{-1}}\left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i},$$

where

$\mu_i = t_i\mu$ and $\alpha = \frac{1}{v}$

$\mu > 0$ is the mean incidence rate of y per unit of exposure.

Exposure may be time, space, distance, area, volume, or population size

ti is the exposure for a particular observation.

α is the heterogeneity parameter

In negative binomial regression, the mean of y is determined by the exposure time t and a set of k regressor variables (the x's). These quantities are related by the expression

$$\mu_i = \exp\left(In(t_i) + \beta_1 x_1 + \beta_2 x_2 +,,, + \beta_k x_{ki}\right),$$

When $x_1 = 1$, β1 is called the intercept (most often $x_1 = 1$) . The regression coefficients $\beta 1, \beta 2, \dots, \beta 1$ with estimate $\hat{\beta}_1$,,$\hat{\beta}_2$, …, $\hat{\beta}_k$, are unknown parameters to be estimated from the research data.

- *.Estimation of the negative binomial model*

The regression coefficients of the negative binomial regression model are estimated using the method of maximum likelihood. Based on the work of [10] the logarithm of the likelihood function is given as:

$$L = \sum_{i=1}^{n}[In[\Gamma(y_i + \alpha^{-1})] - In[\Gamma(\alpha^{-1})] - In[\Gamma(y_i + 1)],$$

$$-\alpha^{-1}In(1 + \alpha\mu_i) - y_i In(1 + \alpha\mu_i) + y_i In(\alpha) + y_i In(\mu_i),$$

C. *zero-inflated poison (zip) regression models*

Zero-inflated regression models are used to model count dependent variables that have a considerable number of zero values. The zero-inflated Poisson (ZIP) regression has been used by many researchers for handling zero-inflated count data. ZIP regression can be obtained by mixing a distribution generated at zero with the poisson distribution, This is done by allowing the incorporation of explanatory variables in both the zero process and the poisson distribution. The p.m.f. of ZIP regression is,

$$\Pr(Y_i = y_i) = \begin{cases} \omega_i + (1-\omega_i)\exp(-\mu_i), & y_i = 0 \\ (1-\omega_i)\frac{\mu_i^y}{y_{i!}}\exp(-\mu_i)\, y_i > 0 \end{cases}$$

Where $0 \le \omega_i < 1$ and $\mu_i > 0$, with mean $E(Y_i) = (1-\omega_i)\mu_i$ and variance

$Var(Y_i) = (1-\omega_i)\mu_i(1+\omega_i\mu_i)$. ZIP regression reduces to Poisson regression when $\omega_i = 0$, exhibits overdispersion when $\omega_i > 0$ and exhibits under-dispersion when$\omega_i < 0$

The covariates can be incorporated by using a log link for μ_i and a logit link for ω_i,. i.e.

$$\log(\mu_i) = X_i^T\beta \text{ and } log\left(\frac{\omega_i}{1-\omega_i}\right) = Z_i^T Y$$

Where x_i and z_i are the vectors of explanatory variables, and β and γ are the vectors of regression parameters. Maximum likelihood estimates can be obtained by maximizing the log likelihood

D. zero-inflated hurdle regression models

The hurdle models are a class of models for count data that help to handle excess zeros and over-dispersed data. To correct for the excess zeros that practically exist in count data, the hurdle models divide the modeling procedure into two distinct processes. The first part of the process determines whether the response outcome is positive via a binary model for the dichotomous event of having zero or positive values. Logistic regression is usually used to allow for the investigation of the effects of covariates on the probability of an observation being zero. If the realization is positive, the hurdle is crossed and the second process models the level of the outcome which is a truncated-at-zero count outcome. Frequent choices for the truncated-at-zero count model are truncated Poisson model, or the truncated negative binomial model.

Although the hurdle model is similar to the zero-inflated model, the hurdle model is more flexible in that, the zero outcomes can be deflated as well as inflated. The probability mass function for the hurdle likelihood is defined by

$$p(y \backslash \theta, \lambda) = \begin{cases} \theta & if\ y = 0\ \ and \\ (1-\theta)\frac{Poisson(y\backslash\lambda)}{1-PoissonCDF(0\backslash\lambda)} & if\ \ y > 0 \end{cases},$$

Where, Poisson CDF is the cumulative distribution function for the Poisson distribution.

F. Model Check: Goodness-of-fit tests

The two frequently used traditional t**ools for** model diagnostics **(**goodness-of-fit) in generalized linear models (GLM), are the Pearson chi-squares and the deviance.

For a model to be correctly specified in count regression analysis, the Pearson chi-square statistic and the deviance divided by their respective degrees of freedoms should be approximately equal to one. Values much larger than one, implies that, the equi-dispersion assumption of the poison regression is violated and that the data are said to exhibit over-dispersion. Whereas, values less than one implies under-dispersion.

1) Deviance

The deviance is a measure of how well the model fits the data. If the model fits well, the observed values Y_i will be close to their predicted mean, μ_i, causing the deviance to be small.The deviance for a fitted Poisson regression is equal to:

$$D = 2\sum_{1=1}^{n}\left\{Y_i log\left({Y_i}/{\mu_i}\right) - (Y_i - \mu_i)\right\},$$

Where

$$\mu_i = exp\left(\hat{\beta}_1 X_, + \hat{\beta}_2 X_2 + \dots\ + X_p\hat{\beta}_p\right),$$

denotes the predicted mean for observation i based on the estimated model parameters.

An alternative measure of goodness of fit is Pearson's chi-squared statistic,

2) Pearson chi-squares

The Pearson chi-square statistic is given by:

$$\chi_p^2 = \sum_{i=1}^{m}\sum_{j=1}^{k+1}\frac{\left(r_{ij}-n_i\hat{p}_{ij}\right)^2}{n_i\hat{p}_{ij}},$$

Where m is the number of subpopulation profiles,

$\boldsymbol{k}$+1 is the number of response levels,

$\boldsymbol{r_{ij}}$ is the total weight associated with ***j***th level responses in the ***i***th profile

$$n_i = \sum_{j=1}^{k+1} r_{ij},$$

p_{ij} is the fitted probability for the ***j***th level at the ***i***th profile.

For a well specified model, the chi-square statistic has $mk - q$ degrees of freedom, where q is the number of parameters estimated.

E. Model Selection Criteria (AIC and BIC)

The Akaike Information Criteria (AIC) [17] and the Bayesian Schwartz Information Criteria (BIC) were the statistical tools used to choose the best fitting model among the competing count regression models used in this work,

The AIC and BIC are both penalized-likelihood criteria. The only difference between the two in practice is the size of the penalty (a penalty for using the sample data to estimate the model parameters).The BIC penalizes model complexity more heavily.

1) Akaike Information Criterion (AIC)

Akaike information criterion (AIC) [17] is a technique based on in-sample fit to estimate the likelihood of a model to predict the future values.

The model with minimum AIC was considered as the best model to fit the data [8].

The following equation was used to estimate the AIC [17] of a given model:

$$AIC = -2 * \ln(L) + 2 * k$$

where L is the value of the likelihood and k is the number of estimated parameters.

2) Bayesian information criterion (BIC)

In selecting the best fitting model, the Bayesian information criterion (BIC) measures the trade-off between model fit and complexity of the model.

A lower BIC value indicates a better fit. Reference [22] also mentioned the Bayesian Information criterion (BIC) as another common fit statistic.

The following equation was used to estimate the BIC :

$$BIC = -2 * \ln(L) + 2 * \ln(N) * k,$$

where L is the value of the likelihood, N is the number of recorded measurements, and k is the number of estimated parameters.

3) Voung test for Comparing two non-nested Models: In addition to the AIC and BIC measured above, the likelihood-ratio based test proposed by [36] that can compare non-nested models was also used in the model comparison.

III. METHODOLOGY

A. Study Area

This study was carried out at Njala University in the Moyamba District, the southern part of Sierra Leone. In terms of academic output, Njala University is one of the leading universities in Sierra Leone. It is situated near the bank of the Tia river in the Korie Chiefdom. The University has two campuses and this study was carried out on the main Njala university campus that hosts most of the academic departments and staff.

B. Population and Sample Selection

The target population consisted of all the academic staff of Njala University. A Stratified random sampling was used to divide the university teaching and research staff into categories as per schools. Random sampling was then used to select the respondents proportionately from each school. A total sample of one hundred and thirteen (113) academic staff was selected. Data was collected from the selected sample using structured questionnaires.

C. Description and Measurement of Variables used in the Model

Many factors were initially considered as potential determinants of the number of research publications made by NU academics in peer-reviewed journals after the Sierra Leone rebel war. The initial factors considered were: age, gender, marital status, residence, teaching work load, University service, number of years in administrative position, internet access and type of discipline.

1) Measurement of Dependent Variable: The dependent variable in the count regression model was the number of publications made by each academic staff after the Sierra Leone rebel war. The measurement of the dependent variable in this work is in line with the work of [24] which stated that, research effectiveness can be measured by simply counting the number of publications in reputable journals.

2) Measurement of Independent Variables:

- *Age:* The age of the academic staff was measured on a continuous scale. Many studies have found, the relationship between age and number of research publications to be curvilinear as the number of publications increases with age and reaches a peak at some point during the career and then declines. ([3], [16], [26]).
- *Gender:* This referred to the gender of the academic staff. In this study, gender was considered to be a dichotomous categorical variable, that is either Male or Female. Gender difference in terms of the number of publications has surfaced in many published work relating to research productivity. Although some studies have shown that female researchers tend to publish fewer publications than their male counterparts (example: [2], [29] and [39], Other studies have reported that there is no statistically significant gender difference (e.g. [7] and [28]).This shows that the difference in the level of scientific productivity between men and women needs further investigation. This research, therefore, considered the gender of the academic staff as a possible determinant of the academic's research productivity.
- *Marital Status:* This referred to the marital status of the academic staff. In this study, marital status was considered to be a dichotomous categorical variable, that is, either married or single. Marital commitments like raising children may be a possible factor that may negatively influence the number of research publications. Reference [23] analyzed the effect of children on the entire careers of academics which is different for men and women.
- *Experience:* This referred to the number of years in teaching service. Acquisition of skills and experiences enhances productivity [18].
- *Residence:* The issue of whether the academic staff resides on the university campus or out of the university campus was also considered as a possible factor that may influence the number of research publications
- *Teaching work load:* The number of courses taught by each academic staff per semester was identified as a possible determinant of the number of publications made by the academic staff. This was supported by the work of [19] who identified institutional factors as a means of discouragement of research productivity. In their work, they commented that teaching workload and administrative duties were time-consuming, fatiguing and divert an academic away from his/her scholarly writing pursuits. Pressure has continued to mount for greater orientation towards teaching as an alternative or complement to publication among academics in universities [9].

  In the measurement of teaching work load, more than five courses to teach per semester was considered as a high teaching workload; less than five courses to teach per semester as less teaching workload and five courses to teach per semester as a normal teaching workload.
- *Number of Hours spent on Research:* The number of hours spent on research each week was identified as an important factor that can influence the research strength of the academics. The conflict between teaching and research is evident in most academic departments at Njala University. University programs are increasing yet the staff establishments remain the same for most academic departments. This has led to an increase in the number of courses taught by each academic staff in most of the academic departments. This clearly points out research productivity as the opportunity cost associated with teaching.
- *University service:* University service was determined by the number of University committees attended by the academic staff.
- *Number of years in Administrative position:* This refers to the number of years in administrative position as head of department or dean of a school.
- *Type of discipline:* This refers to the major subject discipline of the academic staff. A large-scale study confirms that the differences in publication rate and impact are discipline-specific [12]. This study, therefore tried to find out whether the academic's major subject discipline has any influence on the number of publications made by the researcher
- *Qualification:* This referred to the highest academic degree attained by the academic.

IV. EMPIRICAL ANALYSIS

A. Descriptive and Exploratory Data Analysis.

TABLE 1 SUMMARY OF VARIABLES USED IN THE COUNT REGRESSION MODELS (with n =113)

Variable Name	Variable Description	Dependent Variable (DV)/Independent Variable (IV)	Valid Range	Variable Type
Num_Pub	Number of Publications	DV	0-12 and above	Discrete (count)
Yrs_Admin	years in administrative position	IV	0-10 years and above	Continuous
Residence	Residence	IV	On University campus Out of UniversityCampus	Categorical
Gender	Gender	IV	Male, Female	Categorical
Internet	Internet access	IV	University, Private	Categorical
Uni_Service	University Services	IV	0-10 and above	Discrete (count)
Teach_Load	Teaching Load	IV	$<$ 5=high, 5=normal, $> 5 = high$	Categorical
Mar_Status	Marital Status	IV	Married, Single	Categorical
Type_Dis.	Research Area	IV	Art Pure Science and Math Science Agriculture and Others	Categorical
H_Qualif	Highest Academic Qualification	IV	Bachelors' Degree Master's Degree Doctorate Degree	Categorical
Teach_Expe	Teaching Experience	IV	1-15 years	Continuous
Nuhr_Research		IV	0-5 hours	Continuous
Age		IV	26-64 years and above	Continuous

1) Exploratory Data Analysis:

➢ *Distribution of number of Research Publications*

Figure 1 displayed the frequency distribution of the number of research publications (Dependent variable). From figure 1, the difference in the conditional mean and variance displayed the presence of over-dispersion in the count data. Also, from fig 1 the zero number of publications category is much higher than the others. This signaled the necessity for the use of a count regression model that can accommodate both over-dispersion and the excess zeros present in the count data.

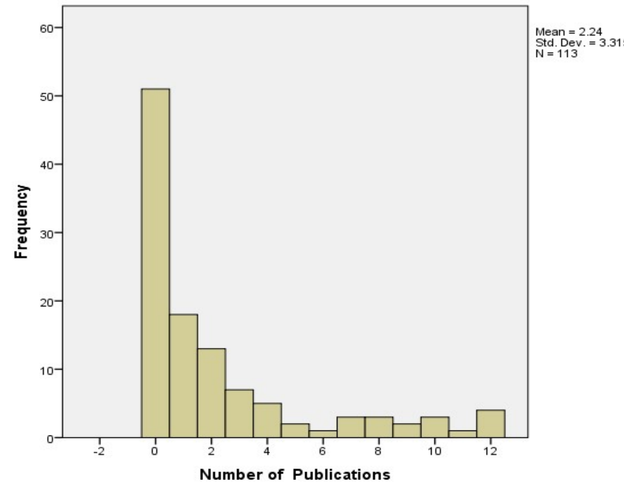

Fig1: Frequency distribution for number of Publications

➢ *Scatter Plots*

Apart from helping to identify outliers, a bivariate scatter plot is an important diagnostic tool that helps the researcher to explore the relationship between two variables, the dependent and one independent variable. In this research work, the scatter plot was used to form a working hypothesis regarding the relationship between each independent variable and the dependent variable. Count regression modeling techniques were finally used to test the working hypothesis to determine the direction and magnitude of each relationship.

Figure 2, contains bivariate exploratory displays for the number of publications (dependent variable) plotted against each of the independent variables. Out of the twelve (12) independent variables initially considered as potential determinants of the number of research publications, only few displayed significant linear relationships with the dependent variable.

Based on the exploratory analysis, the independent variables, academic qualification; hours spent on research; teaching experience and age showed a positive linear relationship with the dependent variable. This means that, an increase in any of these independent variable may increase the number of research publication. All the other independent variables initially considered as potential determinants of the number of research publications but did not show a significant relationship with the dependent variable were not included in the count regression analysis.

From figure 2, scatter plot1, a linear relationship was observed between the number of research publication and the hours spent on research. The R^2 from scatter plot1showed that 26% of the variability in the number of publication was accounted for by the number of hours spent on research

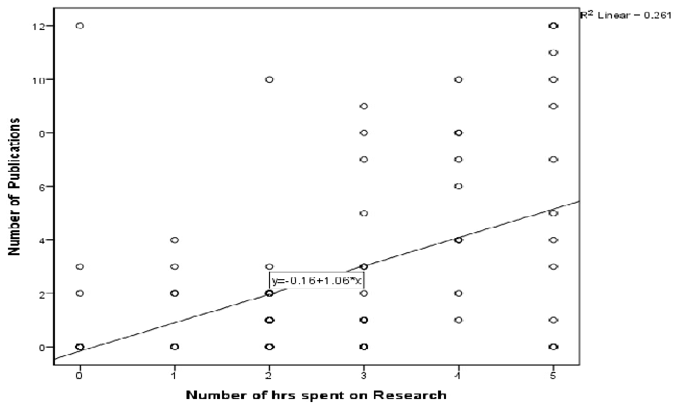

fig 2 **scatter plot 1**

Also from figure 3, scatter plot 2, the R^2 showed that 33% of the variability in the number of publication was accounted for by the number of years spent in teaching services (teaching experience).

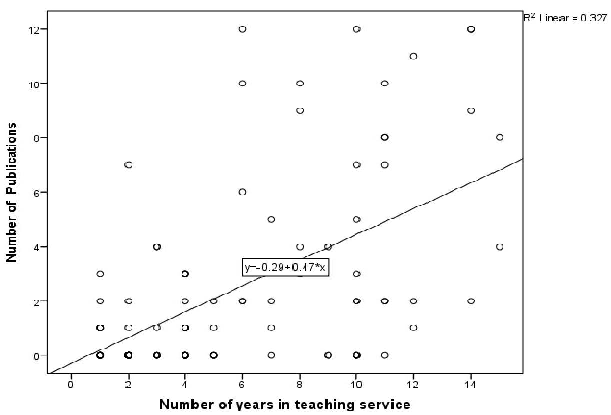

fig 3 **scatter plot 2**

Again, from figure 4, scatter plot 3, the R^2 showed that, the qualification of the academic staff accounted for about 16% of the variability in the number of research publications.

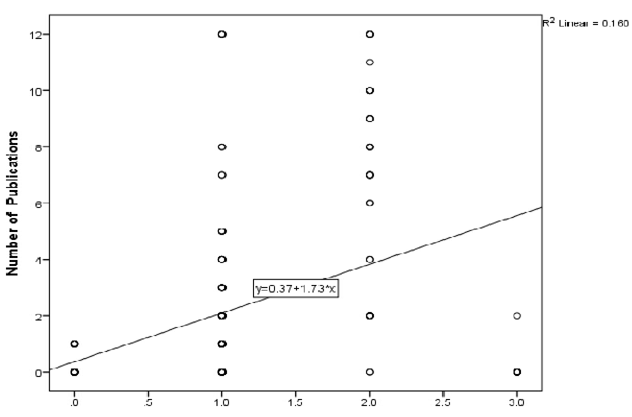

fig 4 **scatter plot 3**

Finally, a positive linear relationship was observed for the independent variable Age (in fig 5 scatter plot 4).The R^2showed that 21% of the variability in the number of research publications was accounted for by the Age of the academic staff.

SCATTER PLOT 4

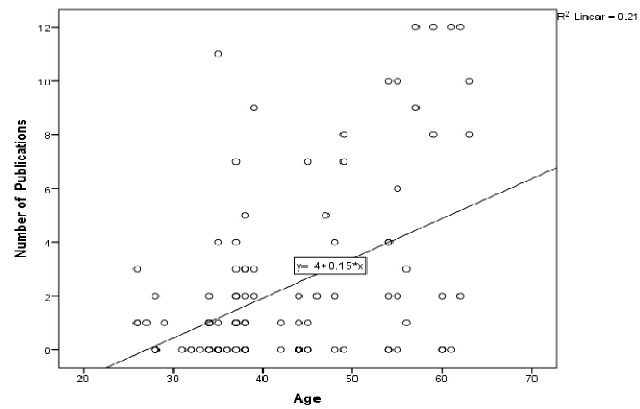

fig 5 **scatter plot 4**

2) Descriptive Analysis: The descriptive statistics in table 2 showed that the variance is greater than the mean. This is an evidence of over-dispersion in the count data.

TABLE Ц: DESCRIPTIVE STATISTICS OF DEPENDENT VARIABLE

Descriptive Statistics

	N	Minimum	Maximum	Mean	Variance
Number of Publications	113	0	12	2.24	10.987
Valid N (listwise)	113				

The categorical variable coding result presented in table 3 showed that majority of the academics selected for the sturdy were master's degree holders.

TABLE III: CATEGORICAL VARIABLE

Categorical Variable Information

			N	Percent
Factor	Highest Accademic Qualification	Doctorate Degree	19	16.8%
		Masters Degree	68	60.2%
		Bachelors Degree	26	23.0%
		Total	113	100.0%

B. Count Data Regression Analysis

Four count data regression models were used in the data analysis. The best model was chosen based on the goodness of fit tests, the Vuong non-nested test and the information criteria. The detailed count regression analyses results (outputs) are presented in the tables below.

1) Poisson Regression Analysis: As a first step towards identifying the major factors influencing research productivity (number of Publications) of Njala university academic staff, A poison regression analysis, often referred to as a baseline count regression analysis was carried out. The result of the analysis is presented in Table 5. First, the preliminary outputs and the goodness of fit tests of the poison regression are presented. Table 4 shows that the dependent variable used in the count regression model is the "Number of publications", the probability distribution is "Poisson" and the link function a natural logarithm (i.e., "Log").

TABLE IV MODEL INFORMATION

Model Information

Dependent Variable	Number of Publications
Probability Distribution	Poisson
Link Function	Log

The categorical variable coding result presented in table 5 shows that majority of the academics selected for the sturdy were master's degree holders.

TABLE V CATEGORICAL VARIABLE CODING

Categorical Variable Information

			N	Percent
Factor	Highest Accademic Qualification	Doctorate Degree	19	16.8%
		Masters Degree	68	60.2%
		Bachelors Degree	26	23.0%
		Total	113	100.0%

Table VI provides the omnibus test used to check if the present (new) model with explanatory variables included is an improvement over the null model with intercept alone. This likelihood ratio chi-square test compares the full model against a null (intercept-only) model. The test result in Table VI is significant. This showed that a model including independent variables, teaching experience, number of hours of research, Qualification and Age fits significantly better than the null (or intercept only) model.

TABLE VI THE OMNIBUS TEST

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
284.785	5	.000

Dependent Variable: Number of Publications
Model: (Intercept), Hi_Qualif, Nuhr_Research, Teach_Expe, Age

a. Compares the fitted model against the intercept-only model.

Table VII presents parameter estimates of the poison regression. All the independent variables (factors influencing research productivity) except Age were statistically significant. More importantly, all the coefficients were positive. Meaning, an increase in any of the independent variables will increase the number of research publications. However, for the result of this analysis to be valid, the main assumption of the poison regression (equi-disperssion) must be met. This assumption therefore, needed to be verified. The result of the model verification called, model goodness of fit test, is presented in Table VIII.

TABLE VII PARAMETER ESTIMATES FROM THE POISON REGRESSION

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
			Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
(Intercept)	-2.502	.4776	-3.438	-1.566	27.444	1	.000	.082	.032	.209
[Hi_Qualif=2]	1.864	.4178	1.045	2.683	19.898	1	.000	6.448	2.843	14.624
[Hi_Qualif=3]	1.113	.4017	.326	1.900	7.675	1	.006	3.043	1.385	6.687
[Hi_Qualif=4]	0[a]	.	.	.	.	.	.	1	.	.
Nuhr_Research	.271	.0487	.175	.366	30.905	1	.000	1.311	1.192	1.442
Teach_Expe	.085	.0190	.048	.122	20.060	1	.000	1.089	1.049	1.130
Age	.015	.0079	-.001	.030	3.555	1	.059	1.015	.999	1.031
(Scale)	1[b]									

Dependent Variable: Number of Publications
Model: (Intercept), Hi_Qualif, Nuhr_Research, Teach_Expe, Age

a. Set to zero because this parameter is redundant.

b. Fixed at the displayed value.

As already mensioned, Table VIII: presents the goodness of fit test statistics for the poisson regression model. For a good fiting model, the value of the deviance divided by the degree of freedom (in Table VIII:) should be as close to 1 as possible. A value considerably less than one indicates underdispersion, greater than one indicates overdispersion and equal to one indicates equidisperssion. From Table VIII:, the value of the deviance divided by the degree of freedom (= 1.59) is considerabely greater than 1.This is an indication of a clear violation of the main assumption of the poison regression..

TABLE VIII MODEL GOODNESS OF FIT FOR POISSON REGRESSION

Goodness of Fit[a]

	Value	df	Value/df
Deviance	188.806	107	1.765
Scaled Deviance	188.806	107	
Pearson Chi-Square	197.531	107	1.846
Scaled Pearson Chi-Square	197.531	107	
Log Likelihood[b]	-186.044		
Akaike's Information Criterion (AIC)	384.088		
Finite Sample Corrected AIC (AICC)	384.880		
Bayesian Information Criterion (BIC)	400.452		
Consistent AIC (CAIC)	406.452		

Dependent Variable: Number of Publications
Model: (Intercept), Hi_Qualif, Nuhr_Research, Teach_Expe, Age

a. Information criteria are in smaller-is-better form.

b. The full log likelihood function is displayed and used in computing information criteria.

The violation of the main assumption of the poison regression (equidispersion) rendered the poison regression inappropriate for this analysis. As a result, an alternative count regression model, the negative binomial regression model that is suitable for over-dispersed count data was used in the second stage of the analysis.

2) Negative Binomial (NB) Regression Analysis: Table **IX** presents information on the probability distribution and link function for the negative binomial regression model.

TABLE IX MODEL INFORMATION

Model Information

Dependent Variable	Number of Publications
Probability Distribution	Negative binomial (1)
Link Function	Log

As already mensioned under the poisson regression, for a good fiting model, the value of the deviance divided by the degree of freedom should be as close to 1 as possible. From Table **X**, the value is 0.778 (which can be rounded up to 1). This value can allow a valid inference to be made on the dependent variable given the independent variables using the negative binomial regression.

TABLE X MODEL GOODNESS OF FIT FOR NB REGRESSION

Goodness of Fit[a]

	Value	df	Value/df
Deviance	83.264	107	.778
Scaled Deviance	83.264	107	
Pearson Chi-Square	83.137	107	.777
Scaled Pearson Chi-Square	83.137	107	
Log Likelihood[b]	-180.502		
Akaike's Information Criterion (AIC)	373.004		
Finite Sample Corrected AIC (AICC)	373.796		
Bayesian Information Criterion (BIC)	389.368		
Consistent AIC (CAIC)	395.368		

Dependent Variable: Number of Publications
Model: (Intercept), Hi_Qualif, Nuhr_Research, Teach_Expe, Age

a. Information criteria are in smaller-is-better form.

b. The full log likelihood function is displayed and used in computing information criteria.

From the test of model effect presented in table1 XI, all the predictors except Age were statistically significant.

TABLE XI TEST OF MODEL EFFECT

Tests of Model Effects

	Type III		
Source	Wald Chi-Square	df	Sig.
(Intercept)	8.454	1	.004
Hi_Qualif	26.665	3	.000
Teach_Expe	14.254	1	.000
Age	.031	1	.859
Nuhr_Research	19.081	1	.000

Dependent Variable: Number of Publications
Model: (Intercept), Hi_Qualif, Teach_Expe, Age, Nuhr_Research

Table XII and Table XIЩ contain the parameter estimates of the two models of the negative binomial regression along with their standard errors, Wald chi-square values, p-values and 95% confidence intervals for the coefficients. These tables also provide the exponentiated values of the coefficients (the "Exp(B)" column). In particular, the column headed B are the estimated negative binomial regression coefficients for the models. A positive coefficient means increase in the independent variable will result in a corresponding increase in the expected log count of the dependent variable (number of research publication). All the coefficients in the tables were positive. This implies that increase in each of the independent variables will result in a corresponding increase in the expected log count of the dependent variable.

From Table XIЩ, the variable, number of hours spent on research (Nuhr_Research) had a coefficient of 0.310, which was statistically significant. This means that, for each one hour increase in the number of hours spent on research (Nuhr_Research), the expected log count of the number of research publications (Num_Pub) increased by 0.310.

Also, the variable, teaching experience (Teach_Expe) had a coefficient of 0.131 which was statistically significant. This means that for each one-year increase in Teach_Expe, the expected log count of the number of research publications (Num_Pub) increased by 0.131.

The categorical variable, highest academic qualification (Hi_Qualif) was another significant factor in determining the number of research papers published by each NU academic. The coding of the categorical variable (or dummy variables) had the lowest qualification level group (Bachelor's degree) serving as the reference category. The regression coefficients for all dummy variables were positive suggesting that, with increasing academic qualification, the number of research publication increased.

In particular, compared to the reference category, level 4 (Bachelor's degree) of Hi_Qualif, the expected log count of the number of research publication for level 2 (master's degree) increased by 1.879. Again, compared to level 4 of Hi_Qualif, the expected log count of the number of research publication for level 3 (doctorate) of Hi_Qualif increased by 1.153.

The Exp(B) column of Table XIЩ helped to identify the change in the incident rate for anyone unit increase in the predictor variable. From Table XIЩ, the incident rate for level 2 (masters degree) of Hi_Qualif was 6.547 times the incident rate for the reference group, level 4 (Bachelor's degree). Likewise, the incident rate for level 3 (doctorate) was 3.166 times the incident rate for the reference group, holding the other variables constant.

Also, from Table XIЩ, the exponentiated value for the independent variable, Nuhr_Research was 1.364. This means that the number of research publications (i.e., the count of the dependent variable) will be 1.364 times greater for each extra hour worked per week. In other words, there is a 36.4% increase in the number of research publications for each extra hour worked by an academic per week. Similarly, from Table

XIЩ, the exponentiated value for Teach_expe was 1.140. This means that the number of research publications (i.e., the count of the dependent variable) will be 1.140 times greater for each extra year spent in the teaching field. In other words, there is a 14.0% increase in the number of publications for each extra year spent in the teaching field.

The model equation for the negative binomial regression with the log of the outcome predicted with a linear combination of the predictors was given as:

$$Log(\text{Num\_Pub}) = \beta_0 + \beta_1 \text{Hi\_Qualif}_2 + \beta_2 \text{Hi\_Qualif}_3 + \beta_3 \text{Nuhr\_Research} + \beta_4 \text{Teach\_Expe}$$

Substituting coefficients from table XIЩ gives:

$$= -2.291 + 1.879 Hi\text{Qualif}_2 + 1.153 Hi\text{Quali}_3 + 0.310\text{Nuhr\_Research} + 0.131\text{Teach\_Expe}$$

This implies:

$$\text{Num\_Pub} = \exp\left(-2.291 + 1.879 Hi\text{Qualif}_2 + 1.153 Hi\text{Qualif}_3 + 0.310\text{Nuhr\_Research} + 0.131\text{Teach\_Expe}\right)$$

$$= \exp(-2.291) * \exp\big(1.879(Hi\_Qualif_2)\big) * \exp\big(1.153(Hi\_Qualif_3)\big) * \exp\,(0.310(Nuhr\_Research)) * \exp\,(0.131(Teach\_Expe))$$

TABLE XII NEGATIVE BINOMIAL REGRESSION MODEL (FULL MODE)

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
			Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
(Intercept)	-2.414	.7625	-3.909	-.920	10.022	1	.002	.089	.020	.399
[Hi_Qualif=2]	1.864	.5620	.763	2.966	11.002	1	.001	6.451	2.144	19.411
[Hi_Qualif=3]	1.156	.4936	.188	2.123	5.483	1	.019	3.177	1.207	8.359
[Hi_Qualif=4]	0[a]	.	.	.	.	.	.	1	.	.
Nuhr_Research	.311	.0878	.139	.483	12.528	1	.000	1.365	1.149	1.621
Teach_Expe	.124	.0459	.034	.214	7.328	1	.007	1.132	1.035	1.239
Age	.004	.0173	-.030	.038	.046	1	.830	1.004	.970	1.038
(Scale)	1[b]									
(Negative binomial)	1[b]									

Dependent Variable: Number of Publications
Model: (Intercept), Hi_Qualif, Nuhr_Research, Teach_Expe, Age

a. Set to zero because this parameter is redundant.

b. Fixed at the displayed value.

Table **XIЩ** presents output for the Negative Binomial Regression with the independent variable, Age excluded (as age has proved to be a non significant predictor of the dependent variable)

TABLE XIЩ NEGATIVE BINOMIAL REGRESSION MODEL (REDUCED MODE)

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
			Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
(Intercept)	-2.291	.4978	-3.266	-1.315	21.172	1	.000	.101	.038	.268
[Hi_Qualif=2]	1.879	.5578	.786	2.972	11.348	1	.001	6.547	2.194	19.535
[Hi_Qualif=3]	1.153	.4933	.186	2.119	5.458	1	.019	3.166	1.204	8.327
[Hi_Qualif=4]	0[a]	.	.	.	.	.	.	1	.	.
Nuhr_Research	.310	.0877	.138	.482	12.505	1	.000	1.364	1.148	1.619
Teach_Expe	.131	.0340	.064	.197	14.801	1	.000	1.140	1.066	1.218
(Scale)	1[b]									
(Negative binomial)	1[b]									

Dependent Variable: Number of Publications
Model: (Intercept), Hi_Qualif, Nuhr_Research, Teach_Expe

a. Set to zero because this parameter is redundant.

b. Fixed at the displayed value.

3*) Comparison of the Negative Binomial regression models*: Two different models were used for the negative binomial regression analysis. Model 1 was the reduced model (from Table XIII) with three independent variables, academic qualification, hours spent on research and teaching experience. Whereas model 2 was the full model (from Table XII) with all the independent variables used in model 1 plus one additional independent variable, age included. The Vuong test for comparing the two models is presented in Table XIV. From Table XIV, the Vuong z-statistic was positive. The positive Voung z-statistics suggested that model 1 (reduced model) was superior to model 2 (full model). However, based on the p-values, the difference was not statistically significant.

TABLE XIV VUONG TEST TO COMPARE NEGATIVE BINOMIAL MODELS

Vuong Hypothesis Test-Statistic			
Model 1 is the reduced model with3 IVs and model 2 is the full model with 4 IVs			
	Vuong z-statistic	H_A	p-value
Raw	0.2252974	model1 > model2	0.41087
AIC-corrected	0.2252974	model1 > model2	0.41087
BIC-corrected	0.2252974	model1 > model2	0.41087

As already mentioned, the negative binomial regression model with 3 independent variables was chosen as the best fitting model based on the Voung test. However, since the value of the deviance divided by the degree of freedom (value/df, as in Table **X)** is less than 0.9, additional count regression models, the Zero inflated regression models were considered in the analysis.

Also, fig1 revealed that, out of the 113 NU academics considered in the study, more than 50 had 0 (zero) peer-reviewed publications. Was this due to the fact that these academics did not attempt to publish at all ("true zeros")? Or did they attempt to publish but due to some infringing factors were not able to publish? To answer these questions and to account for the excess Zeros in the analysis, the zero inflated regression models were also fitted to the data.

4) Zero-inflated models: Using the zero inflated regression models in count data analysis cannot be overruled, as most empirical data often show more zeroes than would be expected under most count data models, including both Poisson and negative binomial regression models. The zero-inflated models are two-part models that attempt to account for excess zeros. There is thought to be two kinds of zeros, "true zeros" and excess zeros. Therefore, the zero-inflated models estimate two equations, one for the count model and one for the excess zeros. This is depicted in the outputs presented in Tables XV and XVII.

The output of the Zero-Inflated poison model presented in Table **XV** consists of two blocks. The first block contains the Poisson regression coefficients for each of the variables along with standard errors, z-scores, and p-values for the coefficients. The second block corresponds to the inflation model. This block includes the logit coefficients for predicting excess zeros along with their standard errors, z-scores, and p-values. Two of the predictors, teaching experience and the number of hours spent on research were each statistically significant in both the count (first block) and inflation portions (second Block) of the model.

The value of the chi-squared test computed on the difference of log likelihoods for the current zero-inflated model with all the predictors and the null model without predictors was 5.490042e-60. Since there are three predictor variables, the degree of freedom for the chi square test was 3. This yielded a high significant p-value. Showing that, the current model fits the data significantly better than the null model.

Table XV PARAMETER ESTIMATES OF ZERO-INFLATED POISON MODEL

Count model coefficients (poisson with log link)				
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.92335	0.20385	-4.530	5.91e-06 ***
Teach_Exp	0.10365	0.01312	7.901	2.76e-15 ***
Hi_Qualif	0.59137	0.09956	5.940	2.85e-09 ***
Nresearch	0.19950	0.03975	5.019	5.20e-07 ***
Zero-inflation model coefficients (binomial with logit link)				
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.13195	0.61601	1.838	0.0661
Teach_Exp	-0.12464	0.05922	-2.105	0.0353 *
Hi_Qualif	0.29514	0.32314	0.913	0.3611
Nresearch	-0.72971	0.17678	-4.128	3.66e-05 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

- *Interpretation of Exponentiated coefficients for the Zero-inflated poison model*

From the 'Zero-inflation model' column of Table XVI, the baseline odds of being among those who never published is 3.10. Each unit (year) increase in teaching experience decreased the odds of being among those who never published by 0.88 and each hour increase in number of hours spent on research decreased it by 0.48. Highest academic qualification increased the odds of being among those who never published, but it is non-significant

Also, from the 'Count model'column of Table XVI, the baseline number of research publication is 0.397 for those who had a chance of publishing research papers (articles or books). A unit (year) increase in teaching experience increased the number by 1.109 times and a unit increase in academic qualification increased it by 1.109. Also, a unit (hour) increase in number of hours spent on research increased the baseline number of research publication by 1.22.

Table XVI EXPONENTIATED COEFFICIENTS FOR ZERO-INFLATED MODEL

	Count_model	**Zero-inflation_model**
(Intercept)	0.3971844	3.1016977
Teach_Exp	1.1092172	0.8828183
Hi_Qualif	1.8064671	1.3433084
Nhr_research	1.2207950	0.4820486

- *Zero-inflated Hurdle Model*

As already mentioned under methodology, the hurdle model is a two-part model. The first part of the model is a binary logit model that models whether an observation takes a positive count or not. The second part of the model can be a truncated Poisson or Negative Binomial model but for the purpose of this work, a truncated poisson model was used and therefore, one part (or process) of the hurdle model governs whether an academic published a research paper or not and the other part governs how many publications were made. The model output for the zero-inflated hurdle model is presented in Table **XVII**. All the predictors in the truncated part of the zero-inflated hurdle model were significant, whiles only teaching experience (Teach_Exp) and the number of hours spent on research (Nhr_Research) were significant in the hurdle part of the model. This model fits the data significantly better than the null model, i.e., the intercept-only model. This was justified by comparing the present model to a null model without predictors using chi-squared test on the difference of log likelihoods. The result, 'log Lik. 4.315565e-34 of the chi-square test is statistically significant, showing that the present zero-inflated hurdle model is an improvement over the intercept only model.

TABLE XVII PARAMETER ESTIMATES OF ZERO-INFLATED HURDLE MODEL

Count model coefficients (truncated poisson with log link):				
	Estimate	**Std. Error**	**z value**	**Pr(>|z|)**
(Intercept)	-0.87417	0.19907	-4.391	1.13e-05 ***
Teach_Exp	0.10569	0.01326	7.969	1.60e-15 ***
Hi_Qualif	0.54329	0.09802	5.543	2.98e-08 ***
Nresearch	0.20245	0.03965	5.106	3.29e-07 ***
Zero hurdle model coefficients (binomial with logit link):				
	Estimate	**Std. Error**	**z value**	**Pr(>|z|)**
(Intercept)	-2.41753	0.41406	-5.839	5.26e-09 ***
Teach_Exp	0.16947	0.05054	3.353	0.0008 ***
Hi_Qualif	0.18964	0.24324	0.780	0.4356
Nresearch	0.72456	0.13199	5.490	4.03e-08 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

log Lik.' 4.315565e-34 (df=9)

➢ *Interpretation of exponentiated coefficients for zero-inflated hurdle model*

From the 'Zero-hurdle model' column of table XVIII, the baseline odds of having a positive count vs zero was 0.089.This odd was increased by 1.18 times by a unit (year) increase in teaching experience and it was increased by 2.064 times by a unit increase in the number of hours spent on research. Highest academic qualification increased it by 1.2 times but does not have a significant effect.

Also, given that the response is positive in the "positive count model' column of Table XVIII, the average count was 0.306. A unit (or year) increase in teaching experience increased this average count by 1.134 times among those who had positive counts and the number of hours spent on research increased it by 1.226. Highest academic qualification increased the average count by 1.849 though not significant.

TABLE XVIII EXPONENTIATED COEFFICIENTS FOR ZERO-INFLATED HURDLE MODEL

	Count_model	Zero_hurdle_model
(Intercept)	0.3061956	0.08914139
Teach_Exp	1.1344579	1.18467165
Hi_Qualif	1.8492469	1.20881888
Nresearch	1.2255595	2.06381259

6) Model comparison for Zero-inflated Models: The zero inflated poison and zero inflated hurdle models were compared to find the best fitting model.

Table1 XIX presents the Akaike information criteria (AIC) to compare the zero-inflated poison and the zero inflated hurdle models. The values of the AIC for the two models showed that the zero inflated hurdle model with the least AIC performed better than the zero-inflated poison model.

TABLE1XIX: AKAIKE INFORMATION CRITERIA (AIC) TO COMPARE ZERO-INFLATED MODELS

No	**Model**	**Link- Function**	**AIC**
1	zero-inflated hurdle	'truncated poisson with log link'and binomial with logit link}	650.8079
2	zero-inflated poison	'poisson with log link' and 'binomial with logit link'	657.5365

5) Vuong test to compare poisson, negative binomial and zero-inflated models

The Vuong test, implemented by the pscl package can test two non-nested models [30]. The null hypothesis was that there is no difference in the models in terms of performance.

➢ *Vuong Non-Nested Hypothesis Tests*

The null hypothesis for the test is that the two models are indistinguishable. From Table XX, Model 1 is the hurdle model and model 2 is the zero-inflated model. The positive value of the Vuong z-statistic suggests that model 1 is superior. However, the difference is not statistically significant.

TABLE XX VUONG TEST TO COMPARE ZERO-INFLATED HURDLE AND ZERO-INFLATED POISSON MODELS

Vuong Non-Nested Hypothesis Test-Statistic			
Model 1 is zero-inflatedhurdle and model 2 is zero-inflatedpoison			
	Vuong z-statistic	H_A	p-value
Raw	0.8358104	model1 > model2	0.20163
AIC-corrected	0.8358104	model1 > model2	0.20163
BIC-corrected	0.8358104	model1 > model2	0.20163

➢ *Model Comparison for the Negative Binomial and the Zero-inflated Models*

Based on [30], the voung non-nested test for comparing two non-nested models was used to compare the zero inflated poison and negative binomial models.. From Table XXI, model 1 was the zero-inflated poison model and model 2 was the negative binomial model. The positive value of the Vuong z-statistic suggested that model 1 was superior, but the p_value showed that the difference between the two models was not statistically significant.

TABLE XXI VUONG TEST TO COMPARE ZERO-INFLATED POISON AND NEGATIVE BINOMIAL MODEL

Vuong Non-Nested Hypothesis Test-Statistic			
Model 1 is zero-inflatedpoison and model 2 is negative binomial			
	Vuong z-statistic	H_A	p-value
Raw	0.6748894	model1 > model2	0.24987
AIC-corrected	0.2250282	model1 > model2	0.41098
BIC-corrected	-0.5267651	model1 > model2	0.29918

The Vuong test presented in table **XXЦ** compares the zero-inflated hurdle model with the negative binomial regression model. From table XXЦ, model 1 is the zero-inflated hurdle model and model 2 is the negative binomial model. The test statistic is significant, indicating that the zero-inflated hurdle model was superior to the negative binomial regression model. If the Raw statistic was used, p = 0.001 gives a strong evidence that model1 (the zero-inflated hurdle model) was superior to model 2 (the negative binomial regression model).

TABLE XXЦ VUONG TEST TO COMPARE ZERO-INFLATED HURDLE AND NEGATIVE BINOMIAL MODELS

Vuong Non-Nested Hypothesis Test-Statistic			
Model 1 is zero-inflatedhurdle and model 2 is negative binomial			
	Vuong z-statistic	H_A	p-value
Raw	3.04416	model1 > model2	0.00116
AIC-corrected	2.08144	model1 > model2	0.01869
BIC-corrected	0.47257	model1 > model2	0.31825
Vuong Non-Nested Hypothesis Test-Statistic			
Model 1 is zero-inflatedhurdle and model 2 is negative binomial			
	Vuong z-statistic	H_A	p-value
Raw	3.0441681	model1 > model2	0.0011666
AIC-corrected	2.0814457	model1 > model2	0.0186966
BIC-corrected	0.4725758	model1 > model2	0.3182579

V. RESULTS AND DISCUSSION

This research was carried out to identify the main factors influencing the research productivity of Njala University academics. It also aimed to point out the effect of each factor on the research output of the academics. In line with the work of [35], research productivity was measured as publication counts and defined as the self-reported number of journal articles and chapters in academic books that the academic (respondent) has published after the Sierra Leone Rebel war.

The data were first analyzed using the baseline count regression model, called the poison regression model. However, due to the presence of over-dispersion in the data, (i.e.,value\df=1.765 >1 in Table VIII), the main assumption of the poison regression model was violated which rendered the poison regression unfit for the data analysis.

To overcome the restrictive assumption (equi-dispersion) of the poison regression and to account for the excess zeros in the count data variable (dependent variable), three additional count regression models were used in the analysis. These were: the negative binomial regression model; the zero-inflated poison regression model and the zero-inflated hurdle regression model. Statistical tests showed that the zero-inflated hurdle model outperformed all the other count regression models, followed by the Zero-inflated poison model and then the negative Binomial regression model.

The results of all the count regression analysis considered in this work showed that: the number of hours spent on research (Nuhr_research), the teaching experience (Teach_Exp) and the highest qualification (Hi_Qualif) of the academic staff were the main factors influencing the number of research publications made by the academics of Njala University. The result of the analysis pointed out that, increase in the number of hours spent on research (Nuhr_Research), increased the expected log count of the number of research publications. Specifically, the exponentiated value of the independent variable, Nuhr_Research presented in Table **XЩ** showed that the number of research publications will be 1.364 times greater for each extra hour spent on research per week. In other words, there is a 36.4% increase in the number of research publications for each extra hour spent on research per week. This is supported by the work of [20], who mentioned in their paper that, the time and energy required to pursue research is limited by the time demands of teaching. Reference [6] also noted that an unsatisfactory classroom performance might result from academics neglecting their teaching responsibilities to pursue research and publications. In other words, effective and committed classroom teaching can be an opportunity cost for an increase in the number of research publications made by academic staff.

Similarly, from Table **XЩ**, the variable Teach_Expe has a positive coefficient of 0.124 which is statistically significant. This means that, for each year increase in Teach_Expe, the expected log count of the number of research publications increased by 0.124 amounts. Also, the exponentiated value for Teach_expe is 1.140. This means that the number of research publications will be 1.140 times greater for each extra year worked in the teaching field. In other words, there is a 14.0% increase in the number of publications for each extra year spent in the teaching field. This finding is in line with the work of [18], which stated that the acquisition of skills and experiences enhances productivity.

The qualification (Hi_Qualif) of the academic is another significant factor in determining the number of research papers published by each NU academic. Qualification was coded as a dummy variable (categorical variable) with the lowest qualification (Bachelor's degree) serving as the reference category. The regression coefficient for all the dummy variables were positive suggesting that increase academic qualification increased the number of research publications. This is supported by [11] who also indicated in research finding that staff qualifications positively influenced research output.

Age, though not a statistically significant factor (see table Table **XII**) in influencing the number of research publications (p-value > 0.05), the regression coefficient for Age is positive, meaning an increase in Age will result to a corresponding increase in the number of research publications. Also, the exploratory analysis showed that age has a positive influence on the number of research publications as exhibited in fig 5 scatter plot 4 under exploratory analysis. This is not surprising as many studies have found that, the average production of publications increases with age and reaches a peak at some point during the career and then declines (see for instance [1], [3], [16] and [26]).

More importantly, from the result (output) of the Zero-inflated poison regression analysis, in the 'Zero-inflation model' column of Table XVI, the baseline odds of being among those who never published was 3.10. Each unit (year) increase in teaching experience decreased the odds of being among those who never published by 0.88 and each hour increase in the number of hours spent on research decreased it by 0.48.

VI. CONCLUSION

The purpose of this work was to identify the main factors influencing research productivity measured by the number of research publications made by each NU academic after the Sierra Leone Rebel War. The dependent variable was a count variable, therefore, in addition to the poison regression, three additional count regression models were used in the analysis. The results of the analysis showed that the number of hours spent on research, teaching experience and the highest qualification of the academics are the main factors influencing the number of research papers (or Books) published by the academics at Njala University. Increase in teaching experience results in an increase in research productivity and the higher the academic qualification, the greater the possibility of publishing more research papers (or books). Also, an increase in the number of hours spent on research is positively related to an increase in the number of research publications. Based on the result of the count regression analysis, there is a 36.4% increase in the number of research publications for each extra hour spent on research per week. This means that any activity, like extra service to the university, administrative activity and teaching workload that may influence the number of hours spent on research may indirectly influence the number of research publications. This is in line with the work of previous authors, who suggested that publication rate also depends on a wide range of factors that cannot easily be measured, such as the availability of research funds; teaching loads; equipment; research assistants; workload policies; departmental culture and working conditions; organizational context; and talent and hard work (e.g. [13], [27] and [33]).

VII. RECOMMENDATIONS

The number of hours spent on research was one of the main factors influencing research productivity. Therefore, any activity like extra service to the university, administrative activity and teaching workload that may influence the number of hours spent on research may indirectly influence the number of research publications made by the academic staff. Though research promotes quality teaching, if 99% of the promotion criteria is tied to publication, effective and dedicated teaching of students may turn out to be the opportunity cost of publishing more research papers by the academics. It is therefore recommended that a conducive research environment be created that will allow the academics to spend more time on research activities or the promotion criteria for academic staff be reviewed such that at most 50% of the total score for promotion be allocated to research while the remaining 50% be allocated to: other university activities; professional activities (e.g. journal reviewing and editing), administrative activities and community services.

REFERENCES

[1]. Aksnes, D. W., Rorstad, K., Piro, F., & Sivertsen, G.(2011). Are FemaleResearchers Less Cited? A large-scale study of Norwegian scientists. Journal of theAmerican Society for Information Science and Technology (JASIST), 62(4), 628–636.

[2]. Baccini& L. Barabesi& M. Cioni& C. Pisani, 2014. “Crossing the hurdle: the determinants of individual scientific performance”. Scientometrics, Springer: AkademiaKiado. Vol.101.(3), pages2035-2062, December

[3]. Barjak, F. (2006). Research productivity in the inter-net era. Scientometrics, 68(3), 343–360.

[4]. Benoit Godin and Yves Gingras, the place of universities in the system of knowledge production Research Policy, 2000, vol. 29, issue 2, 273- ...

[5]. Best, J., and J. Kahn (1998). Research in education (8th ed.).

[6]. Blackburn, Robert T.1974 "The Meaning of Work in ACademia," chapter inMeasuring Faculty Effort, James 1. Dol (ed.), San Francisco: Jossey-Bass.

[7]. Bordons, M., Morillo, F., Fernandez, M. T., & Gomez, I. (2003). One step further in the production of bibliometric indicators at the micro level: Differences by gender and professional category of scientists. Scientometrics, 57(2), 159–173

[8]. Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. Psychometrika, 52, 345370.

[9]. Boyer, E.L. (1990). Scholarship reconsidered: Priorities of the professoriate. The Carnegie Foundation for the Advancement of Teaching. New York: John Wiley and Sons

[10]. Cameron AC, Trivedi PK (2013). *Regression Analysis of Count Data*. Cambridge University Press, Cambridge

[11]. Chepkorir K. R. (2018) Effect of academic staff qualification on research productivity in Kenyan public universities; evidence from moi university. International Journal of Economics, Commerce and Management United Kingdom http://ijecm.co.uk/Vol. VI, Issue 2, February 2018

[12]. Duch, J., Zeng, X. H. T., Sales-Pardo, M., Radicchi, F., Otis, S., Woodruff, T. K., et al. (2012). The possible role of resource requirements and academic career-choice risk on gender differences in publication rate and impact. PLoS ONE, 7(12), 11.

[13]. Dundar, H., & Lewis, D. R. (1998). Determinants of research productivity in higher education. Research in Higher Education, 39, 607–631.Engels,

[14]. Etzkowitz, H., &Leydesdorff, L. (1997). Universities and the global knowledge economy: A triple helix ofuniversity-industry-government relations. Pinter London.

[15]. Godin, B. and Gingras, Y. (2000).The place of universities in the system of knowledge production. Research policy, 29(2), 273-278.

[16]. Gonzalez-Brambila, C., & Veloso, F. M. (2007). The determinants of research output and impact: A study of Mexican researchers. Research Policy, 36(7),1035–1051.

[17]. H. Akaike, (1974) "A new look at the statistical model identification," in IEEE Transactions on Automatic Control, vol.

19, no. 6, pp. 716-723, December 1974, doi:10.1109/TAC.1974.1100705.

[18]. Hatch, N.W. and Dyer, J.H. (2004) 'Human capital and learning as a source of sustainable competitive advantage', Strategic Management Journal, Vol. 25, pp.1155–1178

[19]. Hemmings, B.C., Rushbrook, P. and Smith, E. (2007) 'Academics' views on refereed works: a content analysis', Higher Education, Vol. 54, pp.307–332

[20]. Herbert W. Marsh and John Hattie (2002)The Relation between Research Productivity and Teaching Effectiveness. *The Journal of Higher Education* Vol. 73, No. 5 (Sep. - Oct., 2002), pp. 603-641

[21]. Hilbe, J. (2011). *Negative Binomial Regression*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511973420

[22]. Hilbe, J. (2014). *Modeling Count Data*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139236065

[23]. Hunter L. A, Leahey Erin. Parenting (2010) Research Productivity: New Evidence and Methods. Social Studies of Science 40(3):433-451 ISSN 2348 0386

[24]. Jauch, L.R. &Glueck, W.F. (1975). Evaluation of university professors' research performance.Management Science, 22 (1), 66-75.

[25]. Kristoffer Rørstad, Dag W. Aksnesl (2015) Poblication rate expressed by age gender and cademic Position a- A large scale analysisi of Norwegian academic staff. Journal of Informetrics 9(2015) 317-333

[26]. Kyvik, S. (1990). Age and scientific productivity. Differences between fields of learning. Higher Education, 19, 37–55.

[27]. Kyvik, S. (1993). Academic staff and scientific production. Higher Education Management, 5(2), 191–202

[28]. Lewison, G. (2001). The quantity and quality of female researchers: A bibliometric study of Iceland. Scientometrics, 52(1), 29–43.

[29]. Long, J.S. (1992). Measures of sex differences in scientific productivity. Social Forces, 71(1), 159–178.

[30]. Mangiafico, S.S. 2016. Summary and Analysis of Extension Program Evaluation in R, version 1.18.1

[31]. Meador, M., Walters, S. J. K., & Jordan, J. M. (1992). Academic research productivity. Economics of Education Review 11(2), 161-167.

[32]. Olson, J. E. (1994). Institutional and technical constraints on faculty gross productivity in American doctoral universities. Research in Higher Education 35(5), 549-567.

[33]. Ramesh, B. A., & Singh, Y. P. (1998). Determinants of research productivity. Scientometrics, 43(3), 309–329.

[34]. Robert Wamala Vincent A. Ssembatya , (2015),"Productivity in academia", Quality Assurance in Education, Vol. 23 Iss 2 pp. 184 - 195

[35]. Teodorescu D (2000) Correlates of faculty publication productivity: a cross-national analysis.Higher Edu. 39,201-222.

[36]. Vuong, Quang H. (1989). Likelihood Ratio Tests for Model Selection and non-nested Hypothesis"Econometrica 57(2): 307-333. doi:10:2307/1912557.JSTOR 1912557

[37]. Wamala and Ssembatya (2015), even stated that productivity in the academe is consensually regarded as an indicator of research activity

[38]. Wu, J. (2013). Geographical knowledge diffusion and spatial diversity citation rank. Scientometrics, 94(1),181-201. http//dx.doi.org/10.1007/81192.012.0715.8

[39]. Xie, Y., & Shauman, K. A. (1998). Sex differences in research productivity: New evidence about an old puzzle. American Sociological Review, 63(6), 847–870