Journal of Scientific Research & Reports

24(6): 1-11, 2019; Article no.JSRR.51535
ISSN: 2320-0227

Modelling of Instructors Publication Factors in Ethiopia Public Universities: Advanced Count Regression Models

Alebachew Abebe[1*]

[1]Department of Statistics, College of Computing and Informatics, Haramaya University, P.O.Box 138, Dire Dawa, Ethiopia.

Author's contribution

The sole author designed, analysed, interpreted and prepared the manuscript.

Article Information

DOI: 10.9734/JSRR/2019/v24i630168
Editor(s):
(1) Dr. Luigi dell'Olio, Professor, School of Civil Engineering, Channels and Ports, University of Cantabria, Spain.
(2) Dr. Shunpu Zhang, PhD, Department of Statistics, University of Nebraska–Lincoln, USA.
Reviewers:
(1) Pasupuleti Venakata Siva Kumar, Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering and Technology, India.
(2) Aschalew Kassu, Alabama A&M University, USA.
(3) Chipo Zidana, National University of Science and Technology, Zimbabwe.
Complete Peer review History: https://sdiarticle4.com/review-history/51535

Original Research Article

Received 12 July 2019
Accepted 19 September 2019
Published 01 October 2019

ABSTRACT

Instructors' publication (IP) is one of the major activity in higher education institutes. Currently, IP faced problem both high prevalence and severity in Ethiopia public universities. Even if the problem is common to both developed and developing countries, about 352 (73.9 %) of the instructors employed by public universities in Ethiopia have been affected by a lack of scholarly publications. Since the outcomes from IP factors are mostly discrete variable; they are often modelled using advanced count regression models. The purpose of this study was to model the appropriate count regression model that efficiently fit the IP data and further to identify the key risk factors contributing significantly to IP in public Universities in Ethiopia. The data were collected between November 2015 through November 2016 from selected thirteen (13) public universities in Ethiopia through both questionnaires and interview. The cross-sectional study design was employed using IP data. A simple random sampling technique was applied to the population of Ethiopia public universities to obtain a sample of 13 universities or 476 individual instructors were selected. The average age of the 476 participants was found to be 30 years with 31(6.5%)

**Corresponding author: E-mail: aleb.abebe@yahoo.com; aleb.abebe19@gmail.com;*

being females and 445(93.5%) being males. The count outcomes obtained were modelled using count regression models which included Zero-Inflated Negative Binomial (ZINB), Zero-Inflated Poisson (ZIP) and Poisson Hurdle regression models. To compare the performance and the efficiency of the listed count regression models concerning the IP data, the various model selection methods such as the Vuong Statistic (V) and Akaike's Information Criterion (AIC) were used. The ZINB count regression model concerning the values of the Vuong Statistic and AIC was selected as the most appropriate and efficient count regression model for modelling IP data. Based on the ZINB model the variables age, experience, average work-load, association member and motivation to work were statistically significant risk factors contributing to IP in Ethiopia public universities.

Keywords: Instructors' publication; zero-inflated negative binomial; zero-inflated poisson; poisson hurdle; vuong statistic.

1. INTRODUCTION

Innovation has always had a great impact on economic development. However, it is only recently, (particularly after the late 1950s) that the instrumentality of innovation started to get wide recognition amongst scholars [1]. Understandably, being the most intangible of the factors of production (land, labor, capital), it has remained unaccounted for, by economists for hundreds of years [2]. The transition to the 21st century, which is widely characterized as the knowledge era, however, has amplified the importance attached to innovation [3]. Being cognizant of its instrumentality, private corporations in the industry are allocating a significant sum of their budget in R & D activities [4]. From the governments part also, a trend of building and revitalizing national innovation capacities is evident globally. Likewise, there is an increasing expectation from Higher education institutions (HEIs) to play their part in fostering innovation in today's knowledge age [5,6].

In this token, the sufficiency of the two long-standing objectives of HEIs (teaching and research) has been brought to the spotlight [7]. Creating knowledge for the sake of merely knowing does not suffice anymore. HEIs are expected to apply the knowledge created for solving societal problems and bringing about socio-economic development [8]. At the same time, there is an increasing demand for quality and accountability, among other things [9]. Interestingly enough, such expectations are looming in the face of a declining pattern of government funding; all necessitating entrepreneurial response [5]. According to [10] the survival of university in today's turbulent environment to a great extent depends on their ability to infuse entrepreneurship in their administration, faculty and students [11].

This increasing expectation from universities has also resulted in an increasing number of publications on academic entrepreneurship [12]. According to, [12], the pool of knowledge on university entrepreneurship can generally be classified in to four broad categories: (i) entrepreneurial research university [13] (ii) productivity of technology transfer offices [14] (iii) new firm creation [15] and (iv) environmental context including networks of innovation [16].

Instructors publication is a major problem in Ethiopia public universities yet there is passive researchers are very large or little knowledge of methodology to publish a manuscript. Therefore, imposes a considerable burden not only on the individual instructor but also the economy of Ethiopia to reduce poverty and to improve its' economic growth [5]. Based on this gab, the research seeks to determine robust count regression model that efficiently fits publication data and also use to identify the key risk factors contributing to instructors' publication in Ethiopia public universities.

2. MATERIALS AND METHODS

2.1 Data Source

To address the main objectives of the study, a cross-sectional study design was employed using well-structured and unstructured questionnaires with an interview. The data was collected between November 2015 to November 2016 at selected thirteen (13) Ethiopia public universities. A simple random sampling technique was employed to the population at the universities to select a sample of 476 instructors including senior instructors.

2.1.1 Study sample

The data collection was in two parts: A quantitative survey and document check listed. A cross-sectional study design, the data were collected from thirteen (13) Ethiopian public universities (n=476) were selected for the analysis such as: Addis Ababa 31(6.5%), Bahir Dar 35(7.4%), Gondar 53(11.1%), Mekelle 27(5.7%), Adigrat 34(7.1%), Wollo 42 (8.8%), Semera 40(8.4%), Adama 28(5.9%), Jimma 36(7.6%), Medawolabu 34(7.1%), Hawasa 37(7.8%), Arba Minch 38(8.0%) and Haramaya 41(8.6%) universities were included for this study.

2.1.2 Variables in the study

The purpose of this research is to determine an appropriate advanced count regression model suitable for the analysis and to identify the risk factors considering instructors publication in Ethiopia public universities. The dependent or the outcome variable used for this study was instructors' publication in Ethiopia public universities and it is a continuous and a count data. Also, the independent variables were grouped in three parts below in the part-I (Table 1), part-II (Table 2), and part-III (Table 3) were presented respectively under the results part.

2.1.3 Statistical data analysis

After data were edited, coded and entered to the statistical software SPSS Version-25.0 and SAS Version-9.4 for further analysis. The instructors' publication was treated as the response variable and the remaining were explanatory variables. Descriptive statistics were presented in three parts. Those are multi-categorical variables, dichotomous factors and continuous variables were analyzed. The histogram also was plotted. Inferential Statistics both variables were used for comparison of counted regression models such as Poisson Hurdle Regression Model and Zero-inflated models were used to fit the best robust model by using maximum likelihood estimation techniques.

Part-I: Out of the total of 476 respondents based on the marital status of instructors who have a married status were 180(37.8%), unmarried status was 293(61.6%) and divorce status was 3(0.6%) presented respectively. Similarly, for others, multi-categorical variables have the same fashion of interpretation as presented the Table 1.

Part-II: From the total respondents of 476 instructors who have a gender of male 445(93.5%) and the remaining 31(6.5%) were female respondents. Instructors who have an administrative position at the university level 108(22.7%) and the rest who haven't an administrative position at the university level were 368(77.3%). Likewise, other factors have the same fashion of interpretation.

Table 1. Personal information about multi-categories variables (n=476)

Variables	Categories	Frequency (n)	Percentage
Marital Status	Married	180	37.8%
	Unmarried	293	61.6%
	Divorce	3	0.6%
Higher education level	Bachelor: BSc/BA	125	26.3%
	Master: MSc/MA	323	67.9%
	PhD	27	5.7%
	Specialty certificate	1	0.2%
Present academic rank	Technical assistant	2	0.4%
	Graduate assistant	49	10.3%
	Assistant lecturer	67	14.1%
	Lecturer	319	67.0%
	Assistant professor	34	7.1%
	Associate professor	3	0.6%
	Professor	2	0.4%
Lived in	University apartment on campus	49	10.3%
	University apartment outside campus	110	23.1%
	In rented house	299	62.8%
	In own house	14	2.9%
	Others	4	0.8%

Part-III: Off the total 476 respondents instructors who have no publication was the minimum amount, the maximum amount of publication was 48, the average value of publication 1.16 and with a standard deviation of 4.182. In the same fashion for others covariates there minimum, maximum, mean and standard deviation were presented respectively.

2.2 Inferential Statistics

2.2.1 Regression Models (RMs)

In RMs, count data such as the number of publication in Ethiopian public universities instructors are better modelled using Zero-Inflated Negative Binomial, Zero-Inflated Poisson and Poisson Hurdle Regression Models since it assumes non-negative values, discrete and often Zero-Inflated. These, regression models employed to model the outcomes of the number of publication of instructors are briefly explained as follows:

(i) Poisson Hurdle Regression Model (PHRM)

Many count data exhibit more zero counts and are also over-dispersed. One type of count regression model that is capable of dealing with both excess zeros and over-dispersion in the PHRM, which was proposed [17]. The Poisson Hurdle count regression model is a two-state model; a binary component to predict zeros and a zero-truncated component such as the Poisson to predict the non-zero counts. The probability density function of the Poisson Hurdle model is given by:

$$P\left(Y_i / X_i, Z_i\right) = \frac{(1-\omega_i)\exp(-\mu_i)\mu_i^{\,y}}{(1-\exp(-\mu_i)y!} \quad for\ \ y > 0 \tag{1}$$

Where $\mu_i = \exp(X_i^T \beta)$. The variance and mean of a PHM according to [17] are given as:

$$Var\left(Y_i / X_i, Z_i\right) = \eta(\mu_i - \eta) + \frac{\prod \sigma^2}{1 - P(0;\alpha)} \quad and$$

$$E\left(Y_i / X_i, Z_i\right) = \eta - \frac{\prod \sigma^2}{1 - P(0;\alpha)}$$

The PHM combines a zero-truncated component which is specified more formally as $f_{count}(y, X_i, \beta)$ and the hurdle component, that models the zero counts, which is also specified as $f_{zero}(y, Z_i, \gamma)$ and which as a result is given by the relation:

$$f_{hurdle}(y, X_i, Z_i, \beta, \gamma) = \left\{\begin{matrix} f_{zero}(0, Z_i, \gamma)\ if\, y=0 \\ 1 - f_{zero}(0, Z_i, \gamma) f_{count}(y, X_i, \beta) / f_{count}(0, X_i, \beta)\ if\ y>0 \end{matrix}\right\} \tag{2}$$

The parameters $\gamma\ and\ \beta$ of the PHM can be estimated using the Maximum Likelihood Estimation and the advantage is that the Zero-truncated component and hurdle component can be maximized separately by the likelihood specification. The likelihood function of the PHM has the general form:

$$L = \prod_{i=U_0} \{f_{zero}(0; Z_i, \gamma)\} \prod_{i=U_1}^{\{(1 - f_{zero}(0; Z_i, \gamma))^* f_{count}(y, X_i, \beta)/1 - f_{count}(0; X_i, \beta)\}} \tag{3}$$

Where

$\Omega_0 = (i / y = 0),\ \Omega_1 = (i / y \neq 0)\ \ and\ \Omega_0 \cup \Omega_1 = \{1, 2, ---, N\}$

The log-likelihood, by taking the log of the likelihood function and rearranging the terms, gives the following relation:

$$l = \prod_{i=U_0} \ln\{f_{zero}(0; Z_i, \gamma)\} + \sum_{i=U_1} \{1 - f_{zero}(0; Z_i, \gamma)\} + \sum_{i=U_1} \ln\{f_{count}(y, X_i, \beta)\} - \ln\{1 - f_{count}(0, X_i, \beta)\} \tag{4}$$

The log-likelihood cannot all times be expressed as the sum of the log-likelihoods from the two different models because the likelihood function is separable with regards to the parameter vectors $\beta\ and\ \gamma$. The mean regression relationship can be expressed as: by a Canonical Link.

$$\log(\mu) = X_i\beta + \log(f_{zero}(0; Z_i, \gamma)) - \log(1 - f_{count}(0, X_i, \beta)) \tag{5}$$

(ii) Zero-Inflated Models (ZIM)

Zero-inflated Poisson and Zero-inflated Negative Binomial are Zero-inflated models capable of addressing issues of excess zero counts and over-dispersion [17]. The Zero-inflated models as compared to the Hurdle models also are two-state models that have a count distribution following negative binomial or Poisson and a point mass at zero. The zero counts may come from both the count component and the point mass, indicating the two sources of zero counts. According to [18], if

$$\omega_i = P(i \in (Structural\quad zero) / Z_i)\ and\ \ 1 - \omega_i$$
$$= P(i \in (Sampling\quad zero) / Z_i),$$

then the Zero-inflated Poisson has the distribution:

$$P(Y_i/X_i,Z_i)=\begin{cases}\omega_i+(1-\omega_i)(\dfrac{\theta}{\mu_i+\theta})^{\theta} & if\ y=0\\ (1-\omega_i)\dfrac{\exp(-\mu_i)\mu_i^{y}}{y!} & for\ y>0\end{cases} \quad (6)$$

The variance and the mean of Y_i according to [18] are given respectively as

$$Var(Y_i/X_i,Z_i)=\mu_i(1-\omega_i)(1+\mu_i\,\omega_i)$$
$$and\ \ E(Y_i/X_i,Z_i)=(1-\omega_i)\mu_i.$$

Generally, the Zero-inflated density function is a combination of the count distribution $f_{count}(y,X_i,\beta)$ the probability of observing $and\ a\ po\mathrm{int}\ mass\ at\ Zero\ I_{(0)}(y).$ a zero count is inflated with a probability: $\prod f_{zero}(0;Z_i,\gamma)$

$$f_{zero}(y;X_i,Z_i,\beta,\gamma)=f_{zero}(0,Z_i,\gamma)*I_{(0)}(y)+(1-f_{zero}(0,Z_i,\gamma)*f_{count}(y;X_i,\beta) \quad (7)$$

Where the unobserved probability belongs to the point mass component and $\mathrm{I}(\bullet)$ is the indicator function. The related mean regression equation is formulated as: by a Canonical Link.

$$\mu_i=\prod_i\bullet 0+(1-\prod_i)\exp(X_i^T\beta) \quad (8)$$

2.3 Methods of Parameter Estimation

In estimating the parameters used in the models, the maximum likelihood estimation (MLE) has been considered. It is therefore very necessary to check the significance of the variables included in the models to evaluate the models involved in the study. The regression coefficients estimated have to be statistically significant for a better model.

Table 2. Dichotomous factors included under counted data models (n=476)

Factors	Categories	Frequency (n)	Percentage
Gender	Male	445	93.5%
	Female	31	6.5%
Have you ever assumed administrative position at the university?	Yes	108	22.7%
	No	368	77.3%
Are you a member of any professional association?	Yes	250	52.5%
	No	226	47.5%
Are you motivated to work at the current University?	Yes	183	38.4%
	No	293	61.6%
If you are not motivated do you have the intention to leave the University?	Yes	347	72.9%
	No	129	27.1%

Table 3. Continuous variables included under counted data models (n=476)

Continuous variables	Information	Minimum	Maximum	Mean	Std. deviation
Dependent	Number of publications of journals articles	0	48	1.16	4.182
Covariates	Age	21	66	29.62	6.545
	Family size	1	12	2.01	1.581
	Average semester work load (credit hour)	3	28	11.29	4.042
	Total experience in teaching (in years)	1	38	5.53	4.913
	Number of staff sharing the same office room with you?	1	15	3.24	2.062

2.4 Statistical Model Selection

2.4.1 Selection of zero-inflated models over traditional models

The Score, Likelihood ratio test, Wald test just to mention few are available for testing the Zero-Inflated in the model [19]. For easiness in selecting Zero-Inflated Models over their traditional counterparts, the Vuong Statistic will be considered. In defining the Vuong Statistics, we assume $f_1(Y_i = y / X_i)$ $and\ f_2(Y_i = y / X_i)$ are both the probability density functions of Hurdle or Zero-Inflated models and their traditional models (Poisson regression model and Negative binomial model) respectively whilst $F_1(Y_i = y / X_i)$ $and\ F_2(Y_i = y / X_i)$ as their corresponding cumulative distribution functions. Then, the Vuong Statistic (V) is therefore defined as:

$$V = \frac{\overline{m}}{S_m / \sqrt{n}} \tag{9}$$

Where $\overline{m} = \frac{1}{n}\sum_{i=1}^{n} m_i$ $and\ S_m = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(m_i - \overline{m})^2}$ represents the mean and the standard deviation of the measurement of m_i. m_i , on the other hand, is defined as:

$$m_i = \log\left[\frac{\hat{f}_1(Y_i = y / X_i)}{\hat{f}_2(Y_i = y / X_i)}\right] \tag{10}$$

Where $\hat{f}_1(Y_i = y / X_i)\ and\ \hat{f}_2(Y_i = y / X_i)$ indicate the predicated probability distribution functions $f_1(Y_i = y / X_i)\ and\ f_2(Y_i = y / X_i)$ respectively.

2.4.2 Akaike's information criterion (AIC)

AIC is a measure of the relative quality of a statistical model for a given data [20]. That is, given a collection of models for data, AIC estimates the quality of each model, relative to other models. Hence, AIC provides a means of model selection. For any Statistical model, the AIC value is computed using the relations:

$$AIC = -2L + 2K \tag{11}$$

Where L is the maximized value of the likelihood function and K is the number of parameters in the model. The model with the lowest AIC value among the models being compared is said to be the best-fitted model. In other words, the better the model fit, the smaller the AIC. AIC is used when comparing non-nested models fitted by maximum likelihood estimation.

3. RESULTS

Bar-chart of instructors' publication data showed that the instructors' publication was characterized by many zero-valued observations and additionally positively skewed to the right direction. Furthermore, the instructors' publication descriptively gave a variance of 17.49 which is much greater than the mean of 1.16 indicating the presence of over-dispersion in the data of Ethiopia public universities.

Even if, the distribution of instructors' publication data reveals that the observed number of zero counts in data were 352 out of 476 (73.9%) respondents indicating the presence of excess zeros. Moreover, it can be affirmed that the instructors' publication contains non-negative integer values. Therefore, analyzed results indicate that the best starting point in analyzing instructors' publication is to use advanced count regression models since the observed counts are non-negative.

The zero-inflated Poisson regression model is capable of dealing with over-dispersion in which the Poisson model cannot handle especially dealing with over-dispersion due to excess zero counts, the model was applied to instructors' publication data at a significant level of α=0.05. The variables in the model which include age, experience, average work-load, association member, motivation, and staff share office were all found to be significant contributing factors to instructors publication with respective to their p-values being less than the level of significance 0.05.

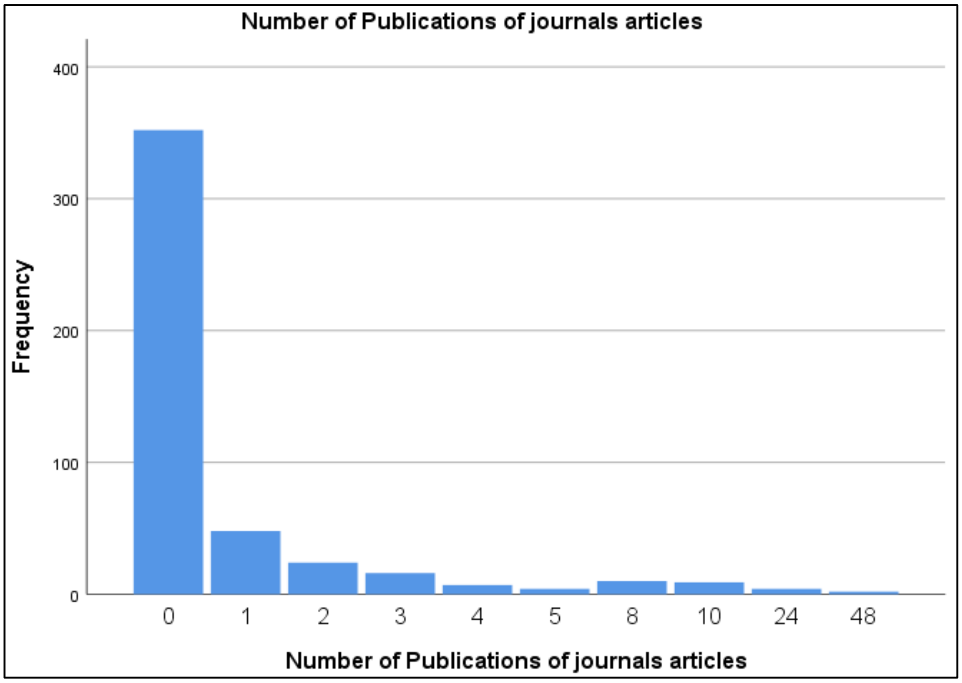

Fig. 1. Bar-chart for distribution of instructors' publication

Table 4. Distribution of instructors' publication data

Number of publication	0	1	2	3	4	5	8	10	24	48	Total
Frequency	352	48	24	16	7	4	10	9	4	2	476.0
Percentage	73.9	10.1	5.0	3.4	1.5	0.8	2.1	1.9	0.8	0.4	100.0
Cumulative percentage	73.9	84.0	89.1	92.4	93.9	94.7	96.8	98.7	99.6	100.0	

A result reveals that gender, family size, experience, average work-load, administrative position, motivation, and staff share office as the significant contributing factors to instructors' publication in Ethiopia public universities. When all the variables in the ZINB regression model were evaluated at zero the model was significant at α=0.05, meaning that, the initial instructors publication equal to intercept term (β_0 =0.494766, P-value=0.0147*). Assessment of the dispersion parameter ϖ showed that there is over-dispersion due to excess zero counts since log(ϖ) is significant at α=0.05.

Finally, the Poisson hurdle regression model which can deal with extra zero counts especially dealing with over-dispersion due to excess zero counts which were fitted at a significant level of α=0.05. Fitting the data to the Poisson Hurdle regression model revealed that the variables: age, gender, family size, experience, average work-load, administrative position, motivation, and staff share office are statistically significant contributing factors to instructors publication in Ethiopia public universities.

Model Evaluation and Comparison

The Vuong test statistic was used to compare the models used in this research since the models are non-nested and was fit the same data. Since many zero-valued models were compared to their traditional counterparts, excess zeros were also tested. AIC was used to compare non-nested models fitted by maximum likelihood to the same data set. The Vuong and AIC test statistics of the various count regression models employed in the study.

To comparing the ZIP and ZINB regression models respectively using the Vuong test statistic, V=9.044266 for ZINB and V=7.834683 for ZIP, shows that ZINB regression models offered a better fit to the instructors' publication data compared to their traditional counterpart regression models with the same data. Evidence of over-dispersion in the instructors' publication data due to excess zero counts as confirmed by the values of the dispersion parameters for ZINB as 6.6212 respectively. Also, AIC values for all the count regression models involved in the study of which the ZINB had the smallest AIC value indicating that ZINB count regression

Table 5. Parameter estimates of the Zero-Inflated Poisson (ZIP) regression model

i) Count model coefficients (pois with log link)

Variables	Estimates	Standard error	Z-Value	Pr(>|Z|)
Intercept	0.809730	0.168216	4.781	1.68×10^{-6}*
Age	0.010753	0.002116	5.083	3.72×10^{-7}*
Gender	0.002321	0.041577	0.056	0.95547
Family size	0.056079	0.066422	0.844	0.39851
Experience	0.165018	0.054615	3.021	0.00252*
Average work-load	0.148342	0.041494	3.575	0.00035*
Association member	0.123997	0.047534	2.609	0.00909*
Administrative position	0.031125	0.051214	0.608	0.54336
Motivation	-0.599701	0.074112	-8.092	5.88×10^{-16}*
Staff share office	-0.263445	0.096013	-2.744	0.00607*

** means significant at α = 0.05*

ii) Zero-inflated model coefficients (binomial with logit link)

Variables	Estimates	Standard error	Z-Value	Pr(>|Z|)
Intercept	.07670	0.74478	2.788	0.0053*
Age	-0.01543	0.01383	-1.116	0.2646
Gender	1.23396	0.29169	-4.230	2.33×10^{-5}*
Family size	.55248	0.37388	9.502	2.31×10^{-16}*
Experience	1.21280	0.27956	-4.338	1.44×10^{-5}*
Average work-load	1.58816	0.30198	-5.259	5.21×10^{-7}*
Association member	0.36588	0.34355	1.065	0.2869
Administrative position	2.28966	0.39204	5.840	1.46×10^{-9}*
Motivation	3.25643	0.83048	-3.921	8.81×10^{-5}*
Staff share office	4.88095	0.84501	-5.776	7.64×10^{-9}*

** means significant at α = 0.05*

Table 6. Parameter estimates of the Zero-inflated Negative Binomial (ZINB) regression model

i) Count model coefficients (negbin with log link)

Variables	Estimates	Standard error	Z-Value	Pr(>|Z|)
Intercept	0.494766	0.202835	2.439	0.0147*
Age	0.012833	0.002744	4.676	2.92×10^{-6}*
Gender	-0.006316	0.050420	-0.125	0.9003
Family size	0.046556	0.081162	0.574	0.5662
Experience	0.215668	0.066069	3.264	0.0011*
Average work-load	0.147140	0.051261	2.870	0.0041*
Association member	0.118588	0.058808	2.017	0.0467*
Administrative position	0.020552	0.062294	0.330	0.7415
Motivation	-0.608800	0.084933	-7.168	7.61×10^{-13}*
Staff share office	-0.171539	0.122214	-1.404	0.1604
$\log(\omega)$	1.854763	0.178263	10.405	2×10^{-16}*

ii) Zero-inflated model coefficients (binomial with logit link)

Variables	Estimates	Standard error	Z-Value	Pr(>|Z|)
Intercept	2.0096369	0.8483111	2.369	0.01784*
Age	-0.000444	0.0200270	-0.022	0.98230
Gender	-1.598473	0.3989885	-4.006	6.17×10^{-5}*
Family size	4.2478853	0.6366233	6.673	2.51×10^{-11}*
Experience	-1.096493	0.3423760	-3.203	0.00136*
Average work-load	-1.841739	0.3584712	-5.138	2.78×10^{-7}*
Association member	0.4215990	0.4310619	0.978	0.32805
Administrative position	2.4995587	0.4213329	5.933	2.98×10^{-9}*
Motivation	-4.656025	1.1365804	-4.097	4.19×10^{-5}*
Staff share office	-6.074075	1.0261714	-5.919	3.24×10^{-9}*

** means significant at α = 0.05*

Table 7. Parameter estimates of the poisson hurdle regression model

i) Count model coefficients (truncated poisson with log link)

Variables	Estimates	Standard error	Z-Value	Pr(>|Z|)
Intercept	0.725658	0.147604	4.916	8.82×10^{-7}*
Age	0.011092	0.002173	5.105	3.32×10^{-10}*
Gender	-0.007831	0.043792	-0.179	0.8581
Family size	0.037822	0.067462	0.561	0.5750
Experience	0.024328	0.054704	0.445	0.6565
Average work-load	0.122532	0.043847	2.795	0.0052*
Association member	0.100041	0.049805	2.009	0.0446*
Administrative position	0.008041	0.054190	0.148	0.8820
Motivation	-0.373023	0.085936	-4.341	1.42×10^{-5}*
Staff share office	-0.162238	0.102320	-1.586	0.1128

** means significant at α = 0.05*

ii) Zero hurdle model coefficients (Binomial with logit link)

Variables	Estimates	Standard error	Z-Value	Pr(>|Z|)
Intercept	-0.588521	0.469452	-1.254	0.20997
Age	0.025407	0.009106	2.790	0.00527*
Gender	0.741820	0.181476	4.088	4.36×10^{-5}*
Family size	-2.358197	0.201381	-11.710	2×10^{-16}*
Experience	1.632692	0.188961	8.640	2×10^{-16}*
Average work-load	1.080031	0.190177	5.679	1.35×10^{-8}*
Association member	0.088776	0.241051	0.368	0.71266
Administrative position	-1.319496	0.236760	-5.573	2.50×10^{-8}*
Motivation	-0.607081	0.239426	-2.536	0.01123*
Staff share office	1.173696	0.243159	4.827	1.39×10^{-6}*

** means significant at α = 0.05*

Table 8. Model comparison and evaluation

Characteristics	ZIP	ZINB	PHURDLE
AIC	4213.256	4013.127	4142.567
Dispersion parameter		6.6212*	
Vuong test	7.83468*	9.044266*	6.890093*

model fits best to the instructors' publication data compared to the other count regression models. Thus, concerning the results from the Vuong test and the AIC, it can be concluded that ZINB as compared to the other regression models employed in the study is the appropriate and efficient model for fitting instructors' publication data to determine key factors that contribute to instructors' publication in Ethiopia public universities. Based on the ZINB regression model, the key factors that contributed significantly to instructors' publication in Ethiopia public universities were age, experience, average work-load, an association member and motivation to work in Ethiopia public universities.

4. DISCUSSION

The response variable, instructors' publication data were characterized by excess zero counts of about 352(73.9%) which is evident by the Vuong test that flavored two-component models as against one component models. The value of the dispersion parameter (6.6212) from ZINB also indicated that there was over-dispersion in the instructors' publication data due to excess zeros in the data. Modelling the instructors' publication data with the various regression models (ZINB, ZIP and Poisson hurdle) showed an agreement with [17] who asserted that ZINB is known to provide robust statistics especially when zero counts are present and in addition to that of ZIP and Poisson Hurdle models. Likewise, there was an agreement with [18] who also found the ZIP model to be better than other models. In this study, based on the ZINB regression model the key factors that contributed significant risk factors for instructors' publication in Ethiopia public universities were age, experience, average work-load, an association member and motivation to work in Ethiopia public universities.

5. CONCLUSION

In this research, an appropriate model suitable for fitting instructors' publication data was determined. Advanced composite count regression models such as ZINB, ZIP and the Poisson Hurdle count regression models gave a more suitable fit to the data with over-dispersion as a result of the high frequency of zero counts. Based on the AIC and Vuong test statistics, the ZINB was selected as the appropriate and significantly efficient model suitable for fitting instructors' publication data which is characterized with over-dispersion and many zero counts. It was finally found that age in year, experience in a year, average work-load in credit hours, an association member and motivation to work were the significant risk factors contributing to instructors' publication in Ethiopia public universities with 352(73.9%) of excess zero counts in the instructors' publication data.

The researcher therefore recommends, that to publish an article it's better to reduce average work-load is a significant root in Ethiopia public universities. While based on experience and by considering an association membership is better to give a chance to follow the next educational program has a positive effect to create young researcher. To end, creating motivation to work in higher education is crucial to publish an article, for instance, health, family, children education, house and so on problems are mandatory to provide them safely.

ACKNOWLEDGEMENT

It's my great pleasure to thank Haramaya University grant for providing me with full financial support for data collection of this study.

COMPETING INTERESTS

Author has declared that no competing interests exist.

REFERENCES

1. Dent D. Innovation and growth-the role of R & D. Dent Associates White Papers 11-02; 2011. Available:http://www.dentassociates.co.uk/wpdivi/wp-content/uploads/2015/09/Innovation-and-Growth-the-role-of-RD.pdf
2. Rosenberg N. Innovation and economic growth; 2004. Available:https://www.oecd.org/cfe/tourism/34267902.pdf
3. Cassiolato JE, Lastres HM, Maciel ML. Systems of innovation and development. UK: Edward Elgar Publishing; 2003.
4. Philips M, Coy P. Look Who's Driving R & D Now: Companies spend to boost productivity, while government cuts back on research. Bloomberg; 2015. Available:http://www.bloomberg.com/news/articles/2015-06-04/look-who-s-driving-r-d-now
5. Mensah IA, Alhassan EA, Affi PO, Baah A, Sarfo BKO. Modelling the occurrence of dental carries in adult population in Ghana; A comparison of competing count regression models. J Biostat Biometric App. 2018;3(2):201.

6. Garomssa HD. The state of entrepreneurialism in a public university in Ethiopia: Status, challenges and opportunities (Master's thesis); 2016.
7. Arnaut D. Towards an entrepreneurial University. International Journal of Euro-Mediterranean Studies. 2010;3(1):135-152. Available:http://www.emuni.si/press/ISSN/1855-3362/3_135-152.pdf
8. Etzkowitz H. Innovation in innovation: The Triple Helix of University-Industry-Government Relations. Social Science Information. 2003;42(3):293-337. DOI: 10.1177/05390184030423002
9. Cowan K. Higher Education's Higher Accountability. American council on education leadership and advocacy; 2013. Available:http://www.acenet.edu/the-presidency/columns-and-features/Pages/Higher-Education's-Higher-Accountability.aspx
10. Gibb A, Haskins G, Hannon P, Robertson I. Leading the Entrepreneurial University: Meeting the entrepreneurial development needs of higher education institutions. University of Oxford; 2012. Available:http://eulp.co.uk/wp-content/uploads/2014/05/EULP-LEADERS-PAPER.pdf
11. Clark BR. Creating entrepreneurial universities: Organizational pathways of transformation. Oxford: IAU PRESS; 1998.
12. Rothaermel F, Agung S, Jiang L. University entrepreneurship: A taxonomy of the literature. Industrial and Corporate Change. 2007;16(4):691-791. DOI: 10.1093/icc/dtm023
13. Lee Y, Gaertner R. Technology transfer from University to industry. A large-scale experiment with technology development and commercialization. Policy Studies Journal. 1994;22(2):384-399. DOI: 10.1111/j.1541-0072.1994.tb01476.x
14. Siegel DS, Waldman DA, Atwater LE, Link AN. Commercial knowledge transfers from universities to firms: Improving the effectiveness of University-industry collaboration. The Journal of High Technology Management Research. 2003; 14(1):111-133. DOI: 10.1016/s1047-8310(03)00007-5
15. De Coster R, Butler C. Assessment of proposals for new technology ventures in the UK: Characteristics of University Spin-off Companies. Technovation. 2005;25(5): 535-543. DOI: 10.1016/j.technovation.2003.10.002
16. Lee SS, Osteryoung JS. A comparison of critical success factors for effective operations of University business incubators in the United States and Korea. Journal of Small Business Management. 2004;42(4):418-426. DOI: 10.1111/j.1540-627x.2004.00120.x
17. Lee AA, Xiang L, Fung WK. Sensitivity of score tests for Zero inflation in count data. Stat Med. 2004;23:2757-69.
18. Mullahy J. Specification and testing of some Modified count data models. J Econometrics. 1986;33:341-65.
19. Lambert D. Zero-inflated poisson regression, with an application to defects in manufacturing. Technometrics. 1992;34: 1-14.
20. Akaike H. Information theory and an extension of the maximum likelihood principle. Proc. 2nd Inter Symposium Information Theory. 1973;267-81.

Peer-review history:
The peer review history for this paper can be accessed here:
https://sdiarticle4.com/review-history/51535