American Journal of Theoretical and Applied Statistics
2019; 8(5): 169-178
http://www.sciencepublishinggroup.com/j/ajtas
doi: 10.11648/j.ajtas.20190805.12
ISSN: 2326-8999 (Print); ISSN: 2326-9006 (Online)

A Comparison of Count Regression Models on Modeling of Instructors Publication Factors: Application of Ethiopian Public Universities

Alebachew Abebe

Department of Statistics, College of Computing & Informatics, Haramaya University, Dire Dawa, Ethiopia

Email address:
aleb.abebe@yahoo.com

To cite this article:
Alebachew Abebe. A Comparison of Count Regression Models on Modeling of Instructors Publication Factors: Application of Ethiopian Public Universities. *American Journal of Theoretical and Applied Statistics*. Vol. 8, No. 5, 2019, pp. 169-178.
doi: 10.11648/j.ajtas.20190805.12

Received: July 10, 2019; **Accepted**: August 5, 2019; **Published**: September 23, 2019

Abstract: Instructors' publication (IP) is one of a major activity in higher education institutes. Currently, IP faced problem both high prevalence and severity in Ethiopian public universities. Publication was affected approximately around 352 (73.9%) instructors have not done publication in Ethiopian public universities even if there is a problem in both developing and developed countries. Since, the outcomes from IP factors are mostly discrete variable; they are often modeled using advanced count regression models. It is therefore, the purpose of this study was to determine 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 Ethiopian. The data were collected between November 2015 through November 2016 from selected thirteen (13) public universities in Ethiopian through both questionnaires and interview. A cross sectional study design was employed using IP data. A simple random sampling technique was applied to the population of Ethiopian public universities to obtain a sample of 13 universities or 476 individual instructors were selected. The average age of the 476 participants were found to be 30 years with 31 (6.5%) being females and 445 (93.5%) being males. The count outcomes obtained were modeled using count regression models which included Poisson, Negative Binomial, Zero-Inflated Negative Binomial (ZINB), Zero-Inflated Poisson (ZIP) and Poisson Hurdle regression models. In order to compare the performance and the efficiency of the listed count regression models with respect to the IP data, the various model selection methods such as the Vuong Statistic (V) and Akaikes Information Criterion (AIC) were used. The ZINB count regression model with reference to the values of the Vuong Statistic and AIC were selected as the most appropriate and efficient count regression model for modeling 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 Ethiopian public universities.

Keywords: IP, ZINB, ZIP, Poisson Hurdle, V

1. Introduction

Innovation has always had a great impact on economic development. However, it is only recently, (particularly after the late 1950's) 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 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.

In this token, the sufficiency of the two long standing objectives of HEIs (teaching and research) have been brought to the spotlight [5]. 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 [6]. At the same time, there is an increasing demand for quality and accountability, among other things [7]. Interestingly enough, such expectations are looming in the face of declining pattern of government funding; all necessitating entrepreneurial response. According to [8] 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 [9].

This increasing expectation from universities has also resulted in an increasing number of publications on academic entrepreneurship [10]. According to, [10], 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 [15] (iii) new firm creation, [16] and (iv) environmental context including networks of innovation [17].

Instructors publication is a major problem in Ethiopian 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 Ethiopian to reduce poverty and to improve its' economic growth. 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 Ethiopian public universities.

2. Methods

To address the main objectives of the study, a cross-sectional study design was employed using well-structured and unstructured questionnaires with interview. The study was conducted between November 2015 to November 2016 at selected thirteen (13) Ethiopian 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. Data Collection

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 or 476 instructors 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.

2.2. Inferential Statistics

Regression Models (RMs)

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

(i). Poisson Regression Model (PRM)

The PRM is the most basic model for count data. If the variance of the counts approximately equals to the mean counts, then the PRM is expressed as:

$$P(Y_i = y / X_i) = \frac{\exp(-\mu_i)\mu_i^y}{y!} \; for\; y = 0,1,2,--- \qquad (1)$$

Where Y_i represents the number of publication for a specific period i and μ_i represents the expected number of publication per given period which can be expressed as:

$$\mu_i = \exp(X_i^T \beta) \qquad (2)$$

Where, β is the vector of unknown regression parameters to be estimated and X_i is the vector of explanatory variables. The equation (2) gives the indication that a unit increase in an explanatory variables increase the expected value μ_i by a multiplicative factor of exp(β) . The main constraint of PRM is that, the mean and the variance are approximately equal that is:

$$E(Y_i = {}^{y}\!/\!{}_{X_i}) = Var(Y_i = {}^{y}\!/\!{}_{X_i}) = \mu \qquad (3)$$

Due to this, in the presence of heterogeneity or over-dispersion (when the variance increase faster than what the PR allows), the PR does not work well hence there is the need to fit a parametric model that is more dispersed than the PM and a natural choice is the Negative Binomial, Poisson Hurdle and the Zero-inflated regression models. The log-likelihood function of the PRM is expressed as:

$$l = \sum_{i=1}^{n} (-\mu_i + y \ln \mu_i - \ln y! \qquad (4)$$

By substituting equation (2) in to equation (4), we further obtain the log-likelihood function as:

$$l = \sum_{i=1}^{n} (-e^{X\beta} + yX_i\beta - \ln y!) \qquad (5)$$

In order to estimate the regression coefficient in the PRM is not obtained from a direct equation but rather the Newton Raphson method used for estimating the unknown parameters in the model.

(ii). Negative Binomial Regression Model (NBRM)

If a PRM doesn't fit the data and it appears that variance of is increasing faster than the PR allows, then a simple scale-factor adjustment is not appropriate. One way to handle this situation is to a fit a parametric PR that is more dispersed than the Poisson. A natural choice is the negative binomial.

Suppose; $y/\lambda \sim Poisson(\mu)\, and\ \lambda \sim Gamma(\alpha, \beta)$ is the gamma distribution with mean $\alpha\beta$ and $\alpha\beta^2$ variance whose density is given by:

$$P(\lambda) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)}\lambda^{\alpha-1}\exp(-\lambda/\beta) \quad (6)$$

For $\lambda > 0$ and zero otherwise. Then it is easy to show that the unconditional distribution of y is negative binomial:

$$P(Y_i = y/X_i) = \frac{\Gamma(\alpha+y)}{\Gamma(\alpha)y!}\left(\frac{\beta}{1+\beta}\right)^{y_i}\left(\frac{1}{1+\beta}\right)^{\alpha} \quad for\ y = 0,1,2,--- \quad (7)$$

This distribution has the mean and variance as; $E(y) = \alpha\beta$ and $Var(y) = \alpha\beta + \alpha\beta^2$ respectively.

In order to build the RM, it is natural to express the negative binomial in terms of the parameters $\mu = \alpha\beta\, and\ \omega = 1/\alpha$ so that $E(y) = \mu\ and\ Var(y) = \mu + \omega\mu^2$ where the variances function is quadratic. The distribution of that is formulated as:

$$P(Y_i = y/X_i) = \frac{\Gamma(\omega^{-1}+y)}{\Gamma(\omega^{-1})y!}\left(\frac{\omega\mu}{1+\omega\mu}\right)^{y_i}\left(\frac{1}{1+\omega\mu}\right)^{\omega^{-1}} \quad (8)$$

Which approaches Poisson $(\mu)\ as\ \omega \to 0$. The negative binomial can accommodate over-dispersion but not under-dispersion with respect to the PM. For regression purposes, it is typically assumed; $y_i\ Negbin(\mu_i, \omega)$ and applies a log-link, so that:

$$\log\mu_i = \eta = X_i^T\beta \quad (9)$$

The log-likelihood function of the negative binomial regression model is obtained from the following equation:

$$l = \sum_{i=1}^{n}\left\{\left[\sum_{i=1}^{y-1}\ln(y_i+\alpha^{-1}) - \ln y_i! + \alpha^{-1}[\ln(\alpha^{-1}) - \ln(\alpha^{-1}+e^{X\beta})] + y_i[X_i\beta - \ln(\alpha^{-1}+e^{X\beta})]\right]\right\} \quad (10)$$

In order to estimate $\beta\ and\ \alpha$ as in the PRM, the iteration procedure or the method of Newton Raphson is applied [18].

(iii). Poisson Hurdle Regression Model (PHRM)

Many count data exhibit more zero counts and are in addition over-dispersed. One type of count regression model that is capable of dealing with both excess zeros and over-dispersion is the PHRM, which was proposed [11]. 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:

$$P(Y_i/X_i, Z_i) = \frac{(1-\omega_i)\exp(-\mu_i)\mu_i^{\,y}}{(1-\exp(-\mu_i))y!} \quad for\ y > 0 \quad (11)$$

Where $\mu_i = \exp(X_i^T\beta)$. The variance and mean of a PHM according to [11] are given as:

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

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:

$$f_{hurdle}(y, X_i, Z_i, \beta, \gamma) = \begin{cases} f_{zero}(0, Z_i, \gamma) & if\ y = 0 \\ 1 - f_{zero}(0, Z_i, \gamma).f_{count}(y, X_i, \beta)/f_{count}(0, X_i, \beta) & if\ y > 0 \end{cases} \quad (12)$$

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:

$$L = \prod_{i=U_0}\{f_{zero}(0; Z_i, \gamma)\}\prod_{i=U_1}\{(1-f_{zero}(0; Z_i, \gamma)) * f_{count}(y, X_i, \beta)/1 - f_{count}(0; X_i, \beta)\} \quad (13)$$

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 log of the likelihood function and rearranging the terms, gives the following relation:

$$l = \prod_{i=U_0} \ln\{f_{zero}(0;Z_i,\gamma)\} + \sum_{i=U_1}\{1-f_{zero}(0;Z_i,\gamma)\} + \sum_{i=U_1}\ln\{f_{count}(y,X_i,\beta)\} - \ln\{1-f_{count}(0,X_i,\beta)\} \tag{14}$$

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 β *and* γ. Hence, the mean regression relationship can be expressed as: by a Canonical Link.

$$\log(\mu) = X_i\beta + \log(f_{zero}(0;Z_i,\gamma)) - \log(1-f_{count}(0,X_i,\beta)) \tag{15}$$

(iv). 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 [12]. 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 $\omega_i = P(i \in (Structural\ zero)/Z_i)$ and $1-\omega_i = P(i \in (Sampling\ zero)/Z_i)$, then the Zero-inflated Poisson has the distribution:

$$P(Y_i/X_i,Z_i) = \begin{cases} \omega_i + (1-\omega_i)(\dfrac{\theta}{\mu_i+\theta})^{\theta} & if\ y=0 \\ (1-\omega_i)\dfrac{\exp(-\mu_i)\mu_i^{y}}{y!} & for\ y>0 \end{cases} \tag{16}$$

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 $f_{count}(y,X_i,\beta)$ *and a point mass at Zero* $I_{(0)}(y)$.

The probability of observing a zero count is inflated with a probability: $\prod f_{zero}(0;Z_i,\gamma)$

$$f_{zero}(y,X_i,Z_i,\beta,\gamma) = f_{zero}(0;Z_i,\gamma) * I_{(0)}(y) + (1-f_{zero}(0;Z_i,\gamma) * f_{count}(y,X_i,\beta) \tag{17}$$

Where the unobserved probability belong to the point mass component and I(•) is the indicator function. The related mean regression equation is formulated as: by a Canconical Link.

$$\mu_i = \prod_i \bullet 0 + (1-\prod_i)\exp(X_i^T\beta) \tag{18}$$

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 in order to evaluate the models involved in the study. The regression coefficients estimated have to be statistically significant for a better model.

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 [18]. 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:

$$V = \frac{\bar{m}}{S_m/\sqrt{n}} \tag{19}$$

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

$$m_i = \log\left[\frac{\hat{f}_1(Y_i=y/X_i)}{\hat{f}_2(Y_i=y/X_i)}\right] \tag{20}$$

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 [14]. That is, given a collection of models for a 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:

$$AIC = -2L + 2K \tag{21}$$

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.

2.5. Statistical Data Analysis

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 to instructors publication in Ethiopia public universities. Out of the total of 476 instructors':- Part-I: in this study 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%)], and 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%)] were utilized for conducted descriptive statistics. Part-II: information on instructors' publication age (in year), family size, service (in year), average work-load, gender (male, female), administrative position (yes, no) and professional association (yes, no) were recorded using a structured questionnaire. Finally, 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. Both variables were used for comparison of counted regression models such as Poisson Regression Model, Negative Binomial Regression Model, Poisson Hurdle Regression Model and Zero-inflated Models were used to fit the best robust model by using maximum likelihood estimation techniques.

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 Ethiopian public universities.

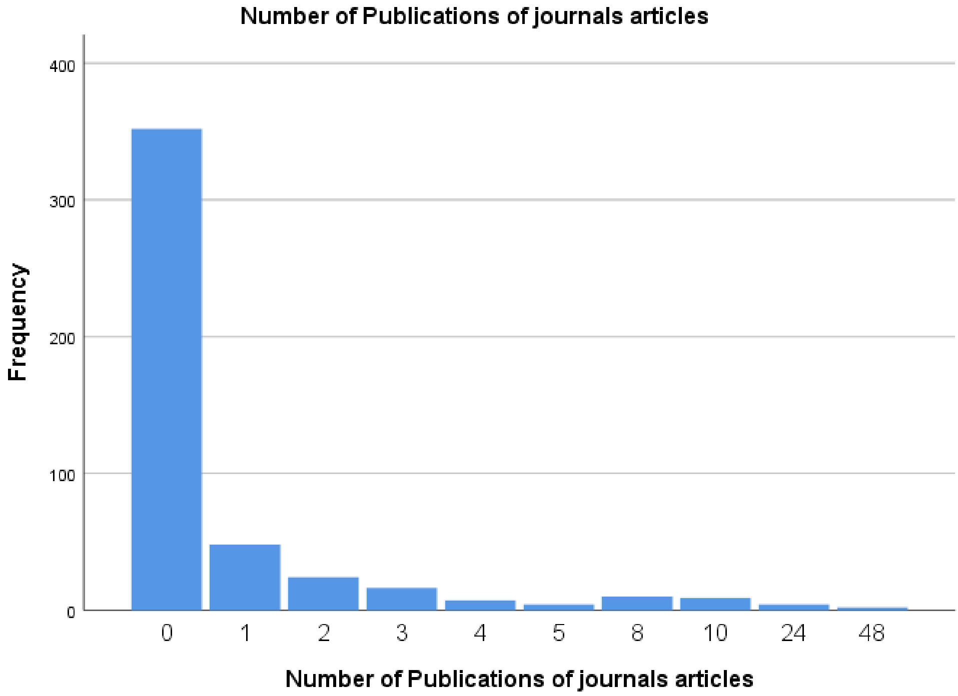

Figure 1. *Bar-chart indicating the distribution of instructors' publication.*

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 clearly indicates that the best starting point in analyzing instructors' publication is to use the Poisson regression model. Since, Poission regression model the data were observed counts are non-negative and it has some extensions that are useful for count data. The outcome variable instructors' publication occurred within a given period of time and did not assume normality hence the use of Poisson count regression model.

Table 1. *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
Percent	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	

According to PRM results revels that age, gender, family size, experience, average work-load, administrative position, association member, motivation, and staff share office with the exception of the intercept were all found to be significant as contributing factors to instructors' publication since their respective p-values are less than the level of significance $\alpha = 0.05$.

Table 2. *Parameter estimates of the Poisson Regression Model for Publication.*

Variables	Estimates	Standard error	Z-Value	Pr(>|Z|)
Intercept	-0.051129	0.114310	-0.334	0.6789
Age	0.0021377	0.001653	4.123	1.3×10^{-11}*
Gender	0.062011	0.026634	3.702	0.0313*
Family size	-0.491620	0.051455	-9.213	1×10^{-13}*
Experience	0.681273	0.030234	5.820	3.8×10^{-14}*
Average work-load	-0.263086	0.020071	-9.190	4.27×10^{-9}*
Administrative position	-0.190912	0.074967	-4.192	4.62×10^{-6}*
Association member	0.561134	0.013754	3.460	0.0141*
Motivation	0.354906	0.036412	1.579	2.54×10^{-7}*
Staff share office	-0.222932	0.099373	-4.305	1.87×10^{-8}*

* means significant at $\alpha = 0.05$

The Negative binomial regression model was estimated using instructors' publication data at a significant level of α=0.05 to account for over-dispersion due to heterogeneity. Based on NBRM results the variables age, family size, experience, average work-load, association member, motivation, and staff share office were found to contribute statistically significant to instructors' publication when their respective p-values were compared to the level of significance α=0.05. The dispersion parameter was obtained as 2.750 which is greater than 1 indicates that the presence of over-dispersion in the instructors publication data. Hence, the PRM will be insufficient in modeling the instructors publication data. The 73.9% of zero counts in the instructors publication data (Table 1) implies the use of Zero-inflated models to account for the excess zeros.

Table 3. *Parameter estimates of the Negative Binomial Regression Model for Publication.*

Variables	Estimates	Standard error	Z-Value	Pr(>|Z|)
Intercept	-0.187671	0.166123	-1.145	0.324350
Age	0.014521	0.001567	4.568	1.62×10^{-9}*
Gender	0.087912	0.033240	1.827	0.096574
Family size	-0.594401	0.081706	-9.123	2×10^{-12}*
Experience	0.246874	0.027807	4.765	1.68×10^{-8}*
Average work-load	-0.332256	0.099305	-6.082	4.54×10^{-4}*
Association member	0.100537	0.005451	1.722	0.000376*
Administrative position	-0.335417	0.089748	-6.288	0.135672
Motivation	0.426995	0.074536	2.671	7.07×10^{-5}*
Staff share office	-0.255019	0.097987	-7.536	0.000209*
ω	2.750			

* means significant at $\alpha = 0.05$

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. Also, worth noting is that, when all the variables in this model were evaluated at zero, the model is statistical significant at α=0.05.

The Zero-inflated negative binomial regression (ZINB) model also noted to be capable of dealing with extra zero counts especially when dealing with over-dispersion due to excess zero counts and heterogeneity, that is assumed to come from both the structural and chance sources, was estimated using instructors' publication data at significant level of 0.05.

Table 4. *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 $\alpha = 0.05$

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

Variables	Estimates	Standard error	Z-Value	Pr(>|Z|)
Intercept	2.07670	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	3.55248	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 $\alpha = 0.05$

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 Ethiopian public universities. When all the variables in the ZINB regression model were evaluated at zero, the model was significant at α=0.05. 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.

Table 5. *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(ϖ)	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 $\alpha = 0.05$

Finally, the Poisson hurdle regression model which has the capability to deal with extra zero counts especially dealing with over-dispersion due to excess zero counts which was fitted at 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 Ethiopian public universities. Also worth noting, the estimated Poisson Hurdle model was statistically significant at α=0.05 when all the variables were evaluated at zero.

Table 6. *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

Model Evaluation and Comparison

The Vuong test statistic was used to compare the models used in this research since the models are non-nested models and were fit to the same data. Since, many zero valued models were compared to their traditional counter parts, 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.

By comparing the Poisson regression model and the NB regression model to the ZIP and ZINB regression models respectively using the Vuong test statistic, V=9.044266 for ZINB versus NB and V=7.834683 for ZIP versus Poisson, shows that both ZINB followed by the ZIP regression models offered a better fit to the instructors' publication data compared to their traditional counterpart regression models with one-component 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 NB and ZINB as 2.750 and 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 model fits best to the instructors' publication data compared to the other count regression models. Thus, with respect to 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 in order to determine key factors that contributes to instructors' publication in Ethiopian 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, association member and motivation to work in Ethiopian public universities.

Table 7. *Model Comparison and Evaluation.*

Characteristic	Poisson	NB	ZIP	ZINB	PHURDLE
AIC	4500.401	4407.738	4213.256	4013.127	4142.567
Dispersion parameter		2.750		6.6212*	
Vuong test			7.834683*	9.044266*	6.890093*

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. Modeling the instructors' publication data with the various regression models (Poisson, Negative Binomial, ZINB, ZIP and Poisson hurdle) showed an agreement with [15] who asserted that the ZIP and ZINB are known to provide robust statistics especially when zero counts are present and in addition to Poisson Hurdle models. Likewise, there was an agreement with [16] who also found the ZIP model to be better than the NB model, and that the NB model was also better than the Poisson model. 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, association member and motivation to work in Ethiopian public universities.

5. Conclusion

In this research, an appropriate model suitable for fitting instructors' publication data were determined. The Poisson count regression model for count data modeling was a good starting point but has a cross-sectional equidispersion assumption. Due to this, the Negative binomial count regression model with more relaxed assumption on variance provided a better solution with evidence of over-dispersion in the instructors' publication data since the variance was greater than the mean.

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 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 year, average work-load in credit hours, association member and motivation to work were the significant risk factors contributing to instructors' publication in Ethiopian 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 memberships are better to give a chance to follow next educational program has 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 in safe manner.

Acknowledgements

It's my great pleasure to thank Haramaya University grant for providing me full financial support for data collection of this study. Likewise, Haramaya University was covered all publication fee for this journal.

References

[1] Dent, D. (2011). *Innovation and Growth-the role of R & D. Dent Associates White Papers 11-02. Retrieved from http://www.dentassociates.co.uk/wpdivi/wp-content/uploads/2015/09/Innovation-and-Growth-the-role-of-RD.pdf*

[2] Rosenberg, N. (2004). *Innovation and economic growth. Retrieved from https://www.oecd.org/cfe/tourism/34267902.pdf*

[3] Cassiolato, J. E., Lastres, H. M., & Maciel, M. L. (2003). *Systems of innovation and development. UK: Edward Elgar Publishing.*

[4] Philips, M., Coy, P. (2015). *Look Who's Driving R & D Now: Companies spend to boost productivity, while government cuts back on research. Bloomberg. Retrieved from http://www.bloomberg.com/news/articles/2015-06-04/look-who-s-driving-r-d-now*

[5] Arnaut, D. (2010). *Towards an Entrepreneurial University. International Journal of Euro- Mediterranean Studies, 3 (1), 135-152. Retrieved from http://www.emuni.si/press/ISSN/1855-3362/3_135-152.pdf*

[6] Etzkowitz, H. (2003). *Innovation in Innovation: The Triple Helix of University-Industry-Government Relations. Social Science Information, 42 (3), 293-337. doi: 10.1177/05390184030423002.*

[7] Cowan, K. (2013). *Higher Education's Higher Accountability. American council on education leadership and advocacy. Retrieved from http://www.acenet.edu/the-presidency/columns-and-features/Pages/Higher-Education's-Higher-Accountability.aspx*

[8] Gibb, A., Haskins, G., Hannon, P., & Robertson, I. (2012). *Leading the Entrepreneurial University: Meeting the entrepreneurial development needs of higher education institutions. University of Oxford. Retrieved from http://eulp.co.uk/wp-content/uploads/2014/05/EULP-LEADERS-PAPER.pdf*

[9] Clark, B. R. (1998). *Creating entrepreneurial universities: Organizational pathways of transformation. Oxford: IAU PRESS.*

[10] Rothaermel, F., Agung S., & Jiang, L. (2007). *University entrepreneurship: a taxonomy of the literature. Industrial and Corporate Change, 16, (4), 691-791 doi: 10.1093/icc/dtm023.*

[11] Mullahy J. (1986). *Specification and Testing of Some Modified Count Data Models. J Econometrics 33: 341-65.*

[12] Lambert D. (1992). *Zero-inflated Poisson Regression, with an Application to Defects in Manufacturing. Technometrics 34: 1-14.*

[13] Lee, Y., & Gaertner, R. (1994). *Technology Transfer from University to Industry. A Large-Scale Experiment with Technology Development and Commercialization. Policy Studies Journal, 22 (2), 384-399. doi: 10.1111/j.1541-0072.1994.tb01476.x.*

[14] Akaike H. (1973). *Information theory and an extension of the maximum likelihood principle. Proc. 2nd Inter Symposium Information Theory 1973: 267-81.*

[15] Siegel, D. S., Waldman, D. A., Atwater, L. E., & Link, A. N. (2003). *Commercial knowledge transfers from universities to firms: improving the effectiveness of university-industry collaboration. The Journal of High Technology Management Research, 14 (1), 111-133. doi: 10.1016/s1047-8310(03)00007-5.*

[16] De Coster, R., & Butler, C. (2005). *Assessment of proposals for new technology ventures in the UK: characteristics of university spin-off companies. Technovation, 25 (5), 535-543. doi: 10.1016/j.technovation.2003.10.002.*

[17] Lee, S. S., & Osteryoung, J. S. (2004). *A Comparison of Critical Success Factors for Effective Operations of University Business Incubators in the United States and Korea. Journal of Small Business Management, 42 (4), 418-426. doi: 10.1111/j.1540-627x.2004.00120.x.*

[18] Lee AA, Xiang L, Fung WK (2004). *Sensitivity of score tests for Zero inflation in count data. Stat Med. 23: 2757-69.*