IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 18, Issue 3 Ser. I (May. – June. 2022), PP 04-13
www.iosrjournals.org

Bayesian Zero/One-Inflated Beta Regression Modeling of Household Expenditure on Education in Marginalized Communities in Kenya

Immaculate Atieno[1], Samuel Mwalili[1], Oscar Ngesa[2],
[1]Department of Statistics and Actuarial Science, Jomo Kenyatta University of Agriculture and Technology, Nairobi 6200-00200, Kenya [2]Department of Mathematics and Physical Sciences, TaitaTaveta University, Voi 635-80300, Kenya

Abstract: *Modeling of bounded outcomes within given intervals is continuously gaining wide applications in various fields as with engineering, economics and health sciences. Major applications have been on the usage of the Beta distribution first introduced by Ospina& Ferrari who used the beta law to define the continuous component of the distribution. This study gives an application of an extension of the Beta distribution incorporating a Bayesian Inference to model the proportion of household expenditure on education in the arid and semi-arid area of Kenya, a case of Isiolo County. The Bayesian Zero/One Inflated Beta Regression model is used in which the household head levels of education, age of household head, household size, household asset value and the distance to an educational facility by a household were evaluated. Data used for the study was secondary data obtained from the Food Agricultural Organization, Kenya for the period 2018-2019. Fitted model parameter estimates were obtained via the Maximum Likelihood Estimation and the Bayesian Inference used for predictive analysis to calculate posterior probabilities for individual covariates. All model covariates were found to be significant in which Distance to an educational facility and the Gender of household head (male) having a negative effect in the estimation and prediction of proportion of household expenditure spent on education. Model diagnostic checks were via the Deviance Residuals, Probability plots and the Chi-Square statistic.*

Date of Submission: 28-04-2022 Date of Acceptance: 10-05-2022

I. Introduction

In regression analysis there is often a problem of modeling outcomes that are proportional in nature and within the interval of zero and one. At times, these outcomes constitute values in the two extremes that is too many observation could be zeroes or ones with few values in between. The Beta regression is applied for (0, 1) outcomes, but again has to be modified to cater for the heavy tailed data at extremes of the distribution [1] which is a concern for this study. This study is concerned with such general class of regression model for modeling continuous proportions when there exists zeros/one inflated outcomes in which we explore the Bayesian Zero/One Inflated Beta Regression model in the modeling of the proportion of household expenditure spent on education.

Here, the household head levels of education, age of household head, household size, household asset value and the distance to an educational facility by a household were evaluated as factors that would influence the proportion of household expenditure spent on education.

Statistical modeling of continuous proportions has received close attention in the last few years as with Xioling et al [7] who used a Zero-Inflated Beta Regression model to identify differentially abundant features between multiple clinical conditions for human metagenomic data, Jonah et al [9], who used the Inflated Beta distribution to estimate plant abundance and Luiz et al [12], who had a process monitoring using Inflated Beta Regression Control Chart for bounded response. They stipulated the goodness of fit of the Beta distribution and its parameterizations in the modeling of bounded continuous (proportional) data. It's in this regard that this study models the proportion of household expenditure on education using the Zero/One Inflated Beta Regression Model.

There is need for improved strategies aimed at increasing the uptake of formal education in the arid and semi-arid areas of Kenya. This is only achievable when there is an efficient analysis of the factors that significantly influence household expenditure on education per household in the arid and semi-arid areas. This study evaluates the levels of education levels of the household head and his/her age,

the household size and asset value and the Distance of a household from a given educational facility. The Bayesian inference is used to determine the individual covariate posterior probability of increasing (decreasing) the proportion of household expenditure on education.

This research study was limited to one county within the Arid and semi-Arid Lands (ASAL) which is Isiolo. This is due to the need of high accountability for improving investments in education in the ASAL areas, how well the government policies on education are impacting the communities on the ground and the lessons learnt for future interventions. The study was also limited on the usage of secondary data collected in the previous years in Isiolo County from the Food and Agricultural Organization-Kenya. The subsequent sections of this study are organized as follows: section two for the literature review, section three for the study methodology, section four for the data analysis and section five for the conclusions & recommendations.

II. Literature Review

In the Generalized Linear Regression Modeling, Achia et al [6] modeled key identifies of household poverty using the Logistic Regression Model as a factor influencing household expenditure. Tony [10] modeled household expenditure using a random assignment scheme incorporating the multiple linear regression model. Manos et al [11] modeled household expenditure on health-care in which they compared estimators for two-part models and the generalized linear models. Buigut et al [13] used the multivariate logistic regression to examine incidence and determinants of catastrophic household expenditure on health-care. Bahadur et al [5] model household expenditure on energy consumption using the Multiple Linear Regression model. Partha et al [14] described estimation and interpretation of the effects of a natural experiment using two classes of non-linear statistical models for household expenditure and use on healthcare.

For the regression models with bounded outcomes, Xiaoling et al [7] proposed a Zero-Inflated Beta Regression model for identifying differentially abundant features between multiple clinical conditions for human metagenomic data. Jonah et al [9] used the Inflated Beta Distributions to estimate plant abundance. Sohom et al [15] developed a new multisite statistical downscaling method based on beta regression for generating synthetic precipitation series, which can preserve temporal and spatial dependence of the data.

Debora et al [16] had a Multivariate Beta Regression with applications in small area estimation process between units especially small samples. Danielle et al [4] used the Zero-Inflated Beta Regression model to evaluate the incidence of Citrus Canker disease in leaves of sweet oranges. Luiz et al [12] had a process monitoring using Inflated Beta Regression Control chart for a bounded response variable.

In the Bayesian Analysis, Omid &Mousa [19] had a Bayesian analysis of regression models using Instrumental Variables to model household income and expenditure based on Markov Chain Monte Carlo (MCMC) simulations. Pudji et al [17] had a hierarchical Bayesian framework for modeling the effect of regional diversity on household expenditure and Brown et al [22] modeled household finances with a Bayesian approach to a multivariate two-part model. In determining the factors influencing household expenditure on education, Mwangi & Kariuki [18] studied household poverty determinants in Kenya. Prahutama et al [8] analyzed factors that influence household expenditure on education using the Multiple Linear Regression Model. Danuta et al [3] examined the nature of household financial decision making with regard to sourcing external finance for household expenditure. Eyre et al [23] had a Bayesian belief network modeling of household food security. Donkoh et al [20] had a study on the determinants of household education expenditure in Ghana. Habanabakize [21] determined the household consumption expenditure resilience towards petrol price, disposable income and exchange rate volatilities.

III. Methodology

Data used in the study was secondary data obtained from Food Agricultural Organization (FAO) for the period 2018-2019. The data attributes included the household head levels of education, age of household head, household size, household asset value and the distance to an educational facility by a household.

Modeling Household Expenditure

In the modeling of Household expenditure on education in the arid and semi-arid areas of Kenya, the Zero/One Inflated Beta Regression model is fit to the data in comparison to the Classic Beta Regression Model.

Beta Distribution

The Beta distribution characterized by two shape parameters α_1 and α_2 was first fit to the data with mean and variance given respectively as $\frac{\alpha_1}{\alpha_1+\alpha_2} = \mu$ and $\frac{\mu(1-\mu)}{1+\phi}$ for $\phi = \alpha_1 + \alpha_2$ [1]. It is parameterized in terms of the mean (μ) and precision parameter (ϕ) as;

$$f(y;\mu,\phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\,\Gamma((1-\mu)\phi)}\, y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1} \quad (1)$$

for $0 < y < 1$, $0 < \mu < 1$ and $\phi > 0$.

For $y_1, y_1, y_1, \cdots y_n$ independent proportions of household expenditure on education for different households, we let $y_t, t = 1,2,\ldots,n$ be household proportion expenditures that follow a Beta distribution with mean μ_t and unknown precision parameter ϕi.e $y \sim Beta\,(\mu_t, \phi)$ and $g(\cdot)$ to be strictly monotone and twice differentiable link function that maps $(0,1)$ to$\mathbb{R}$. The Beta Regression is thus given in terms of a logistic link function as;

$$g(\mu_t) = log\left(\frac{\mu_t}{1-\mu_t}\right) = \sum_{i=1}^{n} X_{ti}\beta_i = \eta_t \qquad (2)$$

Where $\beta = (\beta_1, \beta_2, \ldots, \beta_k)^T \in$ is a vector of unknown regression parameters and $X_{t1}, X_{t2}, \ldots, X_{tn}$ are observations on k covariates $(k < n)$.

Zero/One Inflated Beta Distribution

This model is thought as a mixture model of Beta and Binomial distributions. Here, we suppose y_j as the j^{th} proportion of household income spent on education out of a total of p proportions measured on n independent units, that is, $y_j = (y_{1j}, y_{2j}, \ldots, y_{nj})^T$. In this case, β_{mj} are the linear fixed effects in the link functions $m(m = 1,2,3,4)$ for $j(j = 1,2,\ldots,p)$ responses. $X_{m,ij}$is a design matrix for the model fixed effects and $I_m(Z_{m,ij}, \gamma_{m,i})$ is an indicator function for the presence of random component in the m link functions. $I_m(Z_{m,ij}, \gamma_{m,i}) = Z_{m,ij}, \gamma_{m,i}$if there is a random component in the model link functions and $I_m(Z_{m,ij}, \gamma_{m,i}) = 0$ otherwise. $Z_{m,ij}$is a matrix for the random components of the model link functions. $\gamma_{m,i}$is assumed to be independently and normally distributed with mean Zero and Variance Σ.

Parameter Estimation & Bayesian Inference

In parameter estimation, the Maximum Likelihood Estimation Technique was used. The log-likelihood from equation (1) was given as;

$$l(\beta,\phi) = \sum_{t=1}^{n} \log\Gamma(\phi) - \log\Gamma\big((1-\mu_t)\phi\big) + (\mu_t\phi - 1)log\, y_t + [(1-\mu_t)\phi - 1]\log(1-y_t) \qquad (3)$$

Defining $\Theta = \{\beta_1, \beta_2, \beta_3, \beta_4, \Sigma\}$ as a set of p parameters from the Zero/One Inflated model, the likelihood $p(y|\Theta, \gamma)$ is constructed as;

$$P(y|\Theta,\gamma) \propto \prod_i \prod_j \left(p_{ij}^{I(y_{ij}=0)}(1-p_{ij})^{I(y_{ij}>0)} q_{ij}^{I(y_{ij}=1)}\right) \times \Bigg((1-q_{ij}) \frac{\Gamma(\phi_{ij})}{\Gamma(\phi_{ij}\mu_{ij}\Gamma(\phi_{ij}(1-\mu_{ij})))} (y_{ij})^{\phi_{ij}\mu_{ij}-1}(1-$$

yijϕij1−μij−1Iyij∈0,1 (4)

ϕ_{ij}, μ_{ij}, p_{ij}andq_{ij} are all functions of $\Theta$and$p(y|\Theta,\gamma) \sim N(0,\Sigma)$. All parameters are assumed to be priori independent, thus $f(\Theta) = f(\Sigma)\prod_{j=1}^{p}\beta_{m,j}$.

3.4. Model Diagnostics

In order to analyze the goodness of fit of the fitted model, the Deviance residuals, Probability plots and the Chi-Square test of association shall be used.

IV. Results and Discussions

Preliminary Data Analysis

In the preliminary data analysis of the proportional household expenditure on education, the descriptive and fitted model coefficients are discussed.

Descriptive Statistics

Prior to fitting the Inflated Beta Regression model to the data, descriptive statistics of the data were obtained as given in Table 1.

Table 1. *Household Expenditure on Education Descriptive Statistics*

	Min	**1st Q**	**Med**	**3rd Q**	**Max**
Proportion	0.0001	0.0041	0.0301	0.0866	0.8999
HHH. Education	0.00	0.00	0.00	0.5216	3.00
HH. Asset	-47.75	34.03	84.75	134.95	1831.45

Distance	1.00	1.00	1.00	1.612	10.00
HH. Size	1.00	4.00	6.00	6.00	15.00
HHH. Age	18.00	37.00	48.00	48.76	101.00

The mean household proportion expenditure on education was estimated at 0.0866 with the minimum, 1st Quartile, Median, 3rd Quartile and maximum proportional expenditure given respectively as 0.0001, 0.0041, 0.0301, 0.1210 and 0.8999. As for the individual covariate effect on the proportion of household expenditure on education; the household head levels of education (HHH Edu), the household asset value (HH Asset), distance of a household to an educational center (Distance), household size (HH Size) and the age of household head (HHH Age) had a respective 0.5216, 134.95, 1.612, 6.00 and 48.00 mean effect on the proportion of household expenditure spent on education. The mean proportional expenditure was higher than the median which gave an indication of majority proportional expenditures being to the left of the mean value as given in Figure 1.

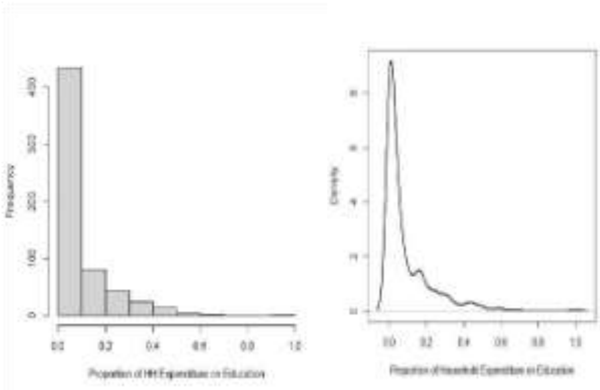

Figure 1. *Histogram and Density of Household Expenditure on Education*

The household levels of education were categorized into None coded as 0, Primary coded as 1, Secondary coded as 2 and Post-Secondary coded as 3 with the minimum level being 0 and maximum 3. The household asset value was computed in terms of thousands and it comprised of all assets in a given household that monetary value could be attached to with the minimum asset value being -47.75 and maximum 1831.45. The distance of a given household to an education facility was measured in Kilometers with minimum and maximum as 1.00 and 10.00 respectively while the household size represented the total number of persons in a given household with a minimum of 1 and maximum 15. The age of the household head was evaluated as the number of years lived up to his/last birthday prior to the survey with a minimum of 18.00 and a maximum of 101.00.

Fitted Model Coefficients

In order to obtain the fitted model coefficients, the Bayesian Inflated Beta Regression model was fit to the data in which the Beta regression coefficients and the Inflated Beta model coefficients were obtained. Table 2 gave a summary of the Beta Regression model coefficients for the covariate, mean and precision models.

Table 2. *Beta Regression Model Coefficients*

	Estimate	**S.Error**	**Z-Value**	**P(>|z|)**
M. Intercept	-3.3423	0.2005	-16.6662	2.3062e-62
P. Intercept	4.8522	0.03299	14.7071	5.7522e-49
Distance	-0.1385	0.0350	-3.9602	7.4898e-05
HH. Asset	0.0004	0.0002	1.8343	6.6616e-02
HHH. Age	0.0089	0.0028	3.1195	1.8117e-03
HHH. Edu	0.1137	0.0518	2.1957	2.8110e-02
HHG. Male	-0.4390	0.0960	-4.5712	4.8494e-06
HH. Size	0.1695	0.0198	8.5793	9.5479e-18

The mean and precision intercepts had a respective -3.3423 and 4.8522 unit decrease/increase for the proportion of household expenditure on education per household for a unit increase in the mean and precision models. A unit increase in distance of a household to an educational facility and the gender of the household head (Male) had a respective -0.1385 and -0.4390 unit decrease effect on the proportion of household expenditure spent on education. A unit increase in the household asset value, age of household head, household head levels of education and the household size had a respective 0.0004, 0.0089, 0.1137 and 0.1695 unit increase effect on the proportion of household expenditure spent on education.

All the covariates had a significant effect on the proportion of household expenditure spent on education with the household size being the most significant. This gave an indication that an increase in the distance of household to an education facility and increase in the number of male household heads led to the decrease in proportion of household expenditure spent on education. Contrary, an increase the household asset value, age of household head, household head levels of education and the household size led to an increase in the proportion of household expenditure spent on education.

Table 3 gave a summary of the Inflated model coefficients for the mean effect (Mean), standard deviation (SD), naive standard deviation (Naive SE) and the time series standard deviation (T-S SE).

Table 3. *Inflated Model Coefficients*

	Estimate	S.D	Naive SE	T-S SE
M.Intercept	-3.3285	0.1915	7.951e-03	6.938e-03
P. Intercept	1.5664	0.0646	2.684e-03	2.686e-03
Distance	-0.1410	0.0338	1.402e-03	1.264e-03
HH. Asset	0.0004	0.0002	9.297e-06	8.863e-03
HHH. Age	0.0086	0.0028	1.182e-04	1.183e-03
HHH. Edu	0.1086	0.0529	2.196e-03	2.197e-03
HHG. Male	0.0984	4.085e-03	-4.5712	4.088e-03
HH. Size	0.1709	0.0197	8.185e-04	8.184e-04

The mean model and the distance of a household to an education facility had a -3.3285 and -0.1410 unit decrease on the proportion of household expenditure on education for a unit increase in each of them. Similarly, a unit increase in the precision model, household asset value, age of household head, education levels of household head, gender of the household head (male) and the household size had a respective 1.5664, 0.0004, 0.0086, 0.1086, 0.0984 and 0.1709 unit increase on the proportion of household expenditure on education per household.

Results Discussion

The study used fitted covariate quantiles, residual and psr functional analyses to aid in the data exploration.

Table 4 gave a summary of the covariate quantiles for fitted model coefficients.

Table 4. *Fitted Covariate Quantiles*

	2.5%	25%	50%	75%	97.5%
M.Intercept	-3.686	-3.4565	-3.3223	-3.2017	-2.9486
P. Intercept	1.449	1.5224	1.5666	1.6048	1.6968
Distance	-2.118e-01	-0.1623	-0.1390	-0.1175	-0.0824
HH. Asset	-1.958e-05	0.0003	0.0004	0.0006	0.0009
HHH. Age	2.863e-03	0.0067	0.0086	0.0107	0.0141
HHH. Edu	1.026e-02	0.0704	0.1083	0.01463	0.2108
HHG. Male	-6.213e-01	-0.5026	-0.4373	-0.3784	-0.2247
HH. Size	1.300e-01	0.1587	0.1709	0.1844	0.2066

The mean and precision intercepts had a respective mean posterior probability of -3.3223 and 1.5666 with lower and upper quantiles being -3.686, 1.449 and -2.9486, 1.6968 respectively. In this regard, the mean model had a -3.4565 average unit effect in the prediction of proportion of household expenditure on education while the precision model had a 1.5666 average effect.

Distance of a given household to an educational facility and the chance of a household head being male had a -0.1390 and -0.4373 respective average unit decrease in the prediction of proportion of household expenditure on education. Consequently, household value of assets, age of household head, education levels of household head and the household size had a respective 0.0004, 0.0086, 0.1083 and 0.1709 average unit increase in the prediction of proportion of household expenditure on education.

Residual Analysis

Table 5 and Figure 2 gave a summary and a graphical visualization of the fitted model residuals. The maximum, minimum and mean residuals were 5.2715, -4.5116 and 0.0004 respectively. The median residuals were higher than the mean residuals which ascertained that majority of the residuals were to the right of the mean value. From Table 5 and Figure 2, there existed an almost symmetry in the fitted model residuals which gave an indication of the goodness of fit of the fitted model in modeling the proportion of household expenditure on education.

Table 5. *Fitted Model Residuals*

Min	1st Q	Med	Mean	3rd Q	Max
-4.5116	-0.4695	0.0518	0.0004	0.5767	5.2715

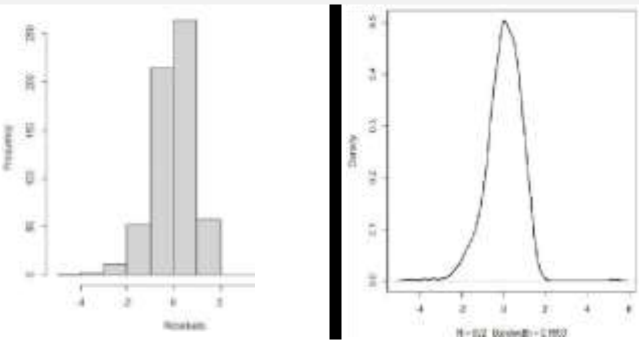

Figure 2. *Histogram and Density of Fitted Model Residuals*

Pseudo Ranks

The Pseudo Rank Function (psrf) was used to check whether multivariate rank value can be calculated for multidimensional model parameters for proportion of household expenditure on education data. It gave a summary statistics on multiple univariate psrf values and the associated box plots as in Table 6 and Figure 3.

Table 6. *Pseudo ranks for proportion of household expenditure spent on education*

	Min	1st Q	Med	Mean	3rd Q	Max
Point Estimate	0.9966	1.0020	1.0070	1.0097	1.0205	1.0220
Upper CI	0.9968	1.0136	1.0200	1.0176	1.0231	1.0286

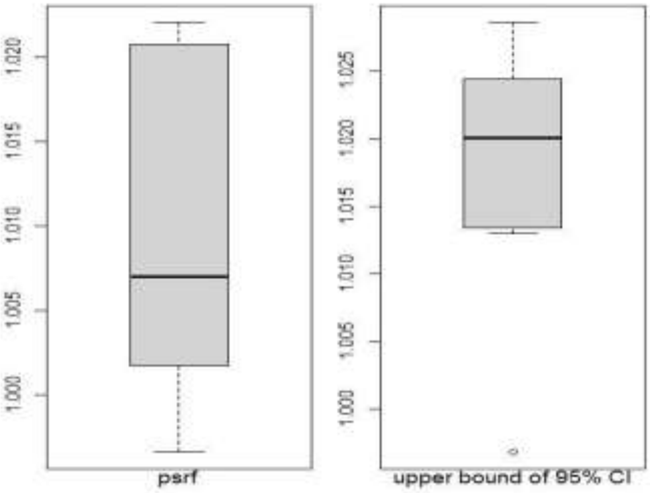

Figure 3. *Box plot for Pseudo ranks of the data*

The maximum and minimum pseudo ranks for the point estimates were 1.0220 and 0.9966 respectively with a mean of 1.0097. The median and mean point estimates were close to each other which gave an implication of majority of the pseudo ranks for the point estimates being centered around the mean. Equally, the maximum and minimum upper confidence intervals for the point estimates were 1.0286 and 0.9968 respectively.

Model Trace and Density Plot

Figure 4 gave the trace and density plot for the most significant covariates in estimating the proportion of household expenditure spent on education. This were education levels of the household head (HHH.Edu), gender of the household head-male (HHG-Male), household size (HH Size) and the mean intercept model. Their densities were symmetrical in nature which gave an indication of their correctness in the estimation and prediction of the proportion of household expenditure spent on education.

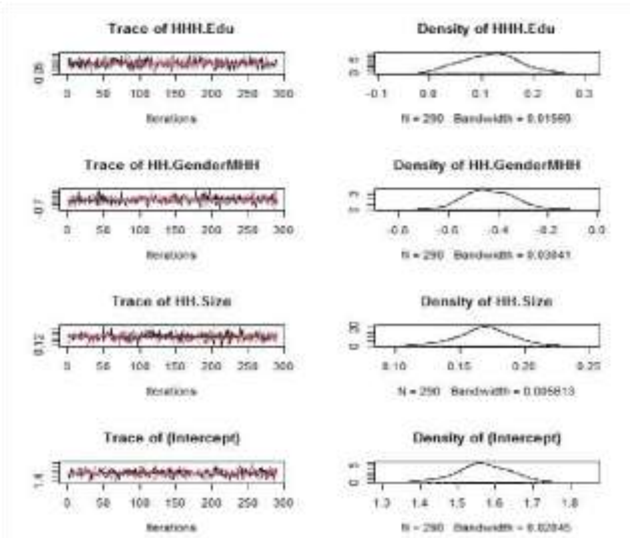

Figure 4. *Trace and Density plot of the data*

4.3. Model Prediction

Table 7. *Household Expenditure on Education Prediction*

Min	1st Q	Med	Mean	3rd Q	Max
0.01698	0.06669	0.09054	0.09708	0.11885	0.30128

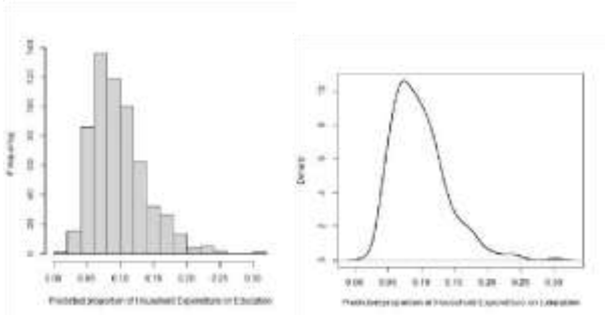

Figure 5. *Histogram and Density of Predicted Household Expenditure on Education*

In doing model predictions, Table 7 gave a summary of the household expenditure on education prediction while Figure 5 gave a graphical visualization of the same. The maximum and minimum predicted proportion of household expenditure on education were 0.30128 and 0.01698 respectively with a mean of 0.09708. Majority of the predicted proportional household expenditures were centered close to the mean as evidenced by the median being close to the mean.

Test for Predicted and Observed Household Expenditure on Education

In order to test for association between the predicted and observed proportion of household expenditure on education, the Chi-square test of association and posterior probability plots were used as given in Table 8 and Figure 6.

Table 8. *Chi-Square test for predicted and observed Proportion of household expenditure spent on education*

X-Squared	Df	P-Value
21.524	6	0.001416

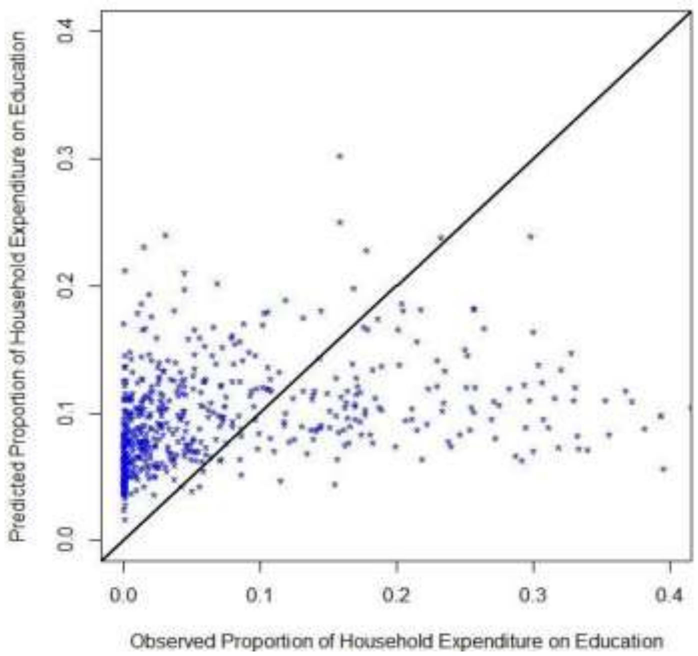

Figure 6. *Posterior Mean Plot of Expected and Predicted Household Expenditure on Education*

The Chi-Square test of association had a p-value of 0.001416 which was less than 0.005 at 6 degrees of freedom. This confirmed a presence of association at an X-Squared value of 21.524 thus giving an indication of the goodness of fit of the Inflated Beta Regression model in the modeling of proportion of household expenditure on education. For the posterior mean plot of expected and predicted proportion of household expenditure on education, majority of the observations were close to zero giving an indication of low household expenditures on education.

V. Conclusions and Recommendations

Individual and societal development with regard to social-economic aspects is greatly influenced by the levels of education of individuals within a given society. It's in this regard that education plays a big and important role in the well-being of society members as educated. Educated persons will always tend to make wise decisions in their day to day activities as opposed to the un-educated ones. In the arid and semi-arid areas of Kenya, there is a low intake of education with a number of factors contributing to the same [2].

This study sought to examine the factors that influence the proportion of household expenditure spent on education in Isiolo County, one of the counties in the arid and semi-arid areas of Kenya. This proportion was computed a ratio of the household expenditure spent on education to the total household expenditure per household. The distance of a household to an educational facility, household asset value, age of household head, education levels of household head, gender of the household head and the household size were evaluated using a Bayesian Inflated Beta Regression model. All the factors had a significant effect on the proportion of household expenditure spent on education with the household size being the most significant.

A total of six hundred and two (602) households were employed in the study in which four hundred and fifty (450) had a male household head and one hundred and fifty two (152) households had a female household head. The highest level of education of a household head was post-secondary and the lowest was zero. The largest household size was fifteen (15) and the smallest one (1). The largest household head age was one hundred and one (101) years while the least was eighteen (18) years. The largest distance of a household to an educational center was ten (10) kilometers and the smallest one (1) kilometer. For household asset value, the highest was 1.8 million while the least was negative four thousand, seven hundred and seventy.

The negative household asset value gave an indication of some households being bankrupt and possessing loans from financial institutions. The model parameters and posterior probabilities were estimated via the maximum likelihood and The Bayesian inference respectively. Model goodness of fit tests were done by the evaluation of the deviance residuals, probability plots and the Chi-Square test of association for predicted and estimated proportion of household expenditure on education. The trace plots, auto-correlation plots and the Gelman plot were used to test the convergence of the Markov Chain Monte Carlo (MCMC) simulations. The model diagnostics showed the presence of association between the estimated and predicted proportion of household expenditure on education.

Recommendations

This study concludes that in the modeling of proportions of household expenditure on education per household, the Bayesian Inflated Beta Regression Model ought to be used. This is as with Luiz et al [12], Danielle et al [4] and Debora et al [16] who used the Inflated Beta Regression model in the modeling of bounded response variables in the interval (0,1). In the determination of significant factors affecting household expenditure, the demographic factors of a household were found to be having a high contribution to an increase in the proportion of household expenditure spent on education.

The government of Kenya ought to come up incentives aimed at addressing the low intake of education in its arid and semi-arid areas. This is as with increasing the number of learning institutions in those regions so as to ensure easy accessibility of learning institutions by the inhabitants. Also, society members ought to be educated and provided with family planning mechanisms so that households can get children they can adequately care and provide for. Much emphasy should also be on educating the members of the general public on the importance of education. This study stipulates the need for further research to determine the association between the different household expenditures per household and determine factors responsible for the same.

Data Availability

This is a quantitative study and real data was used in writing this manuscript. A copy of the data can be obtained by request from the corresponding author.

Conflict of Interest

We as the authors declare no potential conflict of interests

Authors' Contributions

In this work I (Immaculate Atieno) performed the manuscript write up and the data analysis under the supervision of Samuel Mwalili and Oscar Ngesa.

Acknowledgements

I thank Almighty God, my guardians, brothers, sisters and friends for their constant encouragement without which this assignment would not be possible. I take this opportunity to also express my profound gratitude and deep regards to my supervisors Dr. Oscar O. Ngesa and Prof. Samuel .M. Mwalili for their exemplary guidance, monitoring and constant encouragement throughout the course of the study. A deep sense of gratitude also goes to my colleagues at the FAO Resilience Team for Eastern Africa, for their cordial support, valuable information and guidance, which helped me in completing this task through various stages. I am grateful for their cooperation during the period of my research.

References

[1]. S.L.P. Ferrari and R. Ospina, "A general class of zero-or-one inflated beta regression models," *Computational Statistics and Data Analysis,* vol. 56, no. 6, pp. 1609-1623, 2011.
[2]. Government of Kenya (GoK), *Vision 2030 Development Strategy for Northern Kenya and other Arid Lands,* Government of Kenya, 2012, https:www.ndma.go.ke.
[3]. Z. Danuta, S. Agnieszka and K.K. Agnieszka, "Application of logistic regression model to assess household financial decisions regarding debt," *Procedia Computer Science,* Vol. 176, no. 1, pp. 3418-3427, 2020.
[4]. S.P. Danielle, A.G. Terezinha, J. Vanderly, MOG, Aline et al, "Modeling the incidence of citrus canker in leaves of the sweet orange variety," *ActaScientiarum,* Vol. 39, no. 1, pp. 1-8, 2017.
[5]. D.R. Bahadar, K.M. Abdul and A. Akhter, "Household energy consumption and its determinants in timor-leste," Asian *Development Review,* Vol. 34, no. 1, pp. 167-197, 2017.
[6]. T.N.O. Achia, A. Wangombe and N. Khadioli, "A logistic regression model to identify key determinants of poverty using demographic and health survey data," *European Journal of Social Sciences,* Vol. 13, no. 1, pp. 38-45, 2010.
[7]. P. Xiaoling, L. Gang and Z. Liu, "Zero-inflated beta regression for differential abundance analysis with metagenomics data," *Journal of Computational Biology,* Vol. 23, no. 2, pp. 102-110, 2016.
[8]. A. Prahutama, R. Agus, M.M. Abdul and W. Tatik, "Analysis of household expenditures on education in samarang city, indonesia using tobit regression model," *E3S Web of Conferences,* Vol. 125, no. 1, pp. 1-4, 2019.
[9]. L.K. Jonah, P.D. Dewitt, J.F. Jeremy and S.J. Noemie, "Estimating plant abundance using inflated beta distributions; applied learnings from a lichen-caribou ecosystem," *Ecology and Evolution,* Vol. 2017, no. 7, pp. 486-493, 2016.
[10]. L. Tony, "Modeling household spending using a random assignment scheme," *International Journal of Microsimulations,* Vol. 6. No. 2, pp. 56-75, 2013.
[11]. M. Manos, M. Theodore and T. Panos, "Modeling health expenditure at the household level in Greece," *European Journal of Health Economics,* Vol. 10, no. 1, pp. 29-336, 2014.
[12]. M.A. Luiz, L.P. Tarciana, C.S. Tatiene and M.B. Fabio, "Process monitoring using inflated beta regression control chart," *PloSONE,* Vol. 15, no. 7, pp. 105-122, 2020.
[13]. S. Buigut, E. Remare, and D. Amendah, "Catastrophic health expenditure and its determinants in Kenya slum communities," *International Journal for Equity in Health,* Vol. 14, no. 46, pp. 1-12, 2015.
[14]. D. Partha and C.N. Edward, "Modeling health care expenditure and use," *Annual Review of Public Health,* Vol. 39, no. 1, pp. 489-505, 2018.
[15]. M. Sohom, K.S. Roshan and P.S. Slobodan, "Use of beta regression for statistical downscaling of precipitation in the Campbell river-basin, British Columbia, Canada," *Journal of Hydrology,* Vol. 538, no. 2016, pp. 49-62, 2016.
[16]. F.S. Debora and A.S.M. Fernando, "Multivariate beta regression with application in small area estimation," *Journal of official statistics,* Vol. 32, no. 3, pp. 7475-7515, 2016.
[17]. I. Pudji, I. Nur, S. Setiwan and S.U. Brodjol, "Toward a hierarchical bayesian framework for modeling the effect of regional diversity on household expenditure," *Journal of Mathematics and Statistics,* Vol. 8, no. 2, pp. 283-291, 2012.
[18]. P.G. Mwangi and G.K. Kariuki, "Household poverty determinants in Kenya: A demographic and health survey wealth index approach," *Journal of Pan African Studies,* Vol. 9, no. 3, pp. 52-65, 2016.
[19]. A. Omid and G. Mousa, "Bayesian analysis of regression models using instrumental variables: A case study (Iranian rural household income and expenditure data)," *Journal of Statistical Research of Iran,* Vol. 14, no. 1, pp. 53-75, 2017.
[20]. S.A. Donkoh and J.A. Amikuzuno, "The determinants of household education expenditure in Ghana," *Educational Research and Reviews,* Vol. 6, no. 8, pp. 570-579, 2011.
[21]. T. Habanabakize, "Determining the household consumption expenditure resilience towards petrol price, disposable income and exchange rate volatilities," *Economies,* Vol. 9, no. 87, pp. 1-15, 2021.
[22]. S. Brown, P. Ghosh, S. Li and T. Karl, "Modeling household finances: A bayesian approach to a multivariate two-part model," *Journal of Empirical Finance,* Vol. 3, no. 17, pp. 1-18, 2021.
[23]. R.W. Eyre, F.G. Xavier and E.G. Frances, "Bayesian belief network modeling of household food security in rural South Africa," *BMC Public Health,* Vol. 21, no. 93, pp. 1-16, 2021.

Immaculate Atieno, et. al. "Bayesian Zero/One-Inflated Beta Regression Modeling of Household Expenditure on Education in Marginalized Communities in Kenya." *IOSR Journal of Mathematics (IOSR-JM)*, 18(3), (2022): pp. 04-13.