

THE USE OF BAYESIAN QUALITATIVE AND LIMITED DEPENDENT VARIABLE MODELS IN EVALUATING WOMEN PARTICIPATION IN AGRICULTURAL PRODUCTION ACTIVITIES IN OIL RICH AREA OF NIGERIA

FELIX UZOMA OCHUBA;

*Department Of Statistics,
Michael Okpara University of Agriculture Umudike,
Abia State, Nigeria.*

DAVID ADIELE

*Department Of Statistics,
Michael Okpara University of Agriculture Umudike,
Abia State, Nigeria.*

And

CATHERINE ULOAKU OKEREKE

*Department Of Statistics,
Michael Okpara University of Agriculture Umudike,
Abia State, Nigeria.*

Abstract

This study examined the use of Bayesian qualitative and limited dependent variable models in women's participation in agricultural production activities in Ogoni, Rivers State, Nigeria. In specifics, the study aimed to; ascertain the socio-economic characteristics of the women farmers, and use a Bayesian approach to estimate the parameters of our proposed model and compare the Bayesian logistic regression result with the classical logistic regression. Data were analysed using the Bayesian logistic regression model as well as comparing it with the classical logistic regression. The findings of the study revealed a positive and significant relationship between selected socio-economic characteristics of the respondents and women's level of participation in agricultural production activities, except age and cooperative membership which were negatively signed. Holding other variables at fixed value, we would see an average of 110.5% increase in the odds of having high women participation in agriculture production activities for a unit increase in level of education (X_2). There is an average of 100% increase in the odds of getting high women participation in agricultural production activities for a unit increase in household size (X_3). Bayesian estimation performed better than the Classical Logistic Regression (maximum likelihood estimation) in terms of parameter estimates as the Bayesian logistic model had lower standard deviation and standard error of mean (SEM) compared to the Classical logistic (vis-à-vis narrower credible interval versus the confidence interval of the frequentist). However, both the Bayesian and Classical logistic models performed very well in terms of model prediction accuracy with prediction accuracy values of 0.80 and 0.8025 respectively. The leave-one-out cross validation (LOOCV) prediction accuracy was 0.78 or 78%. This is still high enough and implies that the accuracy of the model is 78% likely to hold up when used in a different data set. The study concludes that, predictions in women participation in agricultural production activities is very efficient when using Bayesian Logistic Regression using the independent Cauchy prior as proposed by Gelman et al (2008). It is therefore recommended that Bayesian inference and models can be adopted in estimating predictions in women participation in agricultural

Keywords: Bayesian logistic regression, Independent Cauchy priors, Qualitative data, Odds ratio, Limited Dependent Variables, Agricultural production

Introduction

Bayesian inference is a statistical inference method that uses Bayes' theorem to update a hypothesis' probability when more data or information becomes known. It is a crucial tool in statistics, and particularly in mathematical statistics. Inductive reasoning based on the reconsideration of competing hypotheses in the presence of new evidence is known as Bayesian inference. Bayesian inference takes place in the context of one or more parametric econometric models. Let y_t denote a $p \times 1$ vector of observable random vectors over a sequence of discrete time units $t = 1, 2, \dots$. The history of the sequence $\{y_t\}$ at time t is given by $Y_t = \{y_s\}_{s=1}^t \in \Psi_t; Y_0 = \{\emptyset\}$. A model, A , specifies a corresponding sequence of probability density functions (p.d.f.'s) $p(y_t|Y_{t-1}, \theta, A)$ in which θ is a $k \times 1$ vector of unknown parameters, $\theta \in \Theta \subseteq \mathbb{R}$, and A denotes the model (Lukman *et al.*, 2021).

In probability theory plus statistics, Bayes' theorem designates the probability of an occurrence, contingent on prior knowledge of circumstances that could be connected to the occurrence. People, conversely, seek guidelines to help them make decisions in the face of ambiguity. They want to know how much uncertainty there is about parameter values based on the data they have. That it intends to help with this is the distinctive trademark of Bayesian inference.

A limited dependent variable (LDV) is one whose range of possible values is "significantly confined" (Gelman and Carlin, 2014). When estimating the association between a limited dependent variable of interest plus other variables, the word is widely used in econometrics since it needs methodologies that account for this constraint. Limited dependent variable (LDV) plus discrete choice models are used in a variety of sectors, including economics, finance, marketing, politics, and sociology, to name a few. LDV models, in other words, solve two problems: censoring plus truncation. Censoring occurs when the limit observations are existent in the sample, while truncation occurs when they are not.

The term "qualitative" follows the interpretation which subsumes the following types of variables: unordered categorical data (nominal scale), ordered categorical data (ordinal scale), integer-valued data ("count data") plus censored data ("limited dependent" variables). A LDV is one whereby the dependent variable has a bound or border, and some observations fall within it. A continuous variable having many recurrent observations at the higher or lower bounds is referred to as a limited dependent variable (LDV). Consumption of a product, amount of hours worked by women, and so on are examples. Tobit model (via censored sample), Heckman model (sample selection model), truncated regression (via truncated sample), plus others are examples of LDV models.

Despite their expanding popularity, limited dependent variable approaches appear to be plagued by significant issues such as non-identifiability and problem of collinearity. Also, discrete-data regression can also become unstable from separation, which arises when a linear combination of the predictors is perfectly predictive of the outcome (Gebrie, 2021). In

furtherance to this, agricultural production activity is very crucial to food security and economic growth of any nation (Okah, 2018). Consequently, investigating the participation of women in agricultural production activities in Ogoni, Rivers State will help provide answers to certain agricultural problems in the area. The oil rich area of Ogoni have been saddled with lot of agricultural challenges and out of literature is the women participation in agriculture, since it is believed women in oil rich areas are usually lazy as their farmlands are adversely affected by gas flaring. This work investigated the participation of women in Agriculture and thus bridge this gap in literature.

Therefore, this work shall seek to;

- i. ascertain the socio-economic characteristics of the women farmers in the oil rich area of Rivers State. This paper shall employ
- ii. Bayesian approach to estimate the parameters of our proposed model and compare the Bayesian logistic regression result with the classical logistic regression.

Literature Review

The posterior distribution utilized is a conditional probability density function of β given $Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n$, is denoted by $f(\beta|y_1, y_2, \dots, y_n)$, owing to the Bayesian theorem could be written as;

$$f(\beta|y_1, y_2, \dots, y_n) = \frac{f(y_1, y_2, \dots, y_n|\beta)f_B(\beta)}{g(y_1, y_2, \dots, y_n)} = \frac{f(y_1, y_2, \dots, y_n|\beta)f_B(\beta)}{\int_{-\infty}^{\infty} f(y_1, y_2, \dots, y_n|\beta)f_B(\beta)d\beta} \quad (1)$$

Where; $f_B(\beta)$ is a prior distribution and $f(y_1, y_2, \dots, y_n|\beta)$, is a likelihood function.

Gebrie (2021) investigated the Bayesian regression model and its application to a cross-sectional study of food insecurity at the home level. The major determinants of household food insecurity were identified using Bayesian estimation on binary logistic regression.

The prior distribution of logistic regression parameter is specified by:

$$P(\beta_j) = \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp \left\{ -\frac{1}{2} \left(\frac{\beta_j - \mu_j}{\sigma_j} \right)^2 \right\} \quad (2)$$

The posterior distribution is proportionate to the product of the likelihood function, and the prior distribution over all parameters from the data, as follows:

$$\begin{aligned} \Pi(\beta/\text{data}) &= \prod_{i=1}^n \left[\left(\frac{e^{\beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip}}}{1 + e^{\beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip}}} \right)^{y_i} \left(1 - \frac{e^{\beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip}}}{1 + e^{\beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip}}} \right)^{(1-y_i)} \right] \\ &\times \prod_{j=0}^p \frac{1}{\sqrt{2\pi}\sigma_j} \exp \left\{ -\frac{1}{2} \left(\frac{\beta_j - \mu_j}{\sigma_j} \right)^2 \right\} \end{aligned} \quad (3)$$

Okah (2018) conducted a study to assess women's participation in agricultural production in the remediated Ogoni land of Rivers State, Nigeria. The inferential statistical tools used were ANOVA and multiple regression analysis (OLS estimation). The multiple linear regression equation is as follows:

$$\hat{Y} = b_0 + b_1X_1 + b_2X_2 + \dots + b_pX_p, \quad (4)$$

where: $\hat{Y}$ is the predicted value of the response variable, X_1 through X_p are p separate independent variables, b_0 is the value of Y given that all the independent variables (X_1 through X_p) are equal to zero, while b_1 through b_p are the estimated regression coefficients.

None of these researchers focused on the use of Bayesian qualitative and limited dependent variable models in women participation in agricultural production activities in Ogoni. This constituted a gap in knowledge which this study bridges.

Methodology

The study adopted the Data used in Okah (2018) which were collected from four (4) Local Government Areas in Ogoni, Rivers State (Khana, Tai, Gokhana and Eleme) using multi-stage sampling technique to arrive at a Sample size of 400.

Logistic Regression Model

Assume that y is a binary response variable, and the outcomes are coded $y=1$ or $y=0$ with respective probability p and $1-p$. For a particular subject, the logistic regression equation is expressed as;

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n \quad (5)$$

Where $x_1, x_2, \dots, x_n$ are the subject's feature measurements on a group of explanatory variables, and $\beta_1, \beta_2, \dots, \beta_n$ are parameter estimates for explanatory variables.

Model Specification: Bayesian Logistic Regression Model

Cauchy Prior Distribution

This study adopted the Independent Cauchy Prior with center 0 and scale 10 to the constant term (β_0) and Cauchy with center 0 and scale 2.5 to each of the coefficients ($\beta_1, \beta_2, \dots, \beta_k$) in the logistic regression as proposed by Gelman *et al.* (2008).

The Independent Cauchy Prior suggested by Gelman *et al.* (2008) corresponds to;

$$\beta_j \sim \text{ind}C(0, \sigma_j), \text{ for } j = 1, 2, \dots, p. \quad (6)$$

The Cauchy distribution has the Probability Density Function (PDF):

$$f(x; x_0, \gamma) = \frac{1}{\pi\gamma \left[1 + \left(\frac{x-x_0}{\gamma}\right)^2\right]} = \frac{1}{\pi} \left[\frac{\gamma}{(x-x_0)^2 + \gamma^2} \right], \quad (7)$$

Where x_0 is the location parameter, specifying the location of the peak of the distribution, and γ is the scale parameter which specifies the half-width at half-maximum (HWHM), alternatively 2γ is full width at half maximum (FWHM).

Likelihood Function

The likelihood contribution from the i^{th} subject is binomial and the likelihood function is given by:

$$L(y_i/\beta) = \prod_{i=1}^n \left[(P_i)^{y_i} (1 - P_i)^{1-y_i} \right] = \prod_{i=1}^n \left[\left(\frac{e^{\beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik}}}{1 + e^{\beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik}}} \right)^{y_i} \left(1 - \frac{e^{\beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik}}}{1 + e^{\beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik}}} \right)^{1-y_i} \right] \quad (8)$$

$$\pi(\beta/data) = \prod_{i=1}^n \left[\left(\frac{e^{\beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik}}}{1 + e^{\beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik}}} \right)^{y_i} \left(1 - \frac{e^{\beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik}}}{1 + e^{\beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik}}} \right)^{1-y_i} \right] * \prod \frac{1}{\pi^\gamma \left[1 + \left(\frac{x-x_0}{\gamma} \right)^2 \right]} \quad (9)$$

Method of Data Analysis

Bayesian logit model and the Classical logistics regression were used as well as descriptive statistics. However, the dependent and independent variables for the study include;

Y = Participation in Agric (1 – high participation in agriculture; 0 = low participation)

X₁= Age (years),

X₂= Level of education (Years of formal education),

X₃= Household Size (Actual number),

X₄= Farm Income (Naira),

X₅= Farm size (Dummy; 0.1-5.09 =0, 5.10 & above= 1),

X₆= Marital Status (Dummy; Married = 1, Otherwise = 0),

X₇= farming experience (years),

X₈= cooperative membership (Dummy; member =1, non-member = 0)

X₉ = access to farm inputs (Dummy; Yes = 1, No = 0),

Leave-One-Out-Cross-Validation (LOOCV)

Cross validation addresses the 'over-fitting' problem. The fundamental idea of predictive modeling is that the data are split into two subsets, the training and the test data. The training data are used to construct a model, and the performance of the model on the test data is utilised to check whether the predictions generalize well.

1. Results and Discussion

Evaluation of Socio-Economic Characteristics

Table 1: Distribution according to socio-economic characteristics of respondents in the study area

Variables	Eleme		Gokhana		Khana		Tai		Ogoni-land	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Age										
21-30	11	11.00	2	2.00	16	16.00	10	10.00	39	39.00
31-40	37	37.00	40	40.00	35	35.00	57	57.00	169	42.25
41-50	24	24.00	13	13.00	20	20.00	15	15.00	72	18.00
51-60	12	12.00	29	29.00	18	18.00	10	10.00	69	17.00
61-70	16	16.00	16	16.00	11	11.00	8	8.00	51	12.75
Total									400	100
Marital Status										
Single	23	23.00	11	11.00	33	33.00	8	8.00	75	18.75
Married	50	50.00	52	52.00	56	56.00	75	75.00	233	58.25
Separated/divorced	16	16.00	28	28.00	5	5.00	16	16.00	65	31.00
Total									400	100
Educational Qualification										
Primary	49	49.00	72	72.00	50	50.00	66	66.00	237	59.25
Secondary	34	34.00	19	19.00	40	40.00	31	31.00	124	31.00
Total									400	100
Household size										
1-3	40	40.00	40	40.00	45	45.00	43	43.00	168	42.00
4-6	37	37.00	38	38.00	30	30.00	30	30.00	135	33.75
7-10	15	15.00	11	11.00	10	10.00	12	12.00	48	12.00
11-13	5	5.00	9	9.00	8	8.00	10	10.00	32	8.00
14-16	3	3.00	2	2.00	7	7.00	5	5.00	17	4.25
Total									400	100
Cooperative membership										
Member	79	79.00	67	67.00	74	74.00	54	54.00	274	68.00
Total									400	100
Farming experience										
1-5	11	11.00	24	24.00	17	17.00	15	15.00	67	16.75

The Use of Bayesian Qualitative and Limited Dependent Variable Models in Evaluating Women Participation in Agricultural Production Activities in Oil Rich Area of Nigeria

6-10	26	26.00	30	30.00	23	23.00	20	20.00	99	24.75
11-15	36	36.00	32	32.00	38	38.00	44	44.00	150	37.50
16-20	27	27.00	14	27.00	22	22.00	21	21.00	84	21.00
Total									400	100
Farm size										
0.1-5.09	60	60.00	80	80.00	83	83.00	61	61.00	284	71.00
5.10	40	40.00	20	20.00	17	17.00	39	39.00	116	29.00
Total									400	100
Farm income										
1000-100,000	21	21.00	20	20.00	30	30.00	31	31.00	102	25.50
100,001-200,000	61	61.00	53	53.00	49	49.00	54	54.00	217	54.25
200,001-300,000	10	10.00	15	15.00	13	13.00	5	5.00	43	10.75
300,001-400,000	8	8.00	12	12.00	8	8.00	10	10.00	38	9.50
Total									400	100
Access to farm inputs										
Yes	16	16.00	13	13.00	14	14.00	15	15.00	58	14.50
No	84	84.00	87	87.00	86	86.00	85	85.00	342	85.50
Total									400	100

Classical Logistic Regression

Table 2: Parameters of the classical logistic regression model and their evaluation

Coefficients:					
	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.364e+01	9.758e+02	-0.014	0.989	
Age (X ₁)		-1.391e-01	2.680e-02	-5.188	2.13e-07 ***
Level of edu (X ₂)	5.721e-02	4.817e-02	1.188	0.235	
Household Size (X ₃)	-9.734e-04	3.563e-02	-0.027	0.978	
Farm Income (X ₄)	1.292e-06	1.791e-06	0.722	0.471	
Farm size (X ₅₁)	2.305e+00	1.111e+00	2.075	0.038 *	
Marital Status (X ₆₁)	1.812e+01	9.758e+02	0.019	0.985	
Farming exp (X ₇)	-2.760e-02	3.147e-02	-0.877	0.380	
Coop. member(X ₈₁)	-3.947e-01	5.809e-01	-0.679	0.497	
Access to farm input (X ₉₁)	1.843e+01	9.758e+02	0.019	0.985	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					
Null deviance: 479.92 on 399 degrees of freedom					
Residual deviance: 318.17 on 390 degrees of freedom					
AIC: 338.17					

From Table 2, which revealed the parameters of the classical logistic regression model and their evaluation, it is apparent that only two of the independent variables were significant (X₁ and x₅₁). The negative coefficient for X₁ (age in years), -0.1391, implies that holding other variables at fixed value, we would see an average of 87% decrease in the odds of having high

women participation in agriculture production activities for a unit increase in age of women in years (X_1). This was derived as; $\exp(-0.1391) \times 100 = 87\%$. This result implies that the older the women farmers the lower their participation in agriculture production activities.

Table 3: 95% confidence interval for the parameters

	5 %	95 %
(Intercept)	-3.714483e+02	-2.611036e+02
x1	-1.848050e-01	-9.641452e-02
x2	-2.150174e-02	1.374602e-01
x3	-5.999920e-02	5.756201e-02
x4	-1.673729e-06	4.281261e-06
x51	7.386620e-01	4.655421e+00
x61	1.033582e+02	3.676395e+02
x7	-7.968759e-02	2.408207e-02
x81	-1.329848e+00	5.997636e-01
x91	1.686149e+02	3.712658e+02

These confidence intervals (CI) are ranges of values that are likely to contain the true values of the odds ratios. A 95% CI is the interval that we are 95% sure encompasses the true population value as it might be assessed from a much greater study. Hence, from *Table 3* for X_1 (Age in years) we can be 95% confident that the true estimate of the odds of having high women participation in agriculture production activities for a unit decrease in age of women in years lies between the interval of -0.185 to -0.096 based on hypothesized repeats of the experiment.

Table 4: Classification table under classical logistic regression

cc	0	1
0	256	50
1	29	65

Prediction accuracy for classical logistics regression = **0.8025**

The Prediction accuracy for the classification under the logistics regression is 0.8025. This is about 80% which is high and implies that model has a high predictive accuracy. However, the accuracy discussed above is computed based on the same data that is used to fit the model. To evaluate prediction accuracy more thoroughly, run the calculation using data that were not used to build the model. This evaluates how well the model's accuracy will hold up under "real world" application. This would be demonstrated in the Bayesian approach using the leave one out cross validation (LOOCV).

Bayesian Logistic Regression

Table 5: Bayesian logit estimates

Estimates:	Coefficients (mean)	SD	2.5%	25%	50%	75%	97.5%
(Intercept)	-2.5	3.8	-11.9	-4.3	-1.8	0.1	3.0
Age (X ₁)	-0.1	0.0	-0.2	-0.2	-0.1	-0.1	-0.1
Level of edu (X ₂)	0.1	0.0	0.0	0.0	0.1	0.1	0.2
Household Size (X ₃)	0.0	0.0	-0.1	0.0	0.0	0.0	0.1
Farm Income (X ₄)	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Farm size (X ₅₁)	2.5	1.1	0.6	1.7	2.4	3.2	4.9
Marital Status (X ₆₁)	7.0	3.6	2.5	4.5	6.2	8.5	16.6
Farming exp (X ₇)	0.0	0.0	-0.1	0.0	0.0	0.0	0.0
Coop. member(X ₈₁)	-0.4	0.6	-1.5	-0.8	-0.4	0.0	0.8
Access to farm input (X ₉₁)	7.1	3.6	2.6	4.6	6.2	8.6	16.7

Fit Diagnostics:

Mean_PPD	0.3	0.0	0.2	0.3	0.3	0.3	0.3
Log-posterior	-186.6	2.2	-181.5	-177.8	-176.1	-174.9	-173.2

mean_PPD - the mean of the posterior predictive distribution, which is the mean of the predictions made using the estimated parameters.

log-posterior - the log-posterior is the logarithm of the likelihood times the prior up to some normalizing constant.

From Table 5, the coefficient for X₁ (age in years), -0.1, implies that holding other variables at fixed value, we would see an average of 90.5% decrease in the odds of having high women participation in agriculture production activities for a unit increase in age of women in years (X₁). This was derived as; $\exp(-0.1) \times 100 = 90.5\%$. This result implies that the older the women farmers the lower their participation in agriculture production activities. Also, the coefficient for X₂ (level of education in years) is 0.1. This reveals that holding other variables at fixed value, we would see an average of 110.5% (that is, $\exp(0.1) \times 100 = 110.5\%$) increase in the odds of having high women participation in agriculture production activities for a unit increase in level of education (X₂). This implies that the more educated the women farmers, the higher the odds of having high participation in agriculture production activities in Ogoni land.

The mean of the posterior predictive distribution (mean_PPD) is 0.3. It is the mean of the predictions made using the estimated parameters. The distribution of the outcome that the model predicts after using the observed data to update our presumptions regarding the model's unknown parameters is known as the posterior predictive distribution. To test the model's fit, simulate data from the posterior predictive distribution via the observed predictors.

Table 6: Bayesian logit regression diagnostics

Diagnostics:			
	se	Rhat	n_eff
(Intercept)	0.1	1.0	813
x1	0.0	1.0	3,025
x2	0.0	1.0	3,722
x3	0.0	1.0	3,604
x4	0.0	1.0	3,277
x51	0.0	1.0	2,391
x61	0.1	1.0	738
x7	0.0	1.0	3,237
x81	0.0	1.0	2,390
x91	0.1	1.0	743
mean_PPD	0.0	1.0	3,894
log-posterior	0.1	1.0	1,576

se - standard error; **n_eff** – number of effective sample size

From table 6, the Rhat value for each of the parameters (the independent variables, $x_1, \dots, x_9$) is one (1) which indicates that there is convergence to the equilibrium distribution (that is, since Rhat is 1 then the chains are in equilibrium. In other words, the chains have converged). Meanwhile, the effective sample size (n_eff) is high enough.

Table 7: Posterior uncertainty

	5%	95%
(Intercept)	-9.887	2.292
x1	-0.185	-0.100
x2	-0.021	0.139
x3	-0.062	0.058
x4	-0.000	0.000
x51	0.885	4.497
x61	2.918	0.145
x7	-0.080	0.024
x81	-1.293	0.648
x91	3.024	14.566

Table 7 is the Bayesian posterior uncertainty results showing a matrix with two (2) columns and ten (10) rows as model parameters. For a given value of probability, p , the columns correspond to the lower and upper $100p\%$ interval limits and have the names $100\alpha/2\%$ and $100(1-\alpha/2)\%$, where $\alpha=1-p$. However, the specified probability is 0.9 (a 90% interval) hence the column names "5%" and "95%", respectively. Unlike frequentist confidence intervals which cannot be explained by post-data probabilities the Bayesian uncertainty interval indicates we believe after seeing the data that there is a 0.9 probability that β_2 is between $ci95[1,1]$ and $ci95[1,2]$. Conversely, we could say that there is fundamentally zero probability

that $\beta_2 > 0$, though frequentists cannot make such a claim comprehensibly. Hence, “the true parameter β_2 has a probability of 0.9 of falling in a 95% credible interval.”

Table 8: Leave-one-out cross-validation (LOOCV) table

gg	0	1
0	251	54
1	34	61

Prediction accuracy for Bayesian logistics regression = **0.8**

LOOCV prediction accuracy = 0.78

The Prediction accuracy for the Bayesian logistics regression is 0.8. This is about 80% which is very high and implies that the Bayesian model has a high predictive accuracy. However, this accuracy was computed based on the same data that was utilized to fit the model. This could be bias. Hence, the LOOCV approach of assessing prediction accuracy was performed using part of the data not utilized to create the model. The LOOCV prediction accuracy was 0.78 or 78%. This is still high enough and implies that the accuracy of the model is 78% likely to hold up when utilized in the "real world" or in a different data set.

Comparing the Bayesian Logistics Model and the Classical Logistic Model

In comparison, the Bayesian estimation performed better than the Classical Logistic Regression (maximum likelihood estimation) in terms of parameter estimates as the Bayesian logistic model had lower standard deviation and standard error of mean (SEM) compared to the Classical logistic (vis-à-vis narrower credible interval versus the confidence interval of the frequentist). However, both the Bayesian and classical logistic models performed very well in terms of model prediction accuracy with prediction accuracy values of 0.80 and 0.8025 respectively. However, this accuracy was computed based on the same data that was utilized to fit the model. This could be biased. Hence, the LOOCV approach of assessing prediction accuracy was performed using part of the data not utilized to create the model. The LOOCV prediction accuracy was 0.78 or 78%. This is still high enough and implies that the accuracy of the model is 78% likely to hold up when utilized in the "real world" or in a different data set.

Graphical Diagnostics

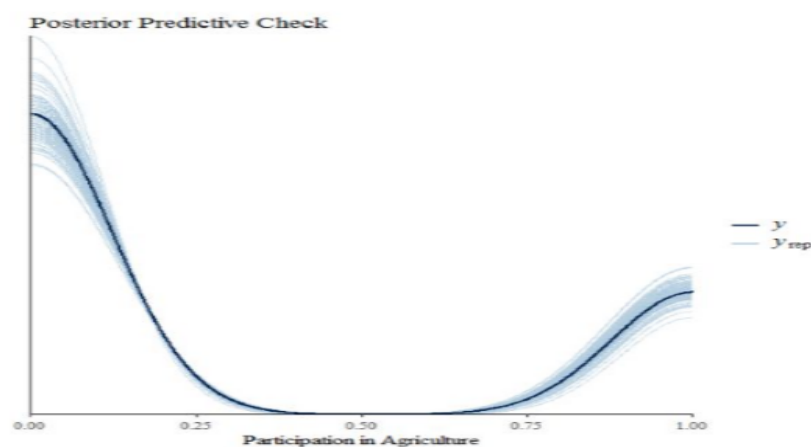


Figure 1: Posterior predictive check

With the posterior predictive check function we can look at how well predictions from the posterior distribution match the true outcome variable that was used to fit the data. In the case of the posterior predictive check, we plotted the 50% uncertainty interval (thick horizontal lines) and the 90% uncertainty interval (thin horizontal lines). In terms of interpretation, the 50% uncertainty interval identifies where 50% of the marginal distribution lies for each parameter. Notice that the graph (Figure 4.1) produced good posterior predictions.

Below are the graphs of the **autocorrelation function** for the chain of each parameter. Notice how, as per the definition of Markov Chains, the past parameter values have no influence on future parameter values. Autocorrelation values close to zero ensure that each value obtained along the chain is random.

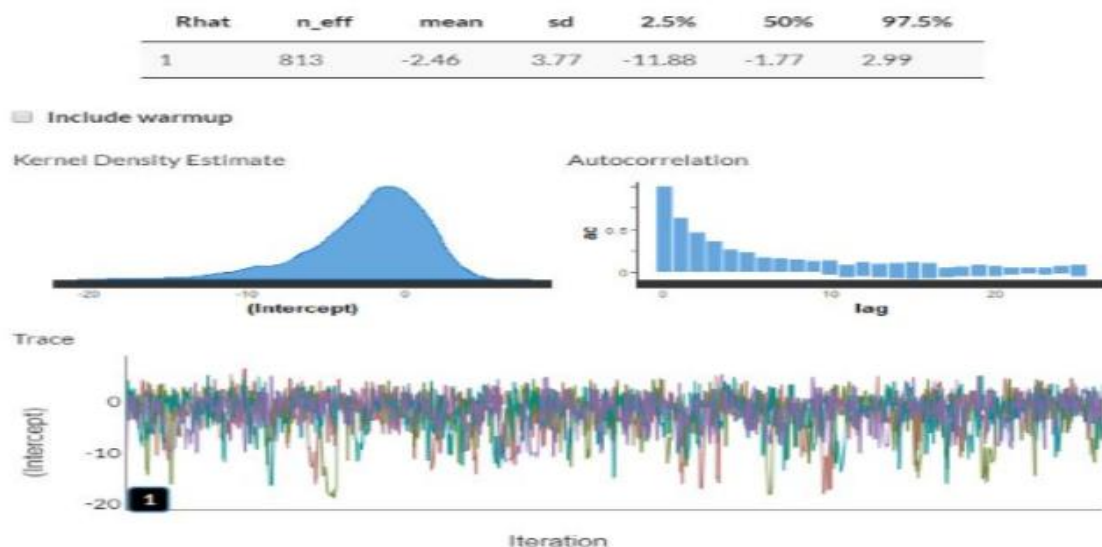


Figure 2: Autocorrelation function for the intercept

Conclusion and Recommendations

The study concludes that, predictions in women participation in agricultural production activities is very efficient when using Bayesian Logistic Regression using the independent Cauchy prior as proposed by Gelman et al (2008). The findings of the study revealed a positive and significant relationship between selected socio-economic characteristics of the respondents and women's level of participation in agricultural production activities, except age and cooperative membership which were negatively signed. Holding other variables at fixed value, we would see an average of 110.5% increase in the odds of having high women participation in agriculture production activities for a unit increase in level of education (X_2). There is an average of 100% increase in the odds of

The Use of Bayesian Qualitative and Limited Dependent Variable Models in Evaluating Women Participation in Agricultural Production Activities in Oil Rich Area of Nigeria

getting high women participation in agricultural production activities for a unit increase in household size (X_3). Also, there is an average of 100% increase in the odds of getting high women participation in agricultural production activities for a unit increase in farm income (X_4). In furtherance, the mean of the posterior predictive distribution is 0.3, which is the mean of the predictions made using the estimated parameters. The Rhat value for each of the parameters (the independent variables, $X_1, X_2, \dots, X_9$) is one (1) which indicates a convergence to the equilibrium distribution. Bayesian estimation performed better than the Classical Logistic Regression (maximum likelihood estimation) in terms of parameter estimates as the Bayesian logistic model had lower standard deviation and standard error of mean (SEM) compared to the Classical logistic (vis-à-vis narrower credible interval versus the confidence interval of the frequentist). However, both the Bayesian and Classical logistic models performed very well in terms of model prediction accuracy with prediction accuracy values of 0.80 and 0.8025 respectively. The leave-one-out cross validation (LOOCV) prediction accuracy was 0.78 or 78%. This is still high enough and implies that the accuracy of the model is 78% likely to hold up when used in a different data set. It is therefore recommended that Bayesian inference and models can be adopted in estimating predictions in women participation in agricultural production activities particularly using Bayesian Logistic Models with independent Cauchy priors.

References

- Gebrie, Y.F. (2021). Bayesian regression model with application to a study of food insecurity in household level: a cross sectional study. *BMC Public Health*, 21:61-90.
- Gelman, A. and Carlin, J. (2014). Beyond power calculations: assessing Type-S (sign) and Type-M (magnitude) errors. *Perspectives on Psychological Science* 9(6): 641-51.
- Gelman, A., Jakulin, A., Pittau, M. G. and Su, Y.S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, 2(4): 1360-1383.
- Gujarati, D. N. (2004). *Basic Econometrics*. 4th ed. New York: McGraw-Hill Higher Education.
- Kleinbaum, D. G. and Klein, M. (2010). *Logistic Regression: A Self-Learning Text*. 3rd ed. Springer Publishers, New York, Inc.
- Koutsoyiannis (2001). *The Theory of Econometrics* (2nd Edition). New York. Palmgrave.
- Laplace, P. S. (1774). Mémoire sur la probabilité des causes par les évènements. *Mémoires de l'Académie Royale des Sciences Présentés par Divers Savants*, 6: 621–656.

- Lukman, P. A., Abdullah, S. and Rachman, A.(2021). Bayesian logistic regression and its application for hypothyroid prediction in post-radiation nasopharyngeal cancer patients. *Journal of Physics: Conference Series*,1725: 1-9.
- Okah, F.N. (2018). Assessment of women's participation in agricultural production activities in the remediated Ogoni land of Rivers State, Nigeria. *An Unpublished Ph.D. dissertation*, Department of Agricultural Extension and Rural Development in the department of Rural Sociology and Extension, Michael Okpara University of Agriculture, Umudike, Nigeria.
- Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge. p. 451. ISBN 0-262-232 19-7. OCLC 47521388.