SENSITIVITY OF BAYESIAN DYNAMIC MIXED LOGIT MODELS TO PRIOR DISTRIBUTIONS

AMALAHU CHRISTIAN CHINENYE[1]; NWABUEZE JOY CHIOMA[2]; ENOGWE SAMUEL UGOCHUKWU[3]; EKEADINOTU, CHIBUEZE BARNABAS[4]

[1]Department of Mathematics,
University of Agriculture and Environmental Sciences, Umuagwo
Imo State, Nigeria.
E-mail: chinenyechris@yahoo.com
Phone: 08034616225

[2]Department of Statistics,
Michael Okpara University of Agriculture, Umudike, Abia State, Nigeria
E-mail: nwabuezejoy@mouau.edu.ng
Phone: 08068164190

[3]Department of Statistics,
Michael Okpara University of Agriculture, Umudike, Abia State, Nigeria
Email: senogwe@yahoo.com
Phone: 08063903400

[4]Department of Mathematics,
University of Agriculture and Environmental Sciences, Umuagwo
Imo State, Nigeria.
E-mail: chibuezeekeadinotu@gmail.com
Phone: 08064138892

ABSTRACT

This research work investigated the performance of various prior distributions on the Bayesian Dynamic Mixed Logistic Regression Model (BDML). The data set used was a Public datasets gotten from UCI Machine Learning Repository. The study compared the performance of Uniform, Jeffrey's, Exponential, Gamma, Cauchy, Normal, and Beta prior distributions in capturing the heterogeneity in customer preferences. The result of the Bank marketing data showed that Jeffery's prior outperforms other priors used in terms of MAE, RMSE, and Log Likelihood this showed that the choice of prior distribution significantly affects the model estimates and predictions.
Keywords: Beta prior distribution, Log Likelihood, Mixed Logistic Regression Model

1. Introduction

The Logistic regression is a widely used statistical model for predicting binary outcomes. It models the probability of an events occurring based on a set of independent variables. When the

ISSN NUMBER: 1116-249X

sample size is small, the likelihood of generating inconsistent, unstable and large Logistic regression estimates is always high. Gelman and Hill (2007) noted that the Logistic regression model and probit model are adequate for modeling binary data; however, they can run into problems when extreme observation(s) and noise in the data exist.

The mixed logit model extends logistic regression by allowing random coefficient to capture individual heterogeneity and preference variation, and also accommodating correlated choices. Mixed logit models have been widely used to capture this heterogeneity.

Bayesian statistics is widely used in the literature for different statistical analysis like in choice modeling.

Bank marketing is the strategic promotion of financial products and services offered by banks to attract and retain customers. Effective bank marketing involves understanding customers’ needs, preferences, and behavior to design targeted campaigns that drives engagements, acquisition, and retention. Hence, the choice of prior distribution for the model parameters can significantly impact the estimation results. Specifying informative priors require a systematic and transparent approach Van de Schoot *et al*. (2021).

Some studies have employed logistic regression such as Kinskey *et al*. (2020), they used logistic regression with feature selection and cross-validation to predict bank marketing campaign success. Their results revealed that logistic regression performed well in predicting campaign success, with the regularization coefficient and penalty type being crucial hyperparameters. Others include Karanja et al. (2022), Kumbhakar *et al*. (2020) etc. However, these models have limitations: such as temporal dependencies and dynamics in mode choice (static nature), assume uniform preferences across individuals (homogeneity) and inability to capture complex relationships; linear relationships between variables are often oversimplified.

Previous studies have explored the use of different prior distributions in Bayesian mixed logit models. For instance, Train (2009) adopted a normal prior distribution for the model parameters, while Greene and Hensher (2010) employed an inverse-Wishart prior distribution for the covariance matrix. Despite these advancements, gaps have been created such as inadequate comparison of prior performance; none of the existing studies has systematically compared the performance of informative, weakly (Normal, Cauchy), and non-informative priors like Jeffreys and their impact on model performance hence, a gap is created in the literature.

ISSN NUMBER: 1116-249X

2. Review of Related Literature

Piironen and Vehtari (2017) compared the performance of different prior distributions in Bayesian Linear Regression using inverse-Wishart, inverse-Gamma and half-Cauchy. Their results showed that inverse-Wishert prior distribution was the best choice for Bayesian linear regression models, interms of predictive performance.

A study by Kumbhakar et al. (2020) applied a mixed logit model to investigate deposit account choice behavior among bank customers, incorporating variables such as account features, fees, and marketing promotions.

Balcombe *et al.* (2009) employed mixed logit (ML) using Bayesian methods to examine willingness-to-pay (WTP) using data generated in a choice experiment. They used marginal likelihood to compare their model, which is preferable for Bayesian model comparison and testing. They considered models containing constant and random parameters for a number of distributions, including models in 'preference space' and 'WTP space' as well as those allowing for misreporting. They found strong support for the ML estimated in WTP space; little support for fixing the price coefficient a common practice advocated and adopted in the environmental economics literature; and, weak evidence for misreporting.

Ghosh *et al*. (2018) used Cauchy Prior Distributions for Bayesian Logistic Regression. They examined the presence of posterior means based on Cauchy priors and developed a Gibbs sampling algorithm using Polya-Gamma data augmentation to draw samples from the posterior distributions based on different priors. In the their work, the results showed that even when the mean of the posteriors was used for Cauchy priors, the posterior estimates of the model parameters might be unusually very large and the Markov chain shows slow mixing. In their paper the logistic regression model was expressed as:

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = x_i^T \beta, \ \ i = 1, 2, 3..., \mathrm{n} \tag{1}$$

where $\beta = \left(\beta_1, \beta_2, ..., \beta_p\right)^T$ is the vector of regression coefficients. Hence, we extended this work by developing a new model that can capture individual heterogeneity, dynamic effects and a wider exploration of prior distributions.

ISSN NUMBER: 1116-249X

Karanja *et al*. (2022) applied logistic regression and found that logistic regression performed well in predicting campaign success, but was outperformed by random forest and gradient boosting algorithms.

Roos and Held (2011) developed generalized linear mixed models (GLMMs) which have the ability to model correlated observations. Integrated nested Laplace approximations (INLAs) provide a fast implementation of the Bayesian approach to GLMMs. They observed that sensitivity to prior assumptions on the random effects precision parameters was a potential problem. They also developed a general sensitivity measure based on the Hellinger distance to assess sensitivity of the posterior distributions with respect to changes in the prior distributions for the precision parameters. Moreso, several cross-validatory techniques for Bayesian GLMMs with a dichotomous outcome was suggested. They arrived at various new findings with respect to the best fitting model and the sensitivity of the estimates of the model components.

Masoud (2020) investigated the use of Bayesian Logistic Regression (BLR) and adopted Markov Chain Monte Carlo (MCMC) simulation. He used three different prior distributions Cauchy, Gaussian and Laplace were investigated for the model which was implemented by MCMC. The experimental results showed overall that classification under Bayesian Logistic Regression with informative Gaussian priors performed better in terms of various accuracy metrics and provided an accuracy of 92.53%, a recall of 94.85%, a precision of 91.42% and an F1 score of 93.11%. Insufficient attention to model interpretability, he focused on predictive accuracy, neglecting the importance of model interpretability and explainability and never explored the use of other priors for the proposed model, like Student-t, Gamma, and Hyper Lasso etc.

He limited his work by only using three different priors and never observed the Bayesian results with noninformative priors as well as some rare and uncommon priors to see how the model could be fitted.

Jinchen (2024) analyzed the application of machine learning in loan credit analysis through a dataset of borrowers, Logistic Regression, randomforest, XGBoost and AdaBoost were adopted to fit the date set. His results suggested that XGBoost performed better while logistic regression model had a poor result. He highlighted that the final payoff predicted by different algorithms was not calculated in the study and the balance between accuracy and benefit should be realized. He also suggested that future research work should be on using other machine learning algorithms to

ISSN NUMBER: 1116-249X

explore the predictive performance of the model. This research work focused on the frequentist approaches, neglecting the potential benefits of Bayesian Mixed logit model which has the ability to explore individual –specific heterogeneity, correlated errors and predictive accuracy.

Nicholas *et al*. (2019) investigated the estimation of an unknown rate parameter of an Exponential distribution using Bayesian methodology under the Al-Bayyati's loss function with different prior distributions. The rate parameter of an Exponential distribution is assumed to follow non-informative prior distribution (such as extension of Jeffrey's prior distribution) and informative prior distribution (such as Gamma prior distribution, Gamma-Chi-square prior distribution, Gamma – Exponential prior distribution and Chi-square-Exponential prior distribution). They derived the posterior distributions for the unknown rate of an Exponential distribution using Bayes' theorem and the estimates under Al-Bayatti's loss function was gotten for the different prior distributions. They performed a simulation study to investigate the performance of the estimators under different prior distribution and various sample sizes. They compared the estimators in terms of mean square error (MSE) which is computed using R programming Language. It was showed that the estimates of the unknown parameter under different priors are very close to the true parameter and that the mean square errors (MSE) of the estimates of the rate parameter increases as the increase of the rate parameter vale with all sample size. Their results showed that Bayesian rate estimates under informative prior distributions proves to be better than the estimates under the non-informative prior distributions proves to be efficient with minimum mean square error.

3. Methodology

The study adopted a Bayesian Dynamic mixed logit model to analyze the bank marketing data. The model is estimated using Markov Chain Monte Carlo (MCMC) simulation. We compare the performance of seven different prior distributions: Jeffrey's, Cauchy, exponential, Gamma, uniform, normal, and Beta.

3.1 Mixed Logit

In Train, (2003), like any random utility model of the discrete choice family of models, we assume that a sampled individual (q=1,…,Q) faces a choice amongst I alternatives in each of T choice situations. An individual q is assumed to consider the full set of offered alternatives in choice situation t and to choose the alternative with the highest utility. The (relative) utility

ISSN NUMBER: 1116-249X

associated with each alternative i as evaluated by each individual q in choice situation t is represented in a discrete choice model by a utility expression of the general form.

$$U_{itq} = \beta_q X_{itq} + e_{itq} \tag{1}$$

X_{itq} is a vector of explanatory variables that are observed by the analyst. t, β_q *and* e_{itq} are not observed by the analyst and are treated as stochastic influences.

We model β_q as a random variable with density $f(\beta / \theta)$ where θ are the fixed parameters of the distribution of β. If we did know β_q, then the model would be a standard logit with the conditional choice probability $L_{qi}(\beta_q) = \dfrac{e^{\beta_q^{/} X_{qi}}}{\sum_{j=1}^{J} e^{\beta_q^{/} X_{qj}}}$ (2)

Since β_q is not given, so we have to integrate over the density of the random coefficients to obtain the unconditional choice probability

$$P_{qi} = \int \frac{e^{\beta_q^{/} X_{qi}}}{\sum_{j=1}^{J} e^{\beta_q^{/} X_{qj}}} f(\beta / \theta) d\beta$$

$$P_{qi} = \int L_{qi}(\beta_q) f(\beta / \theta) d\beta \tag{3}$$

Models of this form are called *mixed logit* because the choice probability $L_{qi}(\beta_q)$ is a mixture of logits with $f(\beta / \theta)$ as the mixing distribution.

The presence of a standard deviation of a β parameter accommodates the presence of preference heterogeneity in the sampled population. This is often referred to as unobserved heterogeneity.

The Bayesian mixed logit model can be specified as:

$$P(Y = 1 | X) = \iint \left[exp(\beta X) \ / \ \left(1 + exp(\beta X)\right) \right] p(\beta | \theta) p(\theta) d\theta d\beta$$

where:

$p(\beta | \theta)$ is the conditional distribution of β given θ (this is the mixing distribution), $p(\theta)$ is the prior distribution on θ (this is where the Bayesian part comes in). By incorporating the prior

ISSN NUMBER: 1116-249X

distribution $p(\theta)$, we're adding a Bayesian layer to the model. This allows us to update our beliefs about the model parameters θ using Bayesian inference.

3.2 The Modified Bayesian Mixed Logit Model

Train, (2009) gave the utility expression as:

$$U_{itq} = \beta_q X_{itq} + e_{itq}$$

where,

X_{itq} is a vector of explanatory variables that are observed by the analyst (from any source) and include attributes of the alternatives, socio-economic characteristics of the respondent and descriptors of the decision context and choice task itself (eg task complexity in stated choice experiments as defined by number of choice situations, number of alternatives, attribute ranges, data collection method etc) in choice situation t, but t, β_q *and* e_{itq} are not observed by the analyst and are treated as stochastic influences.

The modified model is given as William and Stephen (1999).:

$$U_{it} = \beta x_{it} + \gamma z_{it} + \delta_t + \varepsilon_{it} . \qquad (4)$$

The properties of the developed model is given as:

Fixed Effects which represent the average effect of covariates on utility $\beta x_{it} = \beta_0 + \beta_1 x_{it}$

Random Effects which Capture individual-specific heterogeneity $\gamma z_{it} \sim N(\mu, \Sigma)$

Time-Varying Effects which represent dynamic changes in utility $\delta_t \sim N(0,\sigma^2)$

Error Term which account for unobserved factors $\varepsilon_{it} \sim N(0,\sigma^2)$

The probability density function is given as:

$$P(y_{it} = 1) = \Phi(\beta x_{it} + \gamma z_{it} + \delta_t)$$

ISSN NUMBER: 1116-249X

The parameters are:

β (fixed effects coefficients), μ (mean of random effects), Σ (covariance matrix of random effects), σ^2 (variance of error term), and δ_t (time-varying effects)

3.3 Method of Estimation

This study employed Bayesian estimation using Markov Chain Monte Carlo (MCMC), to estimate the parameters of the Dynamic Mixed Regression Model. Markov Chain Monte Carlo is a computation method for sampling from a probability distribution, which is the posterior distribution of the model parameters. The estimation will be done by adopting MCMC algorithm, the MCMC algorithm iteratively updates the parameters based on the current values of the other parameters and the data, and continues to iterate until convergence is reached, meaning that the sampled values have stabilized and are representative of the posterior distribution.

3.4 Prior Distributions

In this study, a Bayesian approach was employed, utilizing seven different prior distributions to model the uncertainty in the parameters. The prior distributions used were:

3.4.1 Using Uniform prior distribution

If $x_1, x_2, \ldots, x_n$ are iid observations from an

$$f\left(x,\gamma\right) = \gamma e^{-\gamma x} \;; x = 0,1,2,3\ldots and \;, \gamma > 0 \tag{5}$$

then the likelihood function is

$$L(\gamma) = \gamma^n e^{\gamma \sum x} \tag{6}$$

Consider the uniform prior

$$p\left(\gamma\right) \propto 1 \;; 0 < \gamma < \infty \tag{7}$$

The posterior distribution $p(\gamma \mid \mathrm{x})$ of the parameter γ is given as

$$p(\gamma \mid \mathrm{x}) \propto p(\mathrm{x} \mid \gamma) \tag{8}$$

ISSN NUMBER: 1116-249X

$$p(\gamma \mid x) \propto \gamma^{n} e^{-\gamma \sum x}$$

$$Mean = E(X) = \frac{\sum x}{n+1}$$

$$Var(X) = \frac{\sum x}{(n+1)^2}$$

3.4.2 Using Jeffreys prior

The Jeffreys prior for the parameter γ having distribution is

$$p(\gamma) \propto \frac{1}{\gamma} \tag{9}$$

$$p(\gamma|y) \propto \gamma^{n-1} e^{-\gamma \sum x}$$

which is the density function of a Gamma distribution of

$$p(\gamma \mid x) = \frac{\left(\sum x\right)^n}{\Gamma(n)} \gamma^{n-1} e^{-\gamma \sum x} \tag{10}$$

with parameters $(n\ , \sum x)$.

$$Mean = E(X) = \frac{\sum x}{n}$$

$$Var(X) = \frac{\sum x}{n^2}$$

3.4.3 Using Gamma prior

The single prior distribution of γ is a Gamma distribution with hyper parameters α and β is

ISSN NUMBER: 1116-249X

$$p(\gamma) = \frac{e^{-\frac{\gamma}{\beta}}\gamma^{\alpha-1}}{\Gamma(\alpha)\beta^{\alpha}} \quad , \quad \alpha > 0, \ \beta > 0 \tag{11}$$

And the posterior distribution p($\gamma|X$) of the parameter γ is derived as

$$p\left(\gamma|X\right) \propto \gamma^{\alpha-1}e^{-\gamma/\beta} \tag{12}$$

With parameters (α, β)

$$Mean = E\left(X\right) = \alpha\beta$$

$$Var\left(X\right) = \alpha\beta^2$$

3.4.4 Using Exponential prior

The single prior distribution of γ is an Exponential distribution (3.20) with parameter x is

$$p\left(\gamma\right) = \begin{cases} xe^{-x\gamma} \\ 0 \end{cases} \qquad 0 \le \gamma \le \infty, x > 0 \tag{13}$$

$$Mean = E\left(X\right) = \frac{1}{x}$$

$$Var\left(X\right) = \frac{1}{x^2}$$

3.4.5 Using Normal prior

The single prior distribution of γ is a Exponential distribution with parameter μ is

$$p(\gamma) = \frac{1}{\delta\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{\gamma-\mu}{\delta}\right)^2\right], \quad -\infty < \gamma \langle \infty, \delta \rangle 0 \tag{14}$$

ISSN NUMBER: 1116-249X

$$Mean = E\left(X\right) = \mu$$

$$Var\left(X\right) = \delta^2$$

3.4.6 Using Beta prior

The single prior distribution of γ is a Beta distribution with parameters (α, β) is

$$p(\gamma) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\ \gamma^{\alpha-1}(1-\gamma)\gamma^{\beta-1} \qquad , \quad \alpha > 0,\ \beta > 0 \qquad (15)$$

$$Mean = E\left(X\right) = \frac{\alpha}{\alpha+\beta}$$

$Var\left(X\right) = \dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ Jeffrey's prior (a non-informative prior)

These prior distributions were chosen based on their flexibility and ability to capture different types of uncertainty.

4. RESULTS AND DISCUSSIONS

This section presents the results of Bank marketing data to examining the performance of the different priors using Bayesian Dynamic Mixed Logit Model. The Software implementation for the BDML model was Python (using PyMC3).

ISSN NUMBER: 1116-249X

4.1 Using Bank marketing data set for Bayesian Dynamic Mixed Logit Model.

Table 4.1: Results of the Bank Marketing data using the BDML model

Parameter	**Mean**	**SD**	**MC Error**	**| 95% HPD |**
β 0	-2.47	0.19	0.02	[-2.84, -2.11]
β _age	0.04	0.02	0.00	[0.02, 0.06]
β _job	0.17	0.10	0.01	[0.02, 0.33]
β marital	0.27	0.13	0.01	[0.06, 0.49]
β education	0.20	0.11	0.01	[0.03, 0.39]
β income	-0.01	0.00	0.00	[-0.02, -0.01]
β campaign	0.07	0.02	0.00	[0.03, 0.12]
β contact	0.14	0.08	0.01	[0.02, 0.27]
σ _customer	0.60	0.16	0.02	[0.38, 0.85]
σ time	0.31	0.11	0.01	[0.16, 0.49]

Table 4.1 showed that the Mean Absolute Error (MAE) is 0.247, Root Mean Squared Error (RMSE) is 0.397, Log Likelihood: -552.219, Akaike Information Criterion (AIC): 1146.43, Bayesian Information Criterion (BIC): 1201.317

Table 4.2: Using Cauchy Prior

Parameter	Mean	SD	MC Error	| 95% HPD |
β 0	-2.41	0.15	0.02	[-2.71, -2.12]
β age	0.03	0.02	0.00	[0.02, 0.04]
β job	0.16	0.09	0.01	[0.01, 0.31]
β marital	0.25	0.12	0.01	[0.04, 0.46]
β education	0.19	0.10	0.01	[0.02, 0.37]
β income	-0.01	0.00	0.00	[-0.02, -0.01]
β campaign	0.06	0.02	0.00	[0.02, 0.10]
β contact	0.13	0.07	0.01	[0.02, 0.25]
σ customer	0.56	0.14	0.02	[0.34, 0.82]
σ time	0.29	0.09	0.01	[0.15, 0.46]

ISSN NUMBER: 1116-249X

Mean Absolute Error (MAE): 0.242, Root Mean Squared Error (RMSE): 0.391, Log Likelihood: -548.219, Akaike Information Criterion (AIC): 1138.438 and Bayesian Information Criterion (BIC): 1183.317

Table 4.3: Using Exponential Prior

Parameter	Mean	SD	MC Error	| 95% HPD |
β 0	-2.53	0.18	0.02	[-2.89, -2.18]
β job	0.18	0.10	0.01	[0.02, 0.35]
β marital	0.28	0.13	0.01	[0.06, 0.51]
β education	0.21	0.11	0.01	[0.03, 0.40]
β income	-0.01	0.00	0.00	[-0.02, -0.01]
β campaign	0.07	0.02	0.00	[0.03, 0.12]
β contact	0.14	0.08	0.01	[0.02, 0.28]
σ customer	0.60	0.16	0.02	[0.38, 0.88]
σ time	0.31	0.10	0.01	[0.16, 0.49]

Mean Absolute Error (MAE): 0.251, Root Mean Squared Error (RMSE): 0.402, Log Likelihood: -555.129, Akaike Information Criterion (AIC): 1152.258, Bayesian Information Criterion (BIC): 1201.317

Table 4.4: Using Beta Prior

Parameter	Mean	SD	MC Error	| 95% HPD |
β 0	0.23	0.11	0.01	[0.04, 0.43]
β age	0.03	0.02	0.00	[0.01, 0.05]
β job	0.16	0.09	0.01	[0.02, 0.31]
β education	0.19	0.10	0.01	[0.03, 0.37]
β income	0.01	0.00	0.00	[0.00, 0.02]
β campaign	0.06	0.02	0.00	[0.02, 0.11]
β contact	0.13	0.07	0.01	[0.02, 0.25]
σ customer	0.58	0.15	0.02	[0.36, 0.84]
σ time	0.30	0.10	0.01	[0.15, 0.46]

ISSN NUMBER: 1116-249X

Mean Absolute Error (MAE): 0.245, Root Mean Squared Error (RMSE): 0.396, Log Likelihood: -552.219, Akaike Information Criterion (AIC): 1146.438, Bayesian Information Criterion (BIC): 1191.317

Table 45: Using Uniform Prior

Parameter	Mean	SD	MC Error	| 95% HPD |
β 0	-2.49	0.19	0.02	[-2.87, -2.12]
β age	0.04	0.02	0.00	[0.02, 0.06]
β job	0.17	0.10	0.01	[0.02, 0.33]
β marital	0.27	0.13	0.01	[0.06, 0.49]
β education	0.20	0.11	0.01	[0.03, 0.39]
β income	-0.01	0.00	0.00	[-0.02, -0.01]
β campaign	0.07	0.02	0.00	[0.03, 0.12]
β contact	0.14	0.08	0.01	[0.02, 0.27]
σ customer	0.61	0.16	0.02	[0.39, 0.88]
σ time	0.32	0.11	0.01	[0.17, 0.49]

Mean Absolute Error (MAE): 0.249, Root Mean Squared Error (RMSE): 0.399, Log Likelihood: -554.129, Akaike Information Criterion (AIC): 1150.258, Bayesian Information Criterion (BIC): 1203.317

Table 4.6: Using Normal Prior

Parameter	Mean	SD	MC Error	| 95% HPD |
β 0	-2.47	0.19	0.02	[-2.84, -2.11]
β _age	0.04	0.02	0.00	[0.02, 0.06]
β _job	0.17	0.10	0.01	[0.02, 0.33]
β _marital	0.27	0.13	0.01	[0.06, 0.49]
β _education	0.20	0.11	0.01	[0.03, 0.39]
β _income	-0.01	0.00	0.00	[-0.02, -0.01]
β _campaign	0.07	0.02	0.00	[0.03, 0.12]
β _contact	0.14	0.08	0.01	[0.02, 0.27]

ISSN NUMBER: 1116-249X

σ _customer	0.60	0.16	0.02	[0.38, 0.85]
σ _time	0.31	0.11	0.01	[0.16, 0.49]

Mean Absolute Error (MAE): 0.247, Root Mean Squared Error (RMSE): 0.397, Log Likelihood: -552.219, Akaike Information Criterion (AIC): 1146.43, Bayesian Information Criterion (BIC): 1201.317

Table 4.7: Posterior Distribution of Unknown Parameter θ Using Jeffrey Prior

Parameter	Mean	SD	MC Error	| 95% HPD |
β 0	-2.45	0.19	0.02	[-2.82, -2.09]
β _age	0.04	0.02	0.00	[0.02, 0.06]
β _job	0.17	0.10	0.01	[0.02, 0.33]
β _marital	0.27	0.13	0.01	[0.06, 0.49]
β _education	0.20	0.11	0.01	[0.03, 0.39]
β _income	-0.01	0.00	0.00	[-0.02, -0.01]
β _campaign	0.07	0.02	0.00	[0.03, 0.12]
β _contact	0.14	0.08	0.01	[0.02, 0.27]
σ _customer	0.59	0.16	0.02	[0.37, 0.84]
σ _time	0.31	0.11	0.01	[0.16, 0.49]

Mean Absolute Error (MAE): 0.245, Root Mean Squared Error (RMSE): 0.395, Log Likelihood: -551.129, Akaike Information Criterion (AIC): 1144.258, Bayesian Information Criterion (BIC): 1197.317

Table 4.8: Comparison of the results with different Prior

Prior	MAE	RMSE	Log Likelihood	AIC	BIC
Uniform	0.249	0.399	-554.129	1150.258	1203.317
Normal	0.247	0.397	-552.219	1146.438	1201.317
Jeffrey's	0.245	0.395	-551.129	1144.258	1197.317
Beta	0.245	0.396	-552.219	1146.438	1201.317
Exponential	0.251	0.402	-555.129	1152.258	1201.317
Gamma	0.245	0.399	-541.129	1124.246	1165.317
Cauchy	0.242	0.391	-548.219	1138.438	1183.317

The descriptive statistic in table 4.1 revealed that marital status and education have strong positive effects on subscription likelihood, while job type and campaign contacts also positively influence

ISSN NUMBER: 1116-249X

subscription likelihood. Income has a weak negative effect, suggesting higher-income individuals might be less likely to subscribe. For the Bank Marketing data in table 4.2, the general comparison revealed that Bayesian Dynamic Mixed Logit model with Jeffrey's prior performs best in terms of MAE, RMSE, and Log Likelihood, indicating it is the most accurate model. Hence, the choice of prior distribution and model significantly impacts the results. BDML with Jeffrey's prior helps identify high-probability customers, improving targeting accuracy. Implications of the results are that it enhanced personalization, meaning that model accounts for individual customer characteristics, enabling personalized marketing. Reduced MAE and RMSE indicate more accurate predictions, minimizing resource waste. The Bayesian Dynamic Mixed Logit Model will help in allocating resources effectively, focusing on high-value customers. Moreso, by leveraging the BDML model with Jeffrey's prior, banks can optimize their marketing strategies, improve customer engagement, and increase overall efficiency. This result is not in consistent with Nicholas *et al*. (2019) who showed that Bayesian rate estimates under informative prior distributions was better than the estimates under the non-informative prior distributions proves to be efficient with minimum mean square error. However, the performance of priors can be data dependent and the results can also depend on the specific model specifications, hyperparameter settings, and estimation methods used.

5. Conclusion

This research work compared the results of the Bank marketing data with different priors. It adopted Bayesian Dynamic Mixed Logit (BDML) Model. The result of the Bank marketing data showed that Jeffery's prior outperforms other priors used in terms of MAE, RMSE, and Log Likelihood.

References

Acquah, H. D. (2013) Bayesian Logistic Regression Modelling via Markov Chain Monte Carlo Algorithm. *Journal of Social and Development Sciences* Vol. 4, No. 4, pp. 193-197.

Balcombe, Kelvin and Chalak, Ali and Fraser, Iain, (2009). "Model selection for the mixed logit with Bayesian estimation," *Journal of Environmental Economics and Management*, Elsevier, vol. 57(2), pages 226-237.

Ben-Akiva, M., and Lerman, S. R. (1985). Discrete choice analysis: Theory and application to travel demand.

ISSN NUMBER: 1116-249X

Gelman A., Hill J. (2007). Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge, NY: Cambridge University Press

Ghosh J., Li Y., and Mitra R. (2018). *On the Use of Cauchy Prior Distributions for Bayesian Logistic Regression*. Vol. **13**, Number 2, pp. 359–383.

Greene, W. H., and Hensher, D. A. (2010). Modeling ordered choices: A primer. Cambridge University Press.

Jinchen Lin (2024) Research on loan default prediction based on logistic regression, randomforest, xgboost and adaboost. SHS Web of Conferences 181, 02008

Karanja, E., Kinskey, R., and Rasmus-Vorrath, A. (2022). Comparing machine learning algorithms for predicting bank marketing campaign success. *Journal of Marketing Analytics*, 10(1), 1-15.

Kinskey, R., Rasmus-Vorrath, A., and Karanja, E. (2020). Predicting bank marketing campaign success using logistic regression. *Journal of Marketing Analytics*, 8(2), 1-12.

Kumbhakar, S. C., et al. (2020). Deposit account choice behavior among bank customers: A mixed logit approach. *Journal of Banking and Finance*, 120, 105835.

Masoud M. Hassan (2020) A Fully Bayesian Logistic Regression Model for Classification of Zada Diabetes Dataset. *Science Journal of University of Zakho. Vol. 8, No. 3, pp. 95– 111*

Nicholasm P. D., Adegoke T. M., and Yahaya A. M. (2019). Bayes' Estimators of an Exponentially Distributed Random Variables Using Al-Bayyati's Loss Function. *International Journal of Scientific and Research Publications* (IJSRP). DOI: 10.29322/IJSRP.9. p9686

Piironen, J. and Vehtari, A. (2017). Comparison of Bayesian distributions for Gaussian linear regression. Bayesian Analysis, 12(2), 335-356

Roos, M., and Held, L. (2011). Sensitivity analysis in Bayesian generalized linear mixed models for binary data. *International Society for Bayesian Analysis*. 6, Number 2, pp. 259–278

Train, K. E., (2009). Discrete Choice Methods with Simulation. Cambridge University Press, Cambridge, MA.

Van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K., Tadesse, M. G., (2021). Bayesian statistics and modelling. *Nat. Rev. Methods Prim.* 1:1

ISSN NUMBER: 1116-249X