FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 9 No. 5, May, 2025, pp 101 - 105
DOI: https://doi.org/10.33003/fjs-2025-0905-3654

APPLYING BAYESIAN DYNAMIC MIXED LOGISTIC REGRESSION TO MOBILITY NETWORKS

***[1]Amalahu, Christian Chinenye, [2]Nwabueze, Joy Chioma,**
[2]Enogwe, Samuel Ugochukwu and [1]Ekeadinotu, Chibueze Barnabas

[1]Department of Mathematics, University of Agriculture and Environmental Sciences, Umuagwo, Imo State, Nigeria.
[2]Department of Statistics, Michael Okpara University of Agriculture, Umudike, Abia State, Nigeria

*Corresponding authors' email: chinenyechris2017@gmail.com

ABSTRACT

This study explored the performance of Bayesian Dynamic Mixed Logistic Regression Model (BDML) with different priors that include; Beta, Gamma, Cauchy, Exponential, Normal, Jeffrey and Uniform prior. The primary objective of the model was to capture time-varying random intercepts and slopes while accommodating dynamic data structure. The major aim of this research was to compare the BDML model with alternative models including the Bayesian Mixed Logit model, mixed logit and logistic regression and to evaluate their performance. Simulated transportation data revealed that the DBML model outperformed other models; with the modified Bayesian Dynamic Mixed Logit (BDML) model achieving the highest accuracy (81.5%) and lowest AIC/BIC values, indicating superior performance. The log likelihood for BDML is (-1534.2), Bayesian Mixed Logit (BML) is -1541.1 and Mixed Logit (ML) is given as -1551.9 BDML model's best fit the data. The implications are that travel time and cost are significant factors in mode choice. The study recommended investments in comfortable and eco-friendly transportation and encourages bike usage through infrastructure development like good roads.

Keywords: Bayesian, Dynamic, Mixed Logistic Regression

INTRODUCTION

The Logistic regression models regression is a widely used statistical technique in data analysis, particularly in modeling binary outcomes. It is a type of generalized linear model that estimates the probability of a binary response variable based on one or more predictor variables. Logistic regression has been extensively applied in various fields, including health care, finance, and education (Joshi & Dhakal, 2021), due to its ability to provide insights into the relationships between predictor variables and binary outcomes.

However, logistic regression has several limitations. One major limitation is its assumption of linearity in the logit, which may not always hold true (Boateng & Abaye, 2019). Additionally, logistic regression can be sensitive to outliers and multicollinearity, which can lead to biased estimates and inaccurate predictions Li *et al*., (2021). Furthermore, logistic regression assumes that the predictor variables are measured without error, which is often not the case in real-world applications.

Nemes *et al*. (2009) demonstrated in their work that; as sample size increases, the size of bias in Logistic regression parameter estimates will approach zero. The following equation based on additive definition of the bias is given as:

$$\hat{\beta} = \beta_{pop} + \frac{b_1(\beta)}{n} \quad (1)$$

as the sample size increases $n \to \infty$, the bias converges to 0 $(lim_{n\to\infty} b_1\,(\beta)n^{-1} = 0)$.

The study shows that when the sample size is small, inferences based on the Logistic regression model's estimates could not be reliable and misleading. However, its limitations include restrictive assumptions of independence, identical preferences across individuals and inability to capture complex relationships. These limitations led researchers to seek for more advanced models like Mixed logit models. The Mixed Logit model extends logistic regression by allowing random coefficient to capture individual heterogeneity and preference variation, and also accommodating correlated choices.

Bayesian Statistics is widely used in the literature for different statistical analysis like in choice modeling. Transportation mode choice modeling has evolved significantly over the years. Early studies employed logistic regression Ben-Akiva and Lerman, (1985) and later, mixed logit models Munizaga (2000), became popular. 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.

Some studies have addressed these limitations using advanced models such as incorporating temporal dependencies using dynamic models (e.g., dynamic logit, Kalman filter) Bhat, (2005). Bayesian estimation has also been applied to mixed logit models to account for heterogeneity Train and Sonnier, (2005) while some researchers explored machine learning techniques like neural networks, random forests for mode choice modeling Majbah *et al*., (2021). Despite these advancements, gaps have been created such as inadequate handling of heterogeneity; Bayesian methods may not fully capture individual-specific preferences. Limited consideration of temporal dependence, existing dynamic models often rely on simplistic assumptions and lack of integration. Dynamic (time –varying effects) and Bayesian approaches are rarely combined. This study addresses the research gap by adopting Bayesian Dynamic Mixed Logit (BDML) model that integrates temporal dependence which captures dynamic relationships between variables. Accounting for individual-specific preferences (heterogeneity) and finally, incorporating uncertainty and prior knowledge (Bayesian estimation).

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:

$$log\left(\frac{\pi_i}{1-\pi_i}\right) = x_i^T\beta, \quad i = 1,2,3\ldots, n \qquad (2)$$

where $\beta = \left(\beta_1, \beta_2, \ldots, \beta_p\right)^T$ 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.

Munizaga, (2000) evaluated of mixed logit as a practical modeling alternative. He presented two numerical applications; one was based on simulation study and the other one with real data set. It was discovered that similar taste parameters' rations within models and strange results were found for the correlation parameters. He defined his model as:

$$U_{in} = V_{in} + \eta_{in} + \varepsilon_{in} \qquad (3)$$

Where $\varepsilon_{in} \sim Gumbel(0, \lambda)$ and $\eta_{in} \sim f(\eta/\theta)$, where f is a general density function and θ are fixed parameters that describe it mean and variance, ε_{in} is iid Gumbel and V_{in} is a deterministic components.

Liu and Cirillo (2020) applied a Bayesian mixed logit model to investigate travel mode choice behavior in the context of autonomous vehicles. The results showed that individuals' preferences for autonomous vehicles varied significantly. The Bayesian mixed logit model was specified as:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \beta_2 Z_{ij} + \varepsilon_{ij} \qquad (4)$$

where Y_{ij} denote the choice of travel mode (autonomous vehicle or traditional vehicle) for individual i on trip j, X_{ij} represents the attributes of the autonomous vehicle, Z_{ij} represents the attributes of the traditional vehicle, and ε_{ij} represents the error term. The results showed that the posterior mean of β_1 was 0.85 (95% CI: 0.56-1.14), indicating a positive preference for autonomous vehicles.

Zhu and Levinson (2020) used a Bayesian mixed logit model to analyze route choice behavior in the presence of traffic information. The results showed that the posterior mean of β_2 was -0.67 (95% CI: -1.03--0.31), indicating a negative preference for routes with heavy traffic. Their results indicated that travelers' preferences for routes varied significantly based on traffic conditions.

Yan *et al*. (2021) proposed a new comprehensive travel impedance model to dynamically analyze the accessibility of freeway entrances and exits. The dynamic accessibility of freeway entrances and exits in Zhengzhou was studied using the proposed comprehensive impedance model, and the calculation results were analyzed. The accessibility of freeway entrances and exits is characterized by dynamic changes; the accessibility during the off-peak evening period is the highest, while that during the morning peak period and evening peak period is lower. Their results of the comprehensive impedance model were roughly consistent with reality. From a location perspective, regardless of the period of time, the accessibility of freeway entrances and exits in the central and surrounding areas of Zhengzhou was always at a lower level, and during the off-peak afternoon period, the accessibility of the eastern part of the city is notably higher than that of the western part. Moreso, the accessibility of freeway entrances and exits is closely related to the traffic status of the road network and the characteristics of regional land use. The implication is that it can provide feedback for planning road networks and provide a reference for road network planning and traffic facility design.

Liu *et al*., (2022), studied the impacts of autonomous vehicles on mode choice behavior in the context of short and medium-distance intercity travel. They developed a structure equation model, integrated it into a random parameter Logit model and also developed a hybrid choice model. Wuhan was used as a case to carry out an empirical study, and the study results revealed that: the utility function, the coefficients of three variables, including in-vehicle time, access and exit and waiting time, and travel cost, are not fixed but follow a normal distribution with a mean of -0.014, -0.008, and -0.010 and with the standard deviations of 0.014, 0.021, and 0.017, respectively. The study revealed that travelers have heterogeneous preferences toward the attributes of the transport service offered by autonomous vehicles, such as in-vehicle time, access/egress and waiting time, and travel costs. It is also found that perceived behavioral control and behavioral attitudes have significantly positive impacts on traveler's choice on autonomous vehicles. Therefore, reducing travel costs and travel time of autonomous vehicles can increase the attractiveness of autonomous vehicles.

MATERIALS AND METHODS

The method adopted in this study is the Bayesian Dynamic Mixed Logistic Regression Model. 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 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.

$$U_{itq} = \beta_q X_{itq} + e_{itq} \qquad (5)$$

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

$$L_{qi}(\beta_q) = \frac{e^{\beta_q^{/} X_{qi}}}{\sum_{j=1}^{J} e^{\beta_q^{/} X_{qj}}} \qquad (6)$$

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

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

$$P_{qi} = \int L_{qi}(\beta_q)\, f(\beta/\theta) d\beta \qquad (7)$$

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:

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

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 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.

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 Train, (2009).:

$U_{it} = \beta x_{it} + \gamma z_{it} + \delta_t + \varepsilon_{it}$. (8)

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 $\varepsilon_{it} \sim N(0, \sigma^2)$

$P(y_{it} = 1) = \Phi(\beta x_{it} + \gamma z_{it} + \delta_t)$The probability density function is given as:

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)

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.

RESULTS AND DISCUSSION

This section presents the results of simulation data to examining the performance of the Bayesian Dynamic Mixed Logit Model. We compared the results of the developed model with alternative models such as Bayesian Mixed Logit Model, Mixed Logit Model. The Software implementation for the BDML model was Python (using PyMC3).

Using a data set of 50 commuters, the simulation analysis employed the Bayesian Dynamic Mixed Logit Model

The summary of the simulation is Number of individuals: 50, Number of choices per individual: 5, Modes of transportation: 4 (Car, Bus, Train, Bike). Attributes: Travel Time (minutes), Cost ($), Comfort (scale: 1-5), Environmental Impact (scale: 1-5). Attribute Values - Travel Time: Car: 20-40 minutes, Bus: 30-50 minutes, Train: 15-30 minutes, Bike: 10-20 minutes, showing the impact of travel time, cost, comfort, and environmental impact on mode choice for commuters in urban centers.

Table 1: Results of the simulated transportation data using the BDML model

Parameter	Mean	SD	MCSE	2.5%	97.5%
Travel Time	-0.093	0.021	0.002	-0.134	-0.052
Cost	-0.211	0.041	0.004	-0.291	-0.131
Comfort	0.301	0.051	0.005	0.201	0.401
Environmental Impact	0.251	0.061	0.006	0.131	0.371

Table 2: Results of the simulated transportation data using the Bayesian Mixed Logit Model

Parameter	Mean	SD	MCSE	2.5%	97.5%
Travel Time	-0.091	0.022	0.002	-0.133	-0.049
Cost	-0.208	0.042	0.004	-0.289	-0.127
Comfort	0.294	0.052	0.005	0.192	0.396
Environmental Impact	0.246	0.062	0.006	0.125	0.367

Log Likelihood (-1541.10, DIC 3125.9, WAIC (3142.1), AIC (3083.1), BIC (3133.1)

Table 3: Results of the simulated transportation data using the Mixed Logit Model

Parameter	Mean	SD	MCSE	2.5%	97.5%
Travel Time	-0.095	0.023	0.002	-0.139	-0.051
Cost	-0.219	0.043	0.004	-0.303	-0.135
Comfort	0.313	0.053	0.005	0.209	0.417
Environmental Impact	0.259	0.063	0.006	0.135	0.383

Log Likelihood (-1551.9), DIC (3141.9), WAIC (3161.1), AIC (3089.1), BIC (3139.1)

Table 4: Results of the simulated transportation data using the Logistics Regression model

Parameter	Estimate	Std. Error	z-value	Pr(>|z|)
Travel Time	-0.104	0.025	-4.16	<0.001
Cost	-0.241	0.051	-4.71	<0.001
Comfort	0.351	0.061	5.75	<0.001
Environmental Impact	0.281	0.071	3.95	<0.001
Constant	2.191	0.421	5.20	<0.001

Null Deviance (2011.1), Residual Deviance (1456.9), AIC (1466.9), BIC (1511.9)

Table 5: Comparison of the Results of the simulated transportation data using the Bayesian Dynamic Mixed Logit Model, with Bayesian Mixed Logit model, and Mixed Logit

Model	Log Likelihood	DIC	WAIC	AIC	BIC
ML	-1551.9	3141.9	3161.1	3089.1	3139.1
DBML	-1534.2	3113.1	3125.3	3071.1	3121.1
BML	-1541.1	3125.9	3142.1	3083.1	3133.1

Table 6: Log-likelihood of values of prior distribution of the simulated transportation data

Prior Distribution	BDML	BML	ML
Gamma	-141.91	-145.67	-151.61
Jeffrey	-143.19	-147.15	-153.09
Exponential	-145.67	-149.61	-155.57
Cauchy	-148.15	-152.09	-158.05
Uniform	-151.61	-155.57	-161.53
Beta	-144.51	-148.09	-154.05
Normal	-142.81	-146.79	-152.79

The table 1 above showed that the coefficient for travel time (-0.093) suggests that longer travel times decrease the probability of choosing a mode. This finding is consistent with Liu and Cirillo (2020), who found that travel time is a significant factor in mode choice decisions. A study by Zhu and Levinson (2020) found that the impact of travel time on mode choice can vary depending on the context, such as the type of transportation mode or the purpose of the trip.

The coefficient for cost (-0.211) suggests that higher costs decrease the probability of choosing a mode. The coefficient for comfort (0.301) suggests that higher comfort increases the probability of choosing a mode. The mean values decrease as the lag increases, indicating temporal dependence. The standard deviation of the random effects (sigma) indicates substantial heterogeneity across individuals; this means that individuals' preferences and behaviors become more varied and less predictable as time passes. In table 5 above, the log likelihood (-1534.2) indicates that the model's fit to the data. The implications are that travel time and cost are significant factors in mode choice, emphasizing the need for efficient and affordable transportation options while comfort and environmental impact play crucial roles in mode choice, indicating investments in comfortable and eco-friendly transportation. This study therefore encourages bike usage through infrastructure development like good roads.

It was also revealed that the BDML model has the best fit to the data (highest log likelihood, lowest DIC and WAIC). The ML model estimates slightly larger effects for cost and comfort. The ML model estimates larger random effects variance for cost. The implications are that travel time and cost are significant factors in mode choice, emphasizing the need for efficient and affordable transportation options, Comfort and environmental impact play crucial roles in mode choice, suggesting investments in comfortable and eco-friendly transportation, and encouraging bike usage through infrastructure development and incentives can reduce congestion and environmental impact. Our findings are consistent with existing literature Liu *et al.*, (2022), which highlights the importance of reducing travel costs and travel time.

The comparison of different priors shows that the Gamma prior provides a more accurate estimation of the model parameters, resulting in better predictive performance. The Gamma prior is particularly effective in capturing the complex relationships between various factors and mobility network performance.

The results of our study have important implications for transportation planning and policy decisions. The use of Bayesian dynamic mixed logit model with Gamma prior can provide a more accurate estimation of mobility network performance, allowing for more effective strategies to improve mobility and reduce congestion.

CONCLUSION

This research work Modelled Mobility Networks with Bayesian Dynamic Mixed Logistic Regression using simulated data. The results of the Bayesian dynamic mixed regression model indicate that: Travel time has a significant negative impact on mode choice decisions, suggesting that longer travel times decrease the probability of choosing a particular mode. Cost also has a significant negative impact on mode choice decisions, indicating that higher costs decrease the probability of choosing a particular mode. Comfort, on the other hand, has a significant positive impact on mode choice decisions, suggesting that higher comfort levels increase the probability of choosing a particular mode. The log likelihood (-1534.2) indicates the model's fit to the data. The transportation data showed that Bayesian Dynamic Mixed Logit (BDML) Model outperforms other models; the Bayesian Dynamic Mixed Logit (BDML) model achieves the highest accuracy (81.5%) and lowest AIC/BIC values, indicating superior performance. our study shows that the Bayesian dynamic mixed logit model with Gamma prior provides a more accurate estimation of mobility network performance. The results highlight the importance of prior selection in Bayesian modeling and demonstrate the effectiveness of Gamma prior in capturing complex relationships between various factors and mobility network performance. These findings suggest that transportation policymakers and planners should prioritize reducing travel times, costs, and improving comfort levels to encourage the use of sustainable transportation modes. The results of this study can inform the development of targeted policies and interventions aimed at promoting more efficient, affordable, and comfortable transportation options.

REFERENCES

Ben-Akiva, M. E., & Lerman, S. R. (1985). Discrete Choice Analysis: Theory and Application to Travel Demand. Cambridge, MA: MIT Press.

Bhat, C. R. (2005). A dynamic logit model of travel mode choice. Transportation Research Part B: Methodological, 39(7), 623-644.

Boateng, E.Y. and Abaye, D.A. (2019). A Review of the Logistic Regression Model with Emphasis on Medical Research. Journal of Data Analysis and Information Processing, 7, 190-207. https://doi.org/10.4236/jdaip.2019.74012

Ghosh J., Li Y., and Mitra R. (2018). On the Use of Cauchy Distributions for Bayesian Logistic Regression. Vol. 13, Number 2, pp. 359–383. https://doi.org/10.1214/17-BA1051

Joshi, R.D. & Dhakal, C.K. (2021). Predicting Type 2 Diabetes Using Logistic Regression and Machine Learning Approaches. *Int. J. Environ. Res. Public Health*, 18, 7346. https://doi.org/10.3390/ijerph18147346

Li, Q., Li, Z., & Zhang, Y. (2021). Robust logistic regression with outliers and multicollinearity. *Journal of Statistical Computation and Simulation*, 91(1), 1-13.

Liu, X., & Cirillo, C. (2020). Investigating travel mode choice behavior in the context of autonomous vehicles using Bayesian mixed logit models. Transportation Research Part C: Emerging Technologies, 111, 102744.

Liu Z., Song Z, Deng W., and Bao D. (2022). Impacts of Autonomous Vehicles on Mode Choice Behavior in the Context of Short- and Medium- Distance Intercity Travel[J]. *Journal of Transport Information and Safety*, 40(2): 91-97. http://dx.doi.org/10.3963/j.jssn.1674-4861.2022.02.011

Majbah U., Sabreena A., & Naveen E., (2021). Modeling Freight Mode Choice Using Machine Learning Classifiers: A Comparative Study Using the Commodity Flow Survey (CFS) Data. Transportation Planning andTechnology 24(5):543-559 http://dx.doi.org/10.1080/03081060.2021.1927306

Munizaga, A. M. (2000). Evaluation of mixed logit as a practical modeling alternative. Universidad De Chile, Casilla 228-3 Santiago, Chile

Nemes, S., Jonasson, J.M., Genell, A. *et al.* Bias in odds ratios by logistic regression modelling and sample size. *BMC Med Res Methodol* 9, 56 (2009). https://doi.org/10.1186/1471-2288-9-56

Train K. & Sonnier (2005). “Mixed Logit with Bounded Distributions of Correlated Partworths”

Train, K., (2003). Discrete Choice Methods with Simulation. Cambridge University Press, Cambridge, MA.

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

Yan, Y.; Guo, T.; Wang, D. (2021). Dynamic Accessibility Analysis of Urban Road-to-Freeway Interchanges Based on Navigation Map Paths. Sustainability 2, 13, 372. https://doi.org/10.3390/su13010372

Zhu, S., & Levinson, D. (2020). Analyzing route choice behavior in the presence of traffic information using Bayesian mixed logit models. Transportation Research Part B: Methodological, 139, 102926.