EFFICIENT MODEL SELECTION FOR HIGH-DIMENSIONAL AND HIERARCHICAL DATA USING BAYESIAN HIERARCHICAL GENERALIZED LINEAR MODELS WITH LASSO REGULARIZATION

[1]Aremu, Z.O., [2]Olayiwola, O.M., [3]Dawodu, G.A. and [4]Oshinuga, I.A.

[1]Department of Mathematical and Computing Sciences, Koladaisi University, Ibadan, Oyo State, Nigeria.
[2,3]Department of Statistics, Federal University of Agriculture, Abeokuta, Ogun state, Nigeria.
[4]Department of Mathematics, Federal University of Agriculture, Abeokuta, Ogun state, Nigeria.
Zainabolaide0808@gmail.com; +2347035084083

Abstract

High-dimensional and hierarchical data structures are increasingly prevalent in real-world applications, posing significant challenges to traditional statistical modeling. This study addresses these challenges by developing a novel Bayesian Hierarchical Generalized Linear Model (BHGLM) that incorporates Lasso regularization for variable selection and improved interpretability. The proposed model extends existing hierarchical Bayesian frameworks to binary outcomes using Bernoulli likelihood and Laplace priors, enabling automatic sparsity in the fixed effects. Comparative simulation diagnostics, bias and coverage analysis, and model selection criteria including AIC, BIC, MDL, ICOMP, WAIC, and LOO demonstrate that the proposed model achieves superior estimation accuracy, parsimony, and predictive performance while maintaining robust MCMC convergence. Although the computational cost is higher relative to traditional BHGLMs, the benefits in regularization and model selection make the proposed approach highly suitable for complex, high-dimensional, hierarchical data environments found in healthcare, genomics, education, and social sciences.

Keywords: Bayesian Hierarchical Generalized Linear Model, High Dimensional data, Information Criteria, Autocorrelation function, Regularization.

Introduction

The emerging of high dimensionality and complexity of real-life datasets pose serious challenges to traditional statistical modeling (George & Foster, 2000). Modern applications ranging from healthcare and genomics to social sciences and finance increasingly result in high-dimensional data where the number of predictors can be comparable to or even more than the size of observations. Also, many datasets possess hierarchical or nested structures, such as patients within hospitals, students within schools, or repeated measures within individuals. These two characteristics; high dimensionality and hierarchy need statistical methodologies that are both flexible and computationally efficient (Pamukçu et al., 2015).

Multilevel data or Hierarchical data are characterized by dependencies among observations that violate the assumption of independence required by classical models (Wirawati et al., 2017). Ignoring such structure can lead to biased estimates, underestimated standard errors, and

ISSN NUMBER: 1116-249X

misleading conclusions (Ebrahim et al., 2024). Multilevel or hierarchical models, also known as mixed-effects models, have been developed to address this issue by introducing random effects that capture variation at different levels of data aggregation. These models are widely used in fields like epidemiology, education, and psychology, where natural groupings exist in the data (Gelman & Hill, 2009).

Meanwhile, high-dimensional data introduce an entirely different set of challenges (Wikle, 2003). When the number of variables becomes large, especially in comparison to the number of observations, traditional models become unstable or overfit the data (Nannen, 2010). This leads to poor generalization and difficulty in identifying truly relevant predictors (Gürünlü Alma, 2013). Furthermore, the computational burden increases dramatically as the dimensionality grows, making inference and model selection more difficult (Zhong & Moraga, 2024)

To handle high-dimensionality, regularization techniques have become indispensable (Mallick et al., 2021). Regularization introduces a penalty term in the modeling objective that constrains model complexity and encourages sparsity in the parameter space. One of the most widely used techniques is the Least Absolute Shrinkage and Selection Operator (LASSO), which performs both variable selection and coefficient shrinkage by penalizing the sum of absolute values of the coefficients (Yuan & Lin, 2005). The Lasso has been particularly effective in high-dimensional settings, where it helps isolate a small subset of relevant predictors from a large pool of candidates (Fu, 2015).

However, combining hierarchical modeling with regularization poses unique methodological and computational challenges. Hierarchical models are typically estimated using iterative sampling techniques such as Markov Chain Monte Carlo (MCMC), which can be computationally expensive, especially when high-dimensional predictors and complex random effects structures are involved (Robert et al., 2018).

Bayesian methods naturally incorporate uncertainty, provide interpretable posterior distributions, and allow for flexible modeling of complex data-generating processes (Carlin et al., 2001). When combined with regularization priors, such as the Laplace prior used in the Bayesian Lasso (Park & Casella, 2008), Bayesian approaches offer a coherent way to perform sparse estimation in hierarchical settings. Bayesian hierarchical generalized linear models (BHGLMs) have emerged as powerful tools for analyzing data with complex structures, especially in a situation where there are multiple levels of variability and dependencies among the data (Günther et al., 2021). These models permit for the incorporation of pre-existing awareness about the phenomenon and hierarchical structuring of data, making them highly suitable for an extensive application for complex information (Yi et al., 2011).

Previous model selection methods, such as those based on goodness-of-fit criteria or cross-validation, may not be sufficient or efficient when dealing with high-dimensional data and

ISSN NUMBER: 1116-249X

complex model structures (Müller et al., 2013). Therefore, there is a need to seek for a robust model method that can efficiently balance model fit and complexity and ensuring the selection of an efficient model that generalizes well to the complex data.

The integration of Bayesian hierarchical modeling with regularization techniques like Lasso represents a promising path forward as the data complexity grows (Yuan & Lin, 2005). Such approaches hold the potential to advance statistical methodology in both theory and practice, enabling more reliable inference and improved decision-making across disciplines (Lasso, 2016).

The rapid advancement of complex statistical modeling and computing to fit real-world data structures need the best model selection criteria (Ebrahim et al., 2024). In modeling statistically, paying attention to model convergence and interpretability; no specific reason to select a single best model according to some criterion(Wang et al., 2004). However, it is more sensible to "deselect" poor models; which might be overfitting or underfitting. Most times, this subset might consider just a model, but sometimes probably not (Bozdogan, 2000).

Model selection is one of major challenge in statistical modeling, especially when dealing with hierarchical models, such as those found in Bayesian statistics. The task of selecting from competing models is typically framed as finding an optimal settlement of differences by arbitration between the model and complexity fit (Bozdogan, 2010). As noted by (Xu et al., 2023), this trade-off is central to selecting models that are both effective in explaining the data and parsimonious in their use of parameters. A model, in this context, is a full specification of probability for all observables, including the likelihood function (or the sampling distribution for the observables conditioned on the model population measures) and the prior distribution for these population measures. Model selection criteria, therefore, depend on how these components are handled and can be classified into conditional criteria, which are "based on the likelihood given the random effects, or marginal criteria", which integrate out these random effects (Ariyo et al., 2020.).

Deciding the best and efficient model, "criteria are that allow model comparisons are necessary". Some researchers agree that it is desirable to have complex models (Gelman & Hill, 2009), while there is preference for balance uncertainty from other researchers, which results from large complex models and bias from simple models. The second method place emphases on parsimony (Aho & Peterson, 2016).

Information criteria are employed to select best-fitted and efficient models. Practically, they are employed in obtaining minimized estimate criterion of the model. The general form of information criterion is given as; $R_A\left(\hat{\beta}_A\right) + \alpha_n\ (c_A)$; R_A is a loss function for models R_1, R_2 respectively for $R_1 \subset R_2$, satisfying $Q_{M_2}(\hat{\beta}_{R_2}) \leq Q_{M_1}(\beta_{R_1})$ with the penalty function α_n; model complexity function (Müller et al., 2013).

ISSN NUMBER: 1116-249X

Various criteria for model selection are available but the ones that will considered in this work are; Akaike Information criteria (BIC), Bayesian Information Criteria (BIC), Minimum Description Length (MDL), Information Complexity Criteria (ICOMP) and Widely Applicable Information Criteria (WAIC). Serious attentions are given to the complexity of the model thereby the model selection criteria penalize the likelihood based on the size of the variable and the sample in the model. Deducting the likelihood value from a specific value depends on the model complexity which results to penalty (Emiliano et al., 2014).

The aim of this research work is to develop a novel Bayesian hierarchical generalized Linear model that balances complexity and fit. The specific objectives are;

(a). compare the proposed model with the existing Bayesian hierarchical generalized Linear model.

(b). investigate the effectiveness of different information complexity criteria tailored for Bayesian hierarchical GLMs by focusing on computationally efficient methods for model selection in high-dimensional, hierarchical data environments.

METHODOLOGY

(Ebrahim et al., 2024) proposed three models whereby Conditional Random Coefficient model; $Logit\left(P_{ij}\right) = \alpha + \alpha_i + \alpha_{ij} + (\beta_i)X_{ij}$ was selected as the best fitted model. It can be assumed that both the random intercepts and coefficient/slope model hyperparameters μ_g, μ_β, σ_g, σ_β and ρ have uniform hyperprior distributions with assumptions suitable for parameters.

$Y_{ij} = Binomial(n_i = 1, P_{ij})$ ..... (i)

$Logit\left(P_{ij}\right) = \alpha + \alpha_i + \alpha_{ij} + (\beta_i)X_{ij}$ ..... (ii)

$$\alpha \sim Normal(0, 10)$$
$$(\beta_1, \ldots., \beta_p) \sim Normal(0, 10)$$
$$\alpha_j \sim Normal(0, \sigma_j)$$
$$\alpha_i \sim Normal(0, \sigma_i)$$
$$\sigma_i \sim halfcauchy(0, 1)$$
$$\sigma_j \sim halfcauchy(0, 1)$$

α is the overall intercept ,α_i and α_{ij} are intercepts σ_i and σ_j variance (standard deviation) components for each individual considering all measurements and for the repeated measures at j respectively.

The random coefficients model, also known as the conditional random coefficient model, incorporates all varying effects by adjusting both the intercept (mean) of the outcome variable and the predictor coefficients (slopes) to account for random effects at any clustering level within hierarchical data. The term "coefficient" (slope) in this model emphasizes that a random slope model estimates both the intercept (mean) and the regression coefficients at the appropriate hierarchical level (Gelman & Hill, 2009).

ISSN NUMBER: 1116-249X

However, while this model effectively captures hierarchical structures, with multiple hierarchical intercepts, the effects of predictors may become obscure, making interpretation difficult and does not include any regularization mechanism, making it susceptible to overfitting when dealing with high-dimensional predictors.

The Proposed Model

The model proposed by (Lu et al., 2018) serves as the foundation for the proposed model in this research work. While their hierarchical Bayesian mixed-effects model is designed for continuous outcomes and employs a Gaussian prior on the fixed effects, this study extends the framework to accommodate binary outcomes by adopting a Bernoulli likelihood. In addition, a Laplace (Lasso) prior on the fixed effects was introduced to enable variable selection. These modifications allow the model to better address interpretability of the data that follow Bernoulli, while retaining the strengths of the original hierarchical approach.

Also, in order to overcome some of the challenges posed by (Ebrahim et al., 2024) model, there is need for an enhanced novel Bayesian hierarchical model that integrates Lasso regularization into the framework which can significantly improves the interpretability and computational efficiency of the model as an improved model over the existing models.

$$logit(\psi_{ij}) = \propto + AX_{ij} + BZ_{ij} + \lambda \sum_k |A_k| \quad \text{(iii)}$$

Where; ψ_{ij} is the binary outcome for the ij^{th} observation, $\propto$ is the global intercept, AX_{ij} represents the fixed effect where A is the vector of fixed effects coefficients, X_{ij} is the vector of predictors for the ij^{th}observation, BZ_{ij} represents the random effects where B is the random coefficient associated with Z_{ij}, a vector of group level or observation- specific predictors, $\lambda \sum_k |A_k|$ is the Lasso penalty term

The proposed model introduces a Lasso penalty term ($\lambda \sum_k |A_k|$), which enforces sparsity in the fixed effects coefficients which helps in automatic variable selection by shrinking insignificant coefficients towards zero, reducing model complexity and preventing overfitting. It also improves interpretability by selecting only the most relevant predictors, making the model easier to interpret.

The model assume that; the likelihood function, the response variable ψ_{ij} follows a Bernoulli distribution: $Y_{ij} \sim Bernoulli\ (\psi_{ij})$ where ψ_{ij} is the predicted probability from the logistic function, the logit transformation of the outcome probability is assumed to be a linear combination of fixed effects, random effects, and the Lasso penalty term $logit(\psi_{ij}) = \propto + AX_{ij} + BZ_{ij} + \lambda \sum_k |A_k|$ which ensures that the relationship between the predictors and the log-odds of the outcome is appropriately modeled, the random effect coefficient (B) and global intercept ($\propto$) are normally distributed ; $\propto \sim N(0,\ \sigma_{\propto}^2),\ B \sim N(0,\ \sigma_B^2)$, the fixed effect coefficients A_k are modelled

ISSN NUMBER: 1116-249X

using a Laplace prior to enforce sparsity; $A_k \sim Laplace\ (0,\ \tau)\ where\ \ \tau = {}^{1}/_{\lambda}$, the variance parameter (σ_B^2) follows an inverse gamma distribution $\sigma_B^2 \sim Inverse\ Gamma\ (a, b)$ and the logit transformation of the probability ψ_{ij} follows a linear combination of fixed and random effects.

The likelihood function of the model; Since the outcome follows Y_{ij} a Bernoulli distribution, the likelihood for a single observation ψ_{ij} is:

$$P(Y_{ij}\,|\psi_{ij}) = \psi_{ij}(y_{ij})(1 - \psi_{ij})^{1-y_{ij}} \quad \text{..... (iv)}$$

$$\psi_{ij} = \frac{\exp(\propto + AX_{ij} + BZ_{ij} + \lambda \sum_k |A_k|)}{1+\exp(\propto + AX_{ij} + BZ_{ij} + \lambda \sum_k |A_k|)} \quad \text{..... (v)}$$

Thus substituting ψ_{ij} :

$$P(Y_{ij} \mid \propto, A, B, X_{ij}, Z_{ij}) = \left[\frac{\exp(\propto + AX_{ij} + BZ_{ij} + \lambda \sum_k |A_k|)}{1+\exp(\propto + AX_{ij} + BZ_{ij} + \lambda \sum_k |A_k|)}\right]^{Y_{ij}} \left[\frac{1}{1+\exp(\propto + AX_{ij} + BZ_{ij} + \lambda \sum_k |A_k|)}\right]^{1-Y_{ij}} \quad \text{..... (vi)}$$

Likelihood for N observations

For N observations, assuming the responses Y_{ij} are independent given the parameters, the joint likelihood is the product of the individual likelihoods:

$$P\ (Y \mid \propto, A, B, X, Z) = \prod_{i=1}^{N} P((Y_{ij} \mid \propto, A, B, X_{ij}, Z_{ij}) \quad \text{..... (vii)}$$

substituting the Bernoulli likelihood;

$$P\ (Y \mid \propto, A, B, X, Z) = \prod_{i=1}^{N} \left[\frac{\exp(\propto + AX_{ij} + BZ_{ij} + \lambda \sum_k |A_k|)}{1+\exp(\propto + AX_{ij} + BZ_{ij} + \lambda \sum_k |A_k|)}\right]^{Y_{ij}} \left[\frac{1}{1+\exp(\propto + AX_{ij} + BZ_{ij} + \lambda \sum_k |A_k|)}\right]^{1-Y_{ij}} \quad \text{..... (viii)}$$

Taking the log-likelihood;

$$\log P\ (Y \mid \propto, A, B, X, Z) = \sum_{i=1}^{N} \left(Y_{ij}\ (\propto + AX_{ij} + BZ_{ij})\right) - \log(1 + \exp(\propto + AX_{ij} + BZ_{ij})) \quad \text{..... (ix)}$$

This log-likelihood function is maximized during Bayesian inference using MCMC to estimate posterior distributions. The likelihood serves as a foundation for Bayesian inference by linking observed data to prior distributions, leading to posterior estimation via Bayes' theorem.

ISSN NUMBER: 1116-249X

In the Bayesian hierarchical generalized linear model with a logistic link function, the posterior distribution is obtained using Bayes' theorem. The posterior distribution combines the likelihood with the priors;

$$P(\propto, A, B, \sigma_B^2 \,|Y, X, Z) \;\propto\; P(Y \mid \propto, A, B, X, Z)\, P(\propto) P(A) P(B|\sigma_B^2) P(\sigma_B^2)$$

By expansion,

$$\prod_{ij} Bernoulli\left(Y_{ij}|\psi_{ij}\right) \times N\left(\propto|0, \sigma_\propto^2\right.$$

$$\times \prod_{K} Laplace\,(A_k|0,\tau) \times N(B|0,\sigma_B^2) \times Inverse-Gamma\,(\sigma_B^2|a,b)$$

This posterior cannot be computed analytically, hence, it will be estimated using MCMC approach.

Comparisons of Results of the Posterior Simulation Techniques of Bayesian Hierarchical Generalized Linear Models (BHGLMs) Under Different Conditions.

Computational Efficiency Diagnostics of BHGLMs

The displayed output summarizes the progress of Bayesian sampling for the two BHGLMs under consideration; Existing and Proposed models using the Stan MCMC engine. The process involved running four independent Markov chains, each responsible for drawing samples from the posterior distribution of the model parameters to the models' convergence and reliability. Four different chains were running with 2000 iteration.

Table 1: A table showing the sampling diagnostic of the BHGLMs under consideration

Model	Gradient Evaluation Time (s)	Warm-up time per chain(s)	Sampling Time per chain (s)	Total sampling time (s)
Existing	0.00048 – 0.00089	14.5 – 16.3	8.5 – 13.1	~24 – 29
Proposed	0.000906 – 0.001796	33.0 – 33.4	24.7 – 30.4	~57.8 – 63.4

From table above, Sampling diagnostics for the two candidate models were examined to assess computational efficiency and sampling stability. The two models completed four parallel MCMC chains without any reported warnings or divergent transitions, indicating stable convergence. The computational time for the Existing model (24–29 seconds per chain), while the Proposed Model (which incorporates lasso regularization and a more complex hierarchical structure) was the most computationally intensive, requiring 58 – 63 seconds per chain. Thus, while two models were stable, the Existing Model is preferable if computational speed is the primary concern, whereas

ISSN NUMBER: 1116-249X

the Proposed Model greater time cost may be warranted if model sparsity or regularization is a priority.

Convergence Diagnostic of the Posterior Distribution of BHGLMs

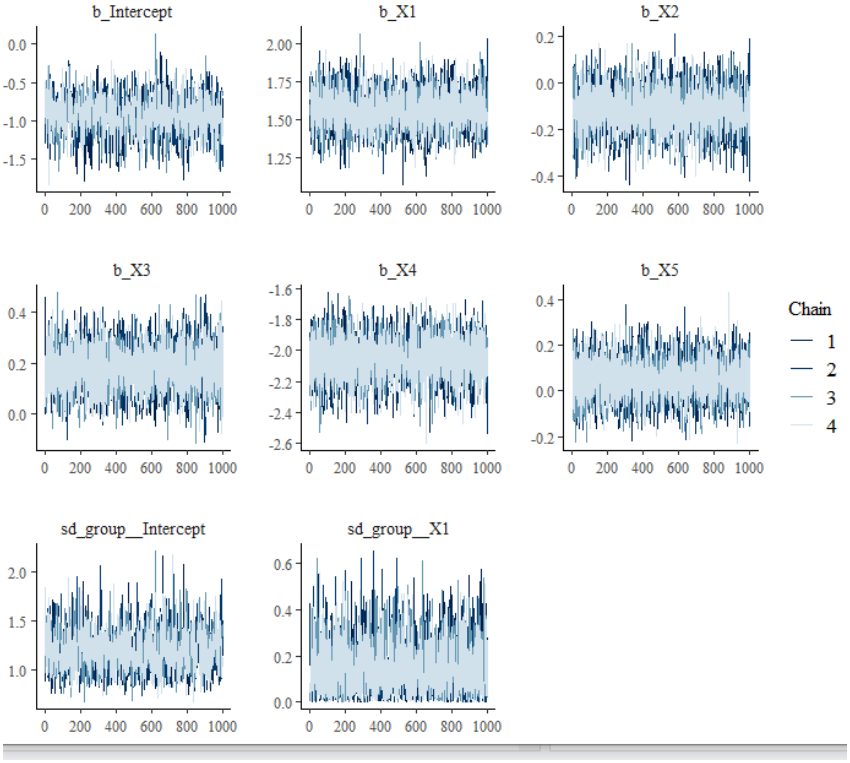

Fig. 1(a) Trace plots for the Existing model

ISSN NUMBER: 1116-249X

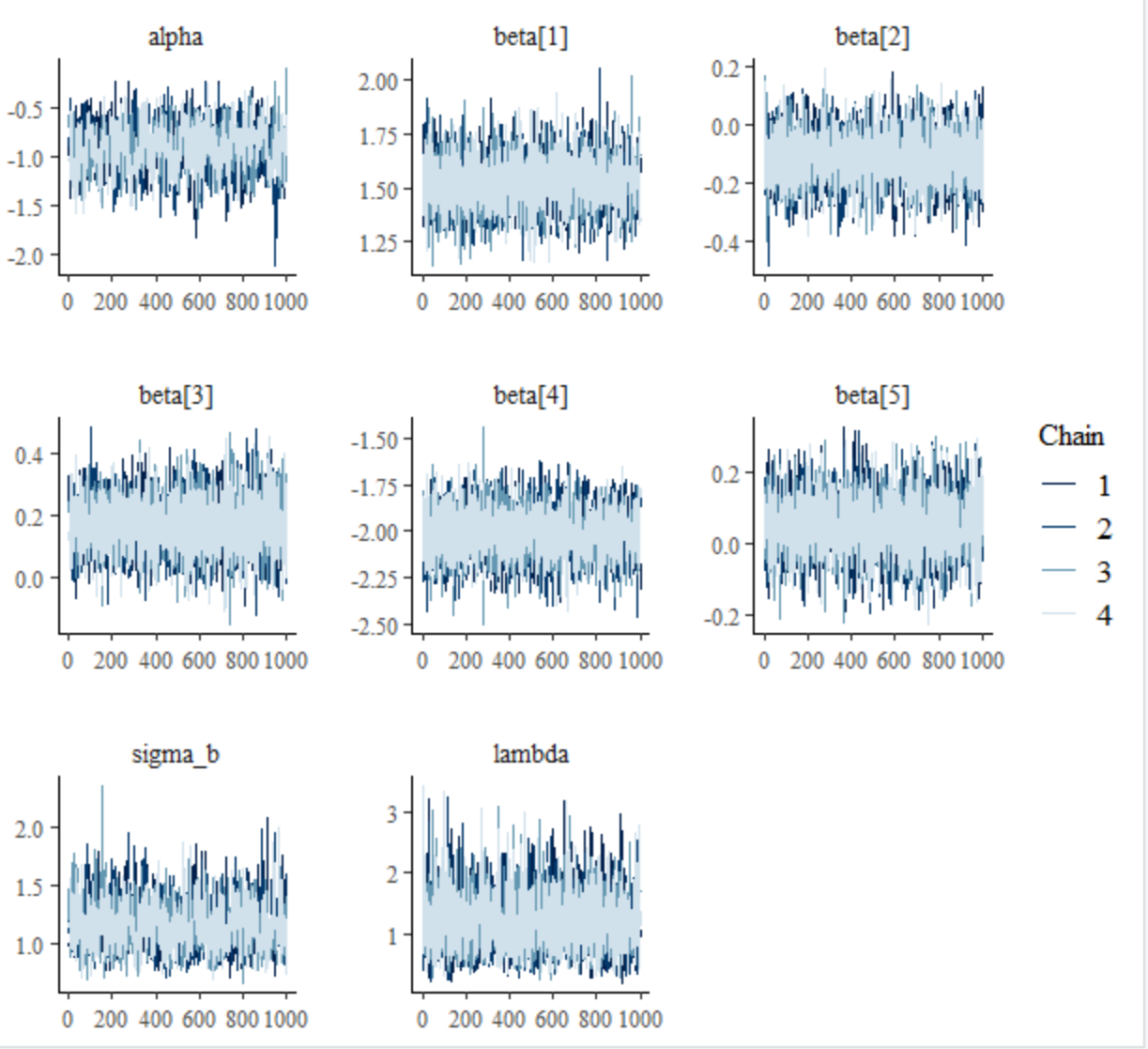

Fig. 1(b) Trace plots for the Proposed Model

Discussion

In fig.1(a), the trace plots shown for the Existing Model, display the Markov Chain Monte Carlo (MCMC) sampling paths for each parameter across four independent chains. Each subplot corresponds to a model parameter: the global intercept, fixed effect coefficients (bX_1 to bX_5), and random effect standard deviations ($sd\ for\ group\ intercept\ and\ sd\ for\ group\ X_1$). The plots revealed a satisfactory sampling behavior for all model parameters. All four MCMC chains demonstrated good mixing, stationarity, and substantial overlap, with no evidence of non-convergence or autocorrelation issues.

In fig.1(b), the plot displays the trace plot for the Proposed Model, shows the MCMC chains for each key parameter: the global intercept ($alpha$), fixed effect coefficients (beta[1] to beta[5]), the

ISSN NUMBER: 1116-249X

random effect standard deviation (sigma-b), and the Lasso regularization parameter (lambda). Each panel shows 4 chains with different colors. All chains for each parameter show rapid movement across the parameter space, with no evidence of chains being stuck or trending. The chains overlap well and move freely, suggesting good mixing and convergence. For every parameter, the sample paths oscillate around a stable mean. Even with increased complexity (Lasso penalty, additional hyperparameters), the chains mix well and converge for all parameters and there is no visible drift or trend, indicating that the sampler has reached a stationary state. In conclusion, all the three models exhibit excellent MCMC diagnostics. However, the Proposed model is especially notable for maintaining excellent mixing despite its higher complexity and the inclusion of the lasso penalty. This suggests the model is both robust and computationally feasible, even when introducing regularization for variable selection.

Autocorrelation Function Curve as a Diagnostic for MCMC Convergence in Bayesian Hierarchical Generalized Linear Models

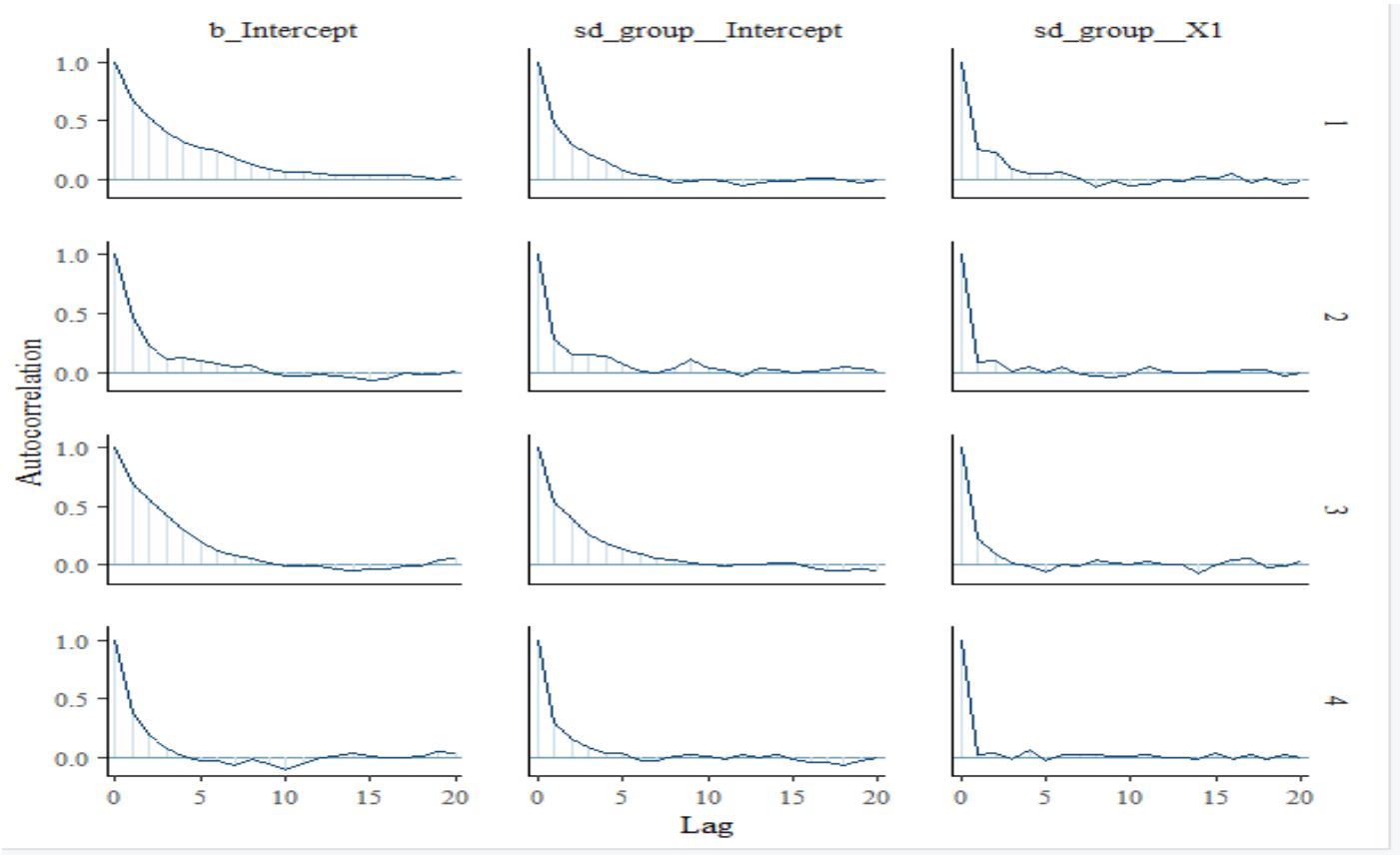

Fig. 2(a): Autocorrelation function graph for the Existing model

ISSN NUMBER: 1116-249X

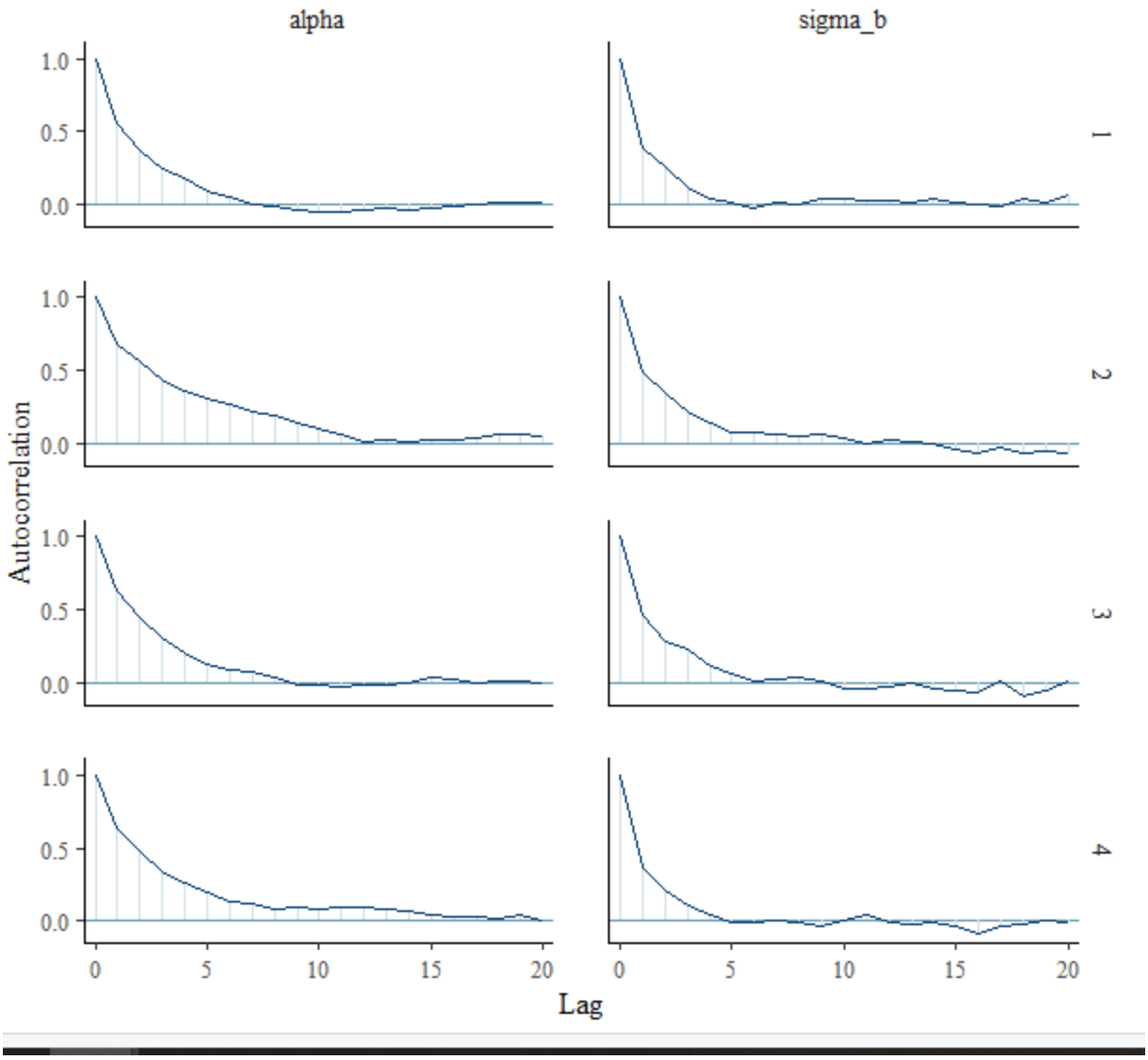

Fig. 2(b): Autocorrelation function graph for the Proposed model

Discussion

In fig. 2(a), the plot shows autocorrelation functions (ACF) for three key parameters of Existing Model (b-Intercept, sd- group Intercept, sd- group X_1) across the four MCMC chains. All parameters and all chains, autocorrelation drops sharply with increasing lag and approaches zero quickly (generally within 5–10 lags). After the initial few lags, the autocorrelation values hover around zero, indicating that consecutive samples are nearly independent after a small number of iterations. The pattern of autocorrelation decay is similar across all four chains and all parameters, showing consistent sampling behavior.

Fig. 2(b), is the autocorrelation functions (ACF) plot for Proposed Model, focusing on the parameters alpha (the global intercept) and sigma- b (the random effect standard deviation) across

ISSN NUMBER: 1116-249X

four chains. The autocorrelation drops rapidly with increasing lag for both parameters and all chains, mirroring the pattern seen in Existing Model. For Parameters; alpha, sigma- b. Autocorrelation also drops sharply to near zero by lag 5–10 for all chains with no evidence of slow mixing or stickiness, even with added model complexity (Lasso penalty and more hierarchical structure).

In summary, Autocorrelation analysis for the two candidate models demonstrated rapid decay of autocorrelation values to near zero within 5–10 lags for all monitored parameters and chains. This pattern was consistent across models, indicating that the Markov chains for each model are mixing efficiently and producing nearly independent samples after a short lag. As a result, the Bayesian inference derived from the two models is expected to be both reliable and precise. Notably, the Proposed Model maintained this high level of sampling efficiency despite its greater complexity and inclusion of Lasso regularization, suggesting that additional hierarchical structure and penalization did not compromise MCMC performance.

Coefficients Comparison Across the BHGLMs

Table 2: A table showing the Posterior Mean Estimates of the Regression Coefficients for each Predictor X_1 to X_5 in BHGLMs under consideration

Variable	Existing Model mean	Proposed Model mean
X_1	1.5685	1.5285
X_2	-0.1169	-0.1098
X_3	0.1806	0.1733
X_4	-2.0483	-2.0110
X_5	0.0541	0.0513

Interpretation

The mean estimates for all coefficients are very similar across the two models. This indicates that the models are producing consistent inferences about the effects of the predictors, regardless of differences in model structure (e.g., random effects, regularization). X_1 has a strong, positive, and stable effect in all models (~1.53 – 1.57), X_2 and X_3 have small effects (X_2 slightly negative, X_3 slightly positive), X_4 consistently shows a strong negative effect (~-2.01 to -2.05) while X_5 shows a very small positive effect in all models. However, the Proposed Model with lasso regularization shows slightly smaller absolute values for all coefficients compared to the Existing Model. This is expected behavior for regularized models, which pull estimates toward zero to avoid overfitting. For the attainment of parsimony, prediction, or protection against overfitting, the proposed Model is preferable due to its regularization, which helps prevent the inflation of spurious effects and typically improves out-of-sample predictive performance.

ISSN NUMBER: 1116-249X

Bias and Coverage Comparison Across BHGLMs

Table 3: A table showing Bias and coverage Level of the BHGLMs

	Bias		Coverage	
Variable	Existing Model	Proposed Model	Existing Model	Proposed Model
X_1	0.0685	0.0285	FALSE	TRUE
X_2	-0.1169	-0.1098	FALSE	TRUE
X_3	0.1806	0.1733	TRUE	FALSE
X_4	-0.0483	-0.0110	FALSE	TRUE
X_5	0.0541	0.0513	FALSE	TRUE

Interpretation

From table 3 above, the Proposed Model is clearly the best model based on bias, as it produces coefficient estimates closest to the true values for all predictors. This suggests the Proposed Model has the highest estimation accuracy between the two models considered, likely due to its use of regularization, which helps reduce overfitting and brings parameter estimates closer to their true values.

In terms of coverage, for the Proposed models only variable X_3 is covered (TRUE); all other variables' intervals miss the true value (FALSE). This indicates that the 95% intervals for $X_1, X_2, X_4 \ and \ X_5$ do not contain the true coefficients which implying these models are systematically underestimating (or overestimating) the uncertainty, or are biased. However, the Coverage is TRUE for $X_1, X_2, X_4 \ and \ X_5$, and FALSE only for X_3. This means that, for most variables, the Proposed model's intervals correctly capture the true value, indicating more honest and reliable uncertainty quantification which is clearly the best model based on coverage. Therefore, the Proposed Model not only estimates coefficients with lower bias, but also provides uncertainty intervals that accurately reflect uncertainty about the true parameter values.

Information Criteria Model Comparison of BHGLMs

Table 4: A table showing the BHGLMs comparison using different Information Criteria

Model	AIC	BIC	MDL	ICOMP	WAIC	LOO
Existing	2908.30	2949.02	178.81	-42.82	901.40	901.7
Proposed	2908.30	2949.02	177.02	-41.10	899.25	899.5

Interpretation

ISSN NUMBER: 1116-249X

Considering AIC and BIC, all two models have identical AIC and BIC values which are 2908.30 and 2949.02 respectively. This implies that the two models fit the data equally well in terms of overall likelihood and basic complexity penalty. Considering MDL, the Proposed model has the lowest value (177.02) among the three models under consideration which suggests that the proposed model concisely the model describes the data. For the ICOMP, the Adjusted model with the value; -43.98 is the lowest and considered best among the three models fitted. For both WAIC, the Proposed model with the value 899.25 is the lowest and best compared to the Existing model with the value 901.40 indicating that the Proposed Model has the best out-of-sample predictive accuracy. In conclusion, based on the full set of information criteria (AIC, BIC, MDL, ICOMP, WAIC, LOO), the Proposed Model offers the best overall balance of fit, complexity, and predictive accuracy.

CONCLUSION

This research introduces an enhanced Bayesian hierarchical generalized linear modeling approach that integrates Lasso regularization to address the dual challenges of high dimensionality and hierarchical data structure. Through rigorous simulation and empirical comparison, the proposed model demonstrates improved estimation accuracy, reduced bias, and better uncertainty quantification compared to existing models. Diagnostics show excellent MCMC convergence and efficient sampling, even with increased model complexity. Model selection criteria consistently favor the proposed approach, highlighting its balance between goodness-of-fit, complexity, and predictive power. While computational demands are greater, the added interpretability and parsimony justify the trade-off, making the model a compelling choice for modern complex datasets. Overall, this work advances statistical methodology and provides practical tools for reliable inference and decision-making in high-dimensional, hierarchical contexts.

While this study demonstrates the advantages of integrating Lasso regularization into Bayesian hierarchical generalized linear models (BHGLMs) for high-dimensional and hierarchical data, several avenues remain open for further exploration. Future research could extend the proposed framework to accommodate other types of outcome variables, such as count data (using Poisson or negative binomial likelihoods), continuous outcomes, or multi-category responses. This would broaden the applicability of the model to a wider range of scientific domains. Investigating the performance of other regularization techniques, such as Ridge, Elastic Net, Horseshoe, or Spike-and-Slab priors within the hierarchical Bayesian context could yield additional insights into variable selection and model performance.

CONFLICT OF INTEREST

On behalf of all authors, the corresponding author states that there is no conflict of interest.

REFERENCES

Bozdogan, H. (2000). Akaike’s Information Criterion and Recent Developments in Information

ISSN NUMBER: 1116-249X

Complexity INTRODUCTION AND PURPOSE. In *Journal of Mathematical Psychology* (Vol. 44). www.idealibrary.com

George, E. I., & Foster, D. P. (2000). Calibration and empirical bayes variable selection. *Biometrika*, *87*(4), 731–747. https://doi.org/10.1093/biomet/87.4.731

Carlin, J. B., Wolfe, R., Hendricks Brown, C., & Gelman, A. (2001). A case study on the choice, interpretation and checking of multilevel models for longitudinal binary outcomes. In *Biostatistics* (Vol. 2, Issue 4).

Wikle, C. K. (2003). Hierarchical Bayesian models for predicting the spread of ecological processes. *Ecology*, *84*(6), 1382–1394. https://doi.org/10.1890/0012-9658(2003)084[1382:HBMFPT]2.0.CO;2

Wang, X., George, E. I., & Wang is Assistant Professor, X. (2004). *A hierarchical bayes approach to variable selection for generalized linear models*. https://www.researchgate.net/publication/228705104.

Yuan, M., & Lin, Y. (2005). Efficient empirical bayes variable selection and estimation in linear models. *Journal of the American Statistical Association*, *100*(472), 1215–1225. https://doi.org/10.1198/016214505000000367.

Park, T. & Casella, G. (2008). The Bayesian Lasso. Journal of the American Statistical

Association 103, 681–686.

Gelman, A., & Hill, J. (2009). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Nannen, V. (2010). *A Short Introduction to Model Selection, Kolmogorov Complexity and Minimum Description Length (MDL)*. *Mdl*, 1–20. http://arxiv.org/abs/1005.2364.

Bozdogan, H. (2010). A new class of information complexity (ICOMP) criteria with an application to customer profiling and segmentation Invited Paper A new class of information complexity (ICOMP) criteria with an application to customer profiling and segmentation. *Dergisi Istanbul University Journal of the School of Business Administration Cilt*, *39*(2), 370–398. www.ifdergisi.org.

Yi, N., Liu, N., Zhi, D., & Li, J. (2011). Hierarchical generalized linear models for multiple groups of rare and common variants: Jointly estimating group and Individual-Variant effects. *PLoS Genetics*, *7*(12). https://doi.org/10.1371/journal.pgen.1002382.

Müller, S., Scealy, J. L., & Welsh, A. H. (2013). *Model Selection in Linear Mixed Models*. *28*(2), 135–167. https://doi.org/10.1214/12-STS410.

Gürünlü Alma, Ö. (2013). PERFORMANCE COMPARISONS OF MODEL SELECTION CRITERIA: AIC, BIC, ICOMP AND WOLD'S FOR PLSR. *Journal of Statistical Research*.

Emiliano, P. C., Vivanco, M. J. F., & Menezes, F. S. De. (2014). *Information criteria : How do they behave in different models ? 69*, 141–153.

Fu, S. (2015). *Hierarchical Bayesian LASSO for a negative binomial regression*.

ISSN NUMBER: 1116-249X

Pamukçu, E., Bozdogan, H., & Çalik, S. (2015). A novel hybrid dimension reduction technique for undersized high dimensional gene expression data sets using information complexity criterion for cancer classification. *Computational and Mathematical Methods in Medicine*, *2015*. https://doi.org/10.1155/2015/370640.

Aho, K., & Peterson, T. (2016). *Model selection for ecologists : The worldviews of AIC and BIC Model selection for ecologists : the worldviews of AIC and BIC*. *July*. https://doi.org/10.1890/13-1452.1.

Lasso, A. N. B. (2016). *HHS Public Access*. *7*(4), 571–582. https://doi.org/10.4310/SII.2014.v7.n4.a12.A.

Wirawati, I., Iriawan, N., & Irhamah. (2017). Bayesian Hierarchical Random Intercept Model Based on Three Parameter Gamma Distribution. *Journal of Physics: Conference Series*, *855*(1). https://doi.org/10.1088/1742-6596/855/1/012061.

Robert, C. P., Elvira, V., Tawn, N., & Wu, C. (2018). *Accelerating MCMC algorithms*. *April*, 1–14. https://doi.org/10.1002/wics.1435.

Lu, S., Zhang, J., Zhou, S., & Xu, A. (2018). Reliability prediction of the axial ultimate bearing capacity of piles: A hierarchical Bayesian method. *Advances in Mechanical Engineering*, *10*(11), 1–11. https://doi.org/10.1177/1687814018811054.

Ariyo, O., Lesaffre, E., Verbeke, G., & Quintero, A. (2020). *Bayesian model selection for longitudinal count data*.

Günther, F., Bender, A., Katz, K., Küchenhoff, H., & Höhle, M. (2021). Nowcasting the COVID-19 pandemic in Bavaria. *Biometrical Journal*, *63*(3), 490–502. https://doi.org/10.1002/bimj.202000112.

Mallick, H., Alhamzawi, R., Paul, E., Svetnik, V., & Sep, M. E. (2021). *The Reciprocal Bayesian LASSO*. 1–35.

Xu, S., Ferreira, M. A. R., Franck, C. T., & Porter, E. M. (2023). *Bayesian model selection for generalized linear mixed models*. *June*, 3266–3278. https://doi.org/10.1111/biom.13896.

Ebrahim, E. A., Cengiz, M. A., & Terzi, E. (2024). The Best Fit Bayesian Hierarchical Generalized Linear Model Selection Using Information Complexity Criteria in the MCMC Approach. *Journal of Mathematics*, *2024*. https://doi.org/10.1155/2024/1459524.

Zhong, R., & Moraga, P. (2024). Bayesian Hierarchical Models for the Combination of Spatially Misaligned Data: A Comparison of Melding and Downscaler Approaches Using INLA and SPDE. *Journal of Agricultural, Biological, and Environmental Statistics*, *29*(1), 110–129. https://doi.org/10.1007/s13253-023-00559-w.

ISSN NUMBER: 1116-249X