International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Advancing Hierarchical Model: Evaluating Performance, Interpretability and Implications

Seth Opoku Larbi[1], Apaka Rangita[2], Joyce Otieno[3]

[1]MSc, Planning Office, Koforidua Technical University, Directorate of Quality Assurance & Academic Planning, Ghana.
[2,3]PhD, Department of Statistics & Actuarial Science, Maseno University, School of Mathematics, Statistics and Actuarial Science, Kenya.

Abstract
Hierarchical data structures emerge when observations are nested within higher-level units or clusters. Existing research often ignores the hierarchical structure of data leading to biased estimates, suboptimal model selection, and challenges in identifying important predictors and dependencies. This study aims to contribute to hierarchical frameworks by improving challenges with the interpretability of the random effect, scalability, and computation feasibility in the traditional hierarchical Bayesian model. The authors' model is an advancement of the Standard hierarchical Bayesian model which introduced a unique variable to the model and parameters to the random effects. The advancements in hierarchical modelling by the authors have significantly improved the accuracy, reliability, and interpretability of the model analysis. Hierarchical Bayesian Information Criteria (HBIC) is the method of selecting variables in the models. The findings of the research indicated that the introduction of Phi (ϕ_j) and Psi (ψ_j) as shrinkage parameters are instrumental in the regulation of parameter estimates towards a common value, resulting in more accurate estimation and less overfitting in the Hierarchical Bayesian Model as well as accounting for the presence of heterogeneity. The introduction of the unique variable (z) allows the model to capture cluster-specific effects associated with the (z). Lastly, the study found that the authors' innovative model outperformed the standard model by improving the accuracy, interpretability, scalability, and regularisation through shrinkage parameters and innovative (z) as the contextual variable.

Keywords: Shrinkage, Hierarchical, Parameters, Prior, Posterior, Uncertainty, Lasso, Unique Variable

Introduction

There has been a surge in modern research and an interest in learning and understanding data-driven decision-making. Having a larger dataset analysis is no more sophisticated, as insights in the underlying huge dataset are no more hidden which makes it exciting and appealing. Despite such advancement in research methodologies, there is a limitation in hierarchical model procedure where potential researchers find it difficult to extract all issues surrounding the data. Most fascinating studies have looked into examining the predictor variables separately ignoring the aspect of interaction and the nested aspect within the dataset (Johnson, 2018). Hierarchical models and structures including the levels of data or clustered observations are common in social science, ecology, and epidemiology areas (Green & White, 2017). Disregarding such structures leads to estimation bias and incorrect conclusions demanding robust statistical tools that address the complexity of data (Rajaraman, 2020). A common approach in analysing multilevel data is through a hierarchical framework by introducing random effects to account for variability in the level of structures (James & Douglas, 2016). The researcher's proposed model adopts a hierarchical structure where the individual-level predictors fixed effects, and group-level variables are captured through the random effects (Klein & Moeschberger, 2018). The researcher's model incorporates interactions within the individual-level explanatory variables and group-level random effects which regulate whether a lower-level factor can moderate that of a higher-level one (Gelman & Hill, 2007). The researchers believe the study will be a game changer in the sphere of hierarchical modelling by accounting for the structures and interactions between lower and higher-level factors. Such insight provides more accurate model that inform policy decision-making through the usage of real datasets.

Hierarchical Bayesian modelling is a formal statistical framework that has established norms that guide its application and interpretation. Hierarchical modelling ensures that the procedure for a model is valid, reproducible, and transparent. The hierarchical model basic norm concentrates specifically on the prior distribution of parameters which should be clearly defined. The knowledge and beliefs of the parameters considered in the model ought to be chosen in the right manner. Hierarchical modelling allows for the nested structures within data and the incorporation of uncertainty, prior information, and variability with the structure settings. The parameters of interest are treated as random variables in Hierarchical modelling. Prior distributions are also defined for each parameter of interest at different levels in the hierarchy to cater for the uncertainty about the parameters before the data is observed. Bayesian theorem is used to observe the posterior distribution of data. The posterior distribution is updated using the combination of likelihood data and prior information.

There has been extensive research nowadays on Hierarchical modelling in the areas of social sciences, economics, epidemiology, and ecology just to mention a few as a result of its ability to handle complex data and beliefs. Due to this, hierarchical modelling analysis has evolved as a powerful tool to handle complex data with uncertainties for decision-making. The use of the most appropriate method in hierarchical Bayesian model selection is to initially identify the model structures that account for the fixed effects, random effects, and interaction which predicts the underlying patterns in a data more accurately (Gelman & Hill, 2007). This is often carried out by comparing alternative models using criteria that weigh the balance that must be taken into account for model fitness and the one which is not overly complicated, such as the Deviance Information Criterion (DID) or the Widely Applicable Information Criterion (WAIC) (Gelman et al., 2023).

The hierarchical Bayesian model has gained several successes although there have been teething limitations and challenges that ought to be considered to formulate a parsimonious model. One of the biggest challenges encountered by researchers in hierarchical modelling is its intensiveness in computing large datasets and at times complex structures. This results in long computational times, high autocorrelation, and slow convergences because of the posterior inferences from the Markov Chain Monte Carlo (MCMC) algorithm. Informative priors that present accurate prior knowledge selection are essential but become a challenge when there is no or slight essential information. Wrong specification of priors results in biased estimates and unreliable conclusions (Klein & Moeschberger, 2018). If the model does not include important predictors or it is not well understood how the data was designed, it can fail to represent significant structures and patterns contained in the data. The hierarchical structure of data is many at times ignored by researchers which leads the model to bias in estimation, suboptimal model selection, and also challenges to be able to identify the essential predictors and dependencies. This problem is encountered by researchers in epidemiology, ecology, and social sciences areas where researchers work with nested structures regularly. Many times, it is difficult to identify the appropriate statistical model because traditional model selection criteria are inadequate for complex datasets and hierarchical relationships. Traditional variable selection procedures are ill-equipped to handle complex data settings with many structures and tens or hundreds of possible predictors to be considered in social and biological networks; this leads researchers to potential suboptimal model construction and interpretation.

However, the major hope for researchers in addressing the aforementioned challenges is through the use of the Hierarchical Bayesian model. Bayesian hierarchical modelling provides a reproducible framework and transparency with articulated guidelines for validating models. There are even practical challenges with model transparency, uncertainty quantification, scalability, interpretability of the random effects, and computational feasibility of which researchers have to be circumspect. Researchers at times ignore the cumbersome dependencies and the correlation within the nested data structures with the traditional model and variable selection procedures. An effective model is produced by the traditional model and variable selection technique; however, the model's predictive performance diminishes, biased estimates are produced, and type I error rates are inflated. Researchers must therefore focus on improving a model's interpretability concerning the relationships under different conditions and hierarchy for the predictor and the criterion variables.

Failure of existing research or statistical methodologies in identifying problems to accommodate hierarchical structures in datasets often leads to suboptimal model and variable selection, and inefficient and biased estimates. Therefore, in order to address hierarchical data analysis challenges and enable accurate model interpretability, transparency, and making informed decisions, there is a need to address model comparison challenges, reproducibility, uncertainty quantification, and model transparency through a robust statistical technique.

The hierarchical structure of the data will be addressed amidst the limitations encountered in hierarchical modelling methodologies by using the researchers' model to enhance the robustness and predictive accuracy through the hierarchical model and variable selection technique. The existing research methodologies regarding model selection procedures focused on individual characteristics; though the researchers' model will address the heterogeneity and dependencies within nested data. The stability of the variable selection in the model, the interpretability of model results, model fitness, and predictive accuracy will be the outcome of the researcher's proposed model. The proposed methodology will address model and variable selection approaches on decision-making and inferences in various aspects and also ensure practical relevancy and applicability.

The general objective of the paper is to advance and evaluate the performance of hierarchical Bayesian modelling, as a way of enhancing accuracy, interpretability, and generalizability while tackling the limitations encountered by current methods. This study has specific aims which are to introduce a unique variable to the standard hierarchical Bayesian model to solve the current limitations and also to enhance on variable selection procedure within the hierarchical model.

The study contributes to scientific knowledge by introducing a hierarchical Bayesian model that is based on a global test (of fixed or random effects) under the likelihood ratio test framework. Secondly, the study advances variable selection techniques within the hierarchical modelling framework to improve the performance of the model. The results of this study will have direct applications in a wide range of domains including education, finance, healthcare, and social sciences, where hierarchical structures and high-dimensional data are common. The study will also enhance the quality and interpretability of methods used in science and other investigations. Specifically, by introducing a unique model the researchers enhance the accuracy and interpretability. This paper aims to make more feasible good decisions where transparency and understanding of model outputs are of utmost priority.

Literature Review

Modern data analysis nowadays has brought a surge in statistical literature on hierarchical models and variable selection techniques in addressing complex data more interpretably and transparently. The technique has gained much attention as a result of the hierarchical structure's ability to provide insight into data patterns and address its selection limitation to approaches. The hierarchical model allows for accurate capture of complex relationships and also models the dependencies among observations within the hierarchies. Improving the model's parsimony and interpretability is ensuring that a model selection technique is employed to elect the right predictors. A valuable contribution was made to the literature on Bayesian hierarchical modelling, by introducing a new approach to improving predictive performance through model selection criteria (Zong & Bradley, 2023). The article's theoretical and empirical contributions provide a valuable resource to researchers seeking to improve the accuracy and efficiency of Bayesian hierarchical models across a diverse set of applications by incorporating an often-used criterion from model selection (Zong & Bradley, 2023). A special case of several information criteria expressions was proposed by Zong & Bradley (2023) which they labeled as Covariance Penalized Error (CPE). By the variance empirical Bayes estimator, a penalized mean value was formed. Zong & Bradley's (2023) prime aim was to obtain a small value of the CPE criterion, by truncating the joint support of the data and the parameter space using Bayesian hierarchical modelling. The authors achieved their objective of minimizing the squared error by identifying a subset parameter space that produces lower values than the Bayesian model averaging yields, provided that there is a non-zero probability value within this truncated set. The limitation in their paper comprises Complexity of Implementation, Interpretability and Transparency, and Dependency on Criterion Selection.

In the publication titled "Income, education, and Other Poverty-related Variables: A Journey Through Bayesian Hierarchical Models," G´omez-M´endez and Chainarong (2023) focuses on addressing the issue of poverty in Thailand. The paper by Gomez-Mendez and Chainarong (2023) examines poverty-related variables such as income, education, and others in the multiresolution governing structural data of Thailand. Bayesian hierarchical models are utilized for this analysis. The authors discuss the progression of their modelling methodology from simple to more complex models and assess each model's effectiveness based on its ability to explain relevant variables while balancing complexity considerations. The models are designed to capture the data's hierarchical structure, incorporating individual and region-level variables. G´omez-M´endez and Chainarong (2023), while Bayesian hierarchical models provide a flexible and powerful framework for analyzing complex data, they have drawbacks, such as challenges in model specification, computational complexity, and overfitting when a large number of parameters are included. The availability, and quality, of data, may also limit the utility of the modelling approach. The methodology in the article "Income, education, and other poverty-related variables: a journey through Bayesian hierarchical models" by Gómez-Méndez and Chainarong (2023) can be summarized in mathematical and statistical terms as follows:

The hierarchical model aims to explain income and other poverty-related variables in Thailand. It can be represented as:

$$Y_{ij} = \beta_0 + \sum_{k=1}^{p} \beta_k X_{ijk} + \phi_j \varphi_{0j} + \varepsilon_{ij} \qquad \textbf{[1]}$$

Where

Y_{ij} is the income or poverty − related variable for observation i in region j,

β_0 is the intercept

β_k are the coefficients for covariates X_{ijk} (such as education level, household characteristics, etc.),

ϕ_j is the regional random effect capturing unobserved heterogeneity at the regional level,

φ_{0j} represents the random effect for region j,

ε_{ij} is the error term

The journey started by the authors in designing the hierarchical models in a simple and then gradually complex form. The performance of each model is assessed based on the ability to explain the variability in income and poverty-related variables. In practice, the preference for simpler models should not only be based on the explanatory power but also the complexity. If two models provide comparable explanatory power a simpler model is preferred to a more complex one. So, the evaluation of model fit, overfitting, and computational complexity for each model should be discussed. Thus, the regularization techniques (e.g., fixed-effects, random-effects, and mixed-effects models), model averaging, and sensitivity analysis for Bayesian hierarchical models. The limitations highlighted in the article as a single policy may not adequately address poverty-related issues in different areas, each with its unique challenges. Custom-made policies for each area separately are unrealistic due to resource limitations. The article points out the limitation of ignoring dependencies or relationships between different geographic areas when formulating poverty-related policies. While Bayesian hierarchical models provide a flexible and powerful framework for analyzing complex data, the authors' model has drawbacks, such as challenges in model specification, computational complexity, and potential overfitting because a large number of parameters were included. The availability, and quality, of data, may also limit the utility of the modelling approach.

The Article by Porter et al., (2023) titled “Objective Bayesian Model Selection for Spatial Hierarchical Models with Intrinsic Conditional Autoregressive Priors” addresses the problem of model selection for Gaussian hierarchical models with intrinsic conditional autoregressive (ICAR) spatial random effects. The problem of model selection is particularly challenging in spatial models as there is spatial dependence and confounding between fixed and spatial random effects. The Article develops a Bayesian model selection approach with fractional Bayes factors to do so, which not only selects regressors but also assesses spatial dependence. Moreover, Porter et al., (2022) also acknowledge the problem of selecting covariates and spatial model structure independently in spatial hierarchical models; previously, methods required one to be fixed a priori and the other to be selected, which necessitated arbitrary decisions. Methods for simultaneous selection solve this problem but there are few Bayesian methods for simultaneous selection; the Article aims to do so, using fractional Bayes factors for model selection under automatic reference priors. In developing this methodological contribution, the Article first develops Bayesian model selection by fractional Bayes factors, so that fixed effects especially are selected simultaneously with spatial model structure; the approach also uses automatic reference priors, so that hyperparameters for priors do not need to be specified. The Article then also establishes the stochastic ordering of two ICAR specifications and its implication for the fractional Bayes factor for the ICAR model under the reference prior. A comparison of the methodology to traditional model selection criteria and its performance in a simulation study is also undertaken.

Porter et al., (2023) hierarchical model under review is represented as:

$$Y_i = \beta_0 + \sum_{k=1}^{p} \beta_k X_{ik} + \varepsilon_i \qquad \textbf{[2]}$$

Where

Y_i is the response varable for observation i
β_0 is the intercept
β_k are the coefficients for covariates X_{ik}
ε_i is the error term

Spatial Random Effects: The model includes spatial random effects represented by the ICAR prior, denoted as ϕ_i.

Fractional Bayes Factors (FBFs): FBFs are used for model selection. M_0 represents the null model with no covariates and only spatial random effects, and M_1 represents the model with covariates and spatial random effects. The Fractional Bayes Factors (FBFs) for comparing models M_0 and M_1 are as follows:

$$FBF_{01} = \frac{BF_{01}}{1+ PBF_{01}} \quad \textbf{[3]}$$

where

BF_{01} is the Bayes factor comparing models M_0 and M_1,

FBF_{01} is the fractional Bayes factor.

The proposed approach offers a novel solution to model selection in spatial hierarchical models. Porter et al., (2023) model in the hierarchical framework has challenges including computational complexity, potential challenges in applying the method to large-scale spatial datasets, and sensitivity to prior specifications that the researchers encountered. Additionally, the approach's performance may vary based on the unique characteristics of each dataset and model specifications. Porter et al., (2023) limitation could be addressed using the authors' model which could be relevant to the article in these aspects.

Methodology

The methodology section of this article outlines the systematic approach employed to achieve the objectives of adding a unique variable to the Bayesian hierarchical model. The methodology section is therefore a transparent framework for undertaking and replicating the effectiveness of the research. The research design provides a road map for the study's overarching plan and strategy to advance the hierarchical Bayesian model. The framework incorporates specific procedures and techniques to identify and develop the desired study objectives. Therefore, the paper advances the hierarchical Bayesian model with a unique variable to the fixed effect and parameters to the random effects respectively. The paper adopted a theoretical design approach; highlighting how the introduction of a unique variable and parameters improve the model's theoretical grounding and possibly align it more closely. This improves the standard Bayesian model by adding a unique variable to the standard Bayesian hierarchical model and assessing the regularization. The theoretical design approach focuses on the conceptual justification that explores model structures conceptually without needing empirical validation thereby accounting for unobserved heterogeneity or the group-level effects. Also, the introduction of parameters ϕ_j, ψ_j, and X_{ijz} as a contextual factor are theoretically significant to be observed in the model. Normal statistical procedures for the determination of population, target population, sample, and sampling technique should be duly followed and researchers must ensure that individual-level and group-level selection are catered for. Techniques such as power analysis or sample size calculation can aid in the determination of optimal sample sizes when adopting empirical validation of the researchers' model. These techniques use different standard statistical parameters such as the effect size, the preferred level of statistical significance, and an estimation of the variability that is expected to be in the data. The rule of thumb for sample size determination in a hierarchical framework ensures that an adequate number of clusters and observations per cluster are present and information is fairly representative in each cluster. This rule of thumb states that at best, a minimum observation ought to be considered, a researcher should have 20 to 30 clusters (i.e. higher-level units) and at least 5 to 10 observations within each cluster. The numbers may vary based on the specific context of the research and the complexity of the model as well as the number of predictors considered in a cluster or unit, these guidelines ensure enough variability at both the individual and group levels to estimate model parameters reliably. The rule of thumb for sample size determination in a hierarchical framework is thus a general guideline proposed by methodologists and researchers wishing to apply the principles underlying multilevel modelling and statistics to applied research. While it is difficult to attribute this rule of thumb to a specific individual or group, it has emerged over time from a collective understanding of the statistical principles underlying multilevel modelling and the practical considerations of designing studies with clustered or hierarchical data structures. Methodologists and researchers alike have been consistent in recommending that a minimum number of clusters and observations per cluster is essential for estimating model parameters reliably and thereby reducing bias due to clustering. This guideline has been widely adopted and used in many different research contexts that utilize multilevel modelling. Mathematically, the rule is expressed as:

$$N \geq n_{clusters(j)} \text{ X } n_{observations(i)} = n_{j(20)} \text{ X } n_{i(5)} \quad \textbf{[4]}$$

Note that the minimum samples for individual - levels and clusters are postulated in Equation 4.

Model Selection

For model selection, researchers must consider adopting the Hierarchical Information Criterion (HIC). The hierarchical information criterion (HIC) reassesses the accuracy versus complexity trade-off to tackle the centralized hierarchical learning issue. Unlike classic information criteria such as AIC and BIC which disregard a hierarchical point of view, HIC considers nested structure by reducing model complication at each tier. Inclusively targeting all levels, it merges information standards while integrating adaptive penalizations corresponding to hierarchy gradation that combats overfitting via weighting level contribution based on intrinsic hierarchal intricacy.

$$HIC = -2\sum_{i=1}^{N} \log\{P(y_i|\theta_i)\} + k\sum_{j=1}^{J} w_j(\log N_j) \quad \textbf{[5]}$$

Where:

N is the total number of observations, $P(y_i|\theta_i)$ is the likelihood for observation i given parameters $\theta_{i,}$

k is a penalty factor for the model complexity,

J is the number of levels in the hierarchy

w_j is the weight associated with level j,

N_j is the number of observation at level j.

However, researchers are not obliged to employ such HIC but can decide to use any other model selection criteria such as Nested Deviance Information Criterion (nDIC), Hierarchy-Adjusted Cross-Validation (HACV), Clustered Bayesian Information Criterion (cBIC), and Bayesian Model Averaging (BMA) could be used in the empirical validation of the model that may account for the nested structures.

Original /Standard Hierarchical Bayesian Model

Below is the Standard Bayesian hierarchical model:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + u_{0j} + u_{0j} X_{ij1} + \varepsilon_{ij} \quad \textbf{[6]}$$

Where:

Y_{ij} is the dependent variable for observatjion i in group j

X_{ij1}, X_{ij2} are the predictor variables,
u_{0j}, u_{1j} are group − specific random effects,
$\beta_0, \beta_1, \beta_2$ are the fixed effects,
ε_{ij} is the individual − level error term.

Introduction of a Unique Variable to the Model

The introduction of a unique variable, X_{ijz}, where z represents the innovative variable capturing a unique aspect within each cluster. The new model then becomes:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \beta_3 X_{ijz} + u_{0j} + u_{0j} X_{ij1} + \varepsilon_{ij} \quad \textbf{[7]}$$

Where:
Y_{ij} is the dependent variable for observatjion i in group j
X_{ij1}, X_{ij2} are the predictor variables,
u_{0j}, u_{1j} are group − specific random effects,
$\beta_0, \beta_1, \beta_2$ β_3 are the fixed effects,
the introduction of X_{ijz} allows the model to capture cluster-specific effects associated with the innovative (z) aspect,
ε_{ij} is the individual − level error term.

New Methodology and Combine Hierarchical Model

A new methodology of incorporating a hierarchical Bayesian Shrinkage prior enhances the estimation of cluster-specific effects. The introduction of hierarchical Bayesian Shrinkage strengthens the estimation of the cluster-specific shrinkage effects, resulting in more robust and stable estimates (i.e., stabilizing the estimates). The unique variable and the new methodology together make a new and unique contribution of the Hierarchical Bayesian model to the field of knowledge by addressing particular substantive aspects of the research domain and enhancing the hierarchical modelling methodology. The substantive role of the unique variable is that the new variable is another predictor that captures something unique in the model (i.e. data). X_{ijz} captures the factor specific which differs to each observation in group j not captured by other predictors (X_{ij1} and X_{ij2}). The addition of the innovative variable in the model paves room to account for individual differences within each cluster that are not explained by the existing covariates. The coefficient of β_3 is the effect of the unique variable on the dependent variable, Y_{ij}. It tells us how a change in Y_{ij} is associated with a unit change in X_{ijz} when other variables are held constant. With the new methodology, the Hierarchical Bayesian Shrinkage model is now:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \beta_3 X_{ijz} + \phi_j u_{0j} + \psi_j u_{0j} X_{ij1} + \varepsilon_{ij} \quad \textbf{[8]}$$

Where:
Y_{ij} is the dependent variable for observatjion i in group j
X_{ij1}, X_{ij2} are the predictor variables,
u_{0j}, u_{1j} are group − specific random effects,
$\beta_0, \beta_1, \beta_2$ β_3 are the fixed effects,
ϕ_j, ψ_j are cluster-specific shrinkage parameters
the introduction of X_{ijz} allows the model to capture cluster-specific effects associated with the novel (z) aspect,

ε_{ij} is the individual − level error term.

Assumptions of Hierarchical Bayesian Shrinkage

Several assumptions are necessary for the estimates of the model to be reliable and valid. Assumptions of Hierarchical Bayesian Shrinkage include:

1. The model assumes a hierarchical structure, where parameters are organized into different levels or clusters (e.g., individual-level, individual, and cluster-level parameters). By allowing for the borrowing of information across groups, this assumption facilitates the generation of more consistent and standardized estimates - an essential valuable feature when there are insufficient observations within particular units or groups.
2. The clusters and levels in the hierarchy are assumed to be exchangeable, which implies that the statistical properties of one cluster are assumed to be identical to any other cluster. The exchangeability assumption is foundational for pooling information across clusters. Information sharing makes it possible to estimate group-specific parameters more accurately by borrowing strength from other clusters when data within the specific cluster is/are limited.
3. The data distributions of the random effects at the cluster level are assumed to be normally distributed; that is the assumption is necessary when using shrinkage methods (Gaussian shrinkage priors). Normality makes the model so easy to control and allows for closed-form solutions. It also facilitates the prediction and interpretation of the hierarchical shrinkage process.
4. The variance of the random effects is homogeneous across all clusters. The assumption of homogeneity is a simplifying device in the model and prevents the shrinkage procedures from overly favouring one particular cluster. This assumption can often be relaxed for more flexible models.
5. Each cluster has enough observations to inform the estimation of group-level parameters adequately. When the number of observations is too small in a cluster or unit, hierarchical shrinkage may not result in much improvement and regularisation; the estimates might be dominated by the prior.
6. Assuming independence among observations within each cluster and across the hierarchies is necessary for maintaining the meaningfulness of the hierarchical structure and accurate differentiation between variability within- versus between groups.

Assumptions Underlying Hierarchical Modelling

For hierarchical Bayesian modelling to perform well, it is essential not to ignore the underlying assumption which provides a concrete foundation. Avoiding the violation of the underlying assumption is by employing diagnostic tools to verify whether the model is devoid of bias estimates or inefficient parameter estimates. The following assumptions underlying hierarchical modelling need to be considered. The observations within each level of the hierarchy are assumed to be independent. This assumption is necessary for the model to accurately estimate the within-level and between-level variation by the researchers doing empirical validation. The random effects at each hierarchy level are assumed to be normally distributed; this assumption allows the model to estimate the mean and variance of the random effects. The relationships between the dependent and explanatory variables, at both the individual and the group levels, are assumed to be linear. This assumption means that the effects of the predictors of the outcome are additive. The variance of the dependent variable is assumed to be constant across different levels of the hierarchy. This assumption ensures that the model captures the variation in the outcome variable. The variance of the random effects at each hierarchy level is assumed to be constant. This assumption implies that the variation between groups is consistent across the hierarchy levels. The random effects at each level are assumed to be normally distributed with a mean of zero. The assumption allows the model to estimate the variation in the outcome variable attributable to the different levels of the hierarchy.

Addressing Potential Bias/Missing Values in the Model

It must be noted that implementing good strategies can minimize potential sources of bias in a model and enhance the reliability and validity of the hierarchical model framework. this can be achieved by ensuring more robust and accurate interpretation and inferences whenever data are involved. Conducting sensitivity analyses and exploring alternative model specifications can help to identify potential biases that may need to be addressed. Addressing missing values in the Bayesian hierarchical framework is essential to allow for the accuracy and robustness of the estimates which is an integral part. Missing values can be treated as unknown parameters (i.e latent variables) and make inferences to them alongside the model parameters by imputing the observed data distribution and its relationships with the Bayesian hierarchal framework. The mode will estimate the missing values iteratively by leveraging information from observed data within the cluster as well as other clusters due to information-sharing criteria in the hierarchical framework. When one's data follows the normal distribution pattern or the missing one, normal information priors can be specified allowing the model to sample plausible values based on prior knowledge and observed data. Whenever a dataset includes non-linear patterns or complex interactions, one must adopt the use of Predictive Mean Matching (PMM) where missing values are imputed by drawing from observed data points that pertain to the data distribution pattern.

Contributions to the Standard Bayesian Model (Findings)

This paper is expected to address Zong & Bradley's (2023) limitation by implementing a straightforward and relatively easy-to-implement model. The estimation procedure and computational algorithms is simplified to facilitate its use by a wider gamut of researchers and practitioners. The authors' model is flexible and adaptable to different datasets and modelling scenarios (e.g., national, urban, rural) and incorporate mechanisms that allow for sensitivity analysis and robustness checks to assess model performance under different conditions and assumptions.

The authors' model is expected to address Gómez-Méndez & Chainarong (2023) limitations by incorporating cluster-specific random effects (u_{0j}) and shrinkage parameters (ϕ_j, ψ_j). The model accounts for variations at the group level (clusters); this allows for capturing heterogeneity among different groups (e.g., regions, communities) in terms of their impact on the dependent variable (e.g., income, poverty-related variables). These parameters control the degree of shrinkage applied to the group-specific random effects, allowing for regularization and improved estimation. Introducing a novel predictor variable (X_{ijz}) allows the model to capture cluster-specific effects associated with z. This helps to identify unique characteristics or factors (e.g., educational programs, healthcare interventions) that may have differential effects on the dependent variable across different clusters. The article does not explicitly mention such an aspect. The model operates within a Bayesian hierarchical framework, which allows for the incorporation of prior information and the estimation of uncertainty. This framework provides a flexible and principled approach to modelling complex relationships and dependencies in a data.

Porter et al., (2022) limitation is expected to be addressed using the model by including both fixed effects (i.e., covariates) and spatial random effects (i.e., ϕ_j and ψ_j), allowing for the simultaneous selection of both spatial dependence and covariate effects, thus addressing the limitation of separately selecting fixed and spatial effects, as described in the article. With the inclusion of terms like X_{ij1}, the model suggests that social context variables like those used in the participant selection process of the article could be included. The inclusion of such variables can help to capture additional spatial dependencies and can potentially improve the accuracy of spatial modelling. The model allows for flexibility in the spatial structure by including spatial random effects as represented by ϕ_j and ψ_j. This flexibility is important as it allows for different spatial dependencies to be explored and to help overcome the challenge of spatial confounding discussed in the article.

The Bayesian hierarchical model developed by the authors overcomes the difficulties and shortcomings of the Standard Bayesian model by adopting a new approach that incorporates a unique variable (X_{ijz}) and interaction terms that explicitly account for unobserved heterogeneity and group-specific effects. This inclusion allows for a more accurate representation of the relationship between the dependent and the covariates especially in sparsely populated groups. Also, the model makes use of adaptive shrinkage priors (i.e. hierarchical shrinkage priors) that allow for more flexible regularization of random effects. These priors can be adjusted based on the magnitude and variability of the effects whilst improving on the parameter estimation and uncertainty quantification compared to traditional priors. Furthermore, by leveraging efficient sampling techniques, the model improves upon the scalability and computational feasibility of Bayesian hierarchical models, especially for high-dimensional data and complex hierarchical structures thereby enhancing transparency and interpretability. Lastly, the introduction of structured group-specific slopes and intercepts modulated by region-specific scaling factors (ϕ_j and ψ_j) random effects enhance the interpretability of the model by allowing an easy understanding of how the relationship between independent and dependent variables varies across different regions or groups.

Conclusion

The introduction of ϕ_j and ψ_j as scaling parameters have brought several improvements over the standard Bayesian hierarchical approach. The ϕ_j and ψ_j have been able to regularize the parameter estimates toward a common value, this model effectively mitigates overfitting and elevates the accuracy of estimates across diverse clusters. This innovative framework facilitates the reduction of variations within clusters while simultaneously drawing strength from other groups with high precision values thereby enhancing the degree of regularization applied to cluster-specific effects. The introduction of the parameters and the unique variable have brought advancement in the model characterized by superior predictive accuracy, enhanced interpretability and also offering clearer insights into the relationships among the variable. The model's ability to account for heterogeneity across clusters indicates that the specificity of each cluster is retained without compromising on the general structure. The model substantially advances the Bayesian hierarchical frameworks as it manages to incorporate parameter estimation with an admirable level of complexity, generalizability, and accuracy in the presence of high levels of cluster heterogeneity.

Recommendation

The following are recommended for further studies:

1. A comparative analysis should be conducted to verify whether indeed the current model outperformance the Standard Bayesian model with real-world datasets in the hierarchical framework.

2. A simulated study could be carried out to find which of the two models promotes more regularisation
3. Other model selection methods like Nested Deviance Information Criterion (nDIC), Hierarchy-Adjusted Cross-Validation (HACV), Clustered Bayesian Information Criterion (cBIC), and Bayesian Model Averaging (BMA) could be used in the empirical validation of the model

Reference

1. Aczel, B., Hoekstra, R., Gelman, A., Wagenmakers, E.-J., Kluglist, I. G., Rouder, J. N., et al. (2018). Expert opinions on how to conduct and report Bayesian inference. https://doi.org/10.31234/osf.io/23m7f
2. Agyemang, S., & Asibey, B. O. (2018). Effect of Education on health care utilization in rural Ghana: The case of selected communities in the Bekwai Municipality. KNUST Journal of Geography and Development, 114.
3. Ahmed T, Roberton T, Vergeer P, Hansen PM, Peters MA, Ofosu AA, et al. (2022) Healthcare utilization and maternal and child mortality during the COVID-19 pandemic in 18 low- and middle-income countries: An interrupted time-series analysis with mathematical modeling of administrative data. PLoS Med 19(8): e1004070. https://doi.org/10.1371/journal.pmed.1004070
4. Arijit Chakrabarti, Jayanta K. Ghosh, (2011): Philosophy of Statistics. Volume 7 in Handbook of the Philosophy of Science.
5. Auguie, B. (2017). gridExtra: Miscellaneous functions for "grid" graphics. Retrieved from https://CRAN.R-project.org/packa ge=gridExtra
6. Aust, F., & Barth, M. (2018). Papaja: Create APA manuscripts with R markdown. Retrieved from https://github.com/crsh/papaja
7. Bates, D., & Maechler, M. (2019). Matrix: Sparse and dense matrix classes and methods. Retrieved from https://CRAN.R-project. org/package=Matrix
8. Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01
9. Betancourt, M. (2018). A conceptual introduction to Hamiltonian Monte Carlo. arXiv PreprinarXiv:1701.02434.
10. Cheru Atsmegiorgis Kitabo and Ehit Tesfu Damtie (2020). Bayesian Multilevel Analysis of Utilization of Antenatal Care Services in Ethiopia. Computational and Mathematical Methods in Medicine Volume 2020, Article ID 8749753, 11 pages https://doi.org/10.1155/2020/8749753
11. Danaher, T. S., Sweeney, J. C., & Johnson, L. W. (2007). A Hierarchical Model of Health Service Quality: Scale Development and Investigation of an Integrated Model. Journal of Service Research, 10(2), 123-142. DOI: 10.1177/1094670507309594
12. D'Errigo, P., Tosti, M. E., Fusco, D., Perucci, C. A., & Seccareccia, F. (2007). Use of hierarchical models to evaluate the performance of cardiac surgery centers in the Italian CABG outcome study. BMC Medical Research Methodology, 7(29). doi:10.1186/1471-2288-7-29