Behaviormetrika
https://doi.org/10.1007/s41237-025-00264-7

ORIGINAL PAPER

Estimating context effects in small samples while controlling for covariates: an optimally regularized Bayesian estimator for multilevel latent variable models

Valerii Dashuk[1] · Martin Hecht[2] · Oliver Lüdtke[3,4] · Alexander Robitzsch[3,4] · Steffen Zitzmann[1]

Received: 18 December 2024 / Accepted: 11 June 2025

Abstract

In this article, we extend the regularized Bayesian estimator of multilevel latent variable models (Dashuk et al. 2024) to improve the estimation of the between-group parameter β_b in two-level latent variable models with covariates. Specifically, we allow for the inclusion of covariates and optimize the estimator's accuracy in terms of the Mean Squared Error (MSE). Simulation results indicate that the regularized Bayesian estimator consistently outperforms both standard and transformed maximum likelihood (ML) estimators, especially in scenarios with small sample sizes and low intraclass correlations. While the estimator achieves lower Root Mean Square Error (RMSE) values and relative bias, it exhibited poorer coverage rates and less reliable standard errors compared to ML estimators. To address these limitations, we propose some strategies. Furthermore, a transformed ML estimator is utilized to enhance the accuracy of the estimation of the covariate parameter γ. Finally, we provide a step-by-step tutorial demonstrating how to apply the extended regularized Bayesian estimator to a real dataset using the `MLOB` package in R.

Keywords Regularized estimation · Multilevel model · Latent variable model · Mean squared error (MSE) · Small sample · Low intraclass correlation (ICC) · Bayesian multivariate model · Covariates

---

Communicated by Masaki Uto.

✉ Martin Hecht
martin.hecht@hsuhh.de

Valerii Dashuk
Valerii.Dashuk@research-development-innovation.de

1 Department of Psychology, Medical School Hamburg, Hamburg, Germany

2 Institute of Psychology, Helmut Schmidt University, Hamburg, Germany

3 Leibniz Institute for Science and Mathematics Education (IPN), Kiel, Germany

4 Centre for International Student Assessment, Kiel, Germany

1 Introduction

Bayesian approaches for estimating multilevel models have become increasingly popular in psychology, the education sciences, and related fields. These methods offer solutions to common challenges by incorporating prior information, which can enhance the accuracy and stability of parameter estimates (Hamaker and Klugkist 2011; Lüdtke et al. 2013; Muthén and Asparouhov 2012; Zitzmann et al. 2016). Their growing use has been facilitated by the availability of user-friendly software, such as Mplus, which support Bayesian estimation and provide accessible tools for researchers (Asparouhov and Muthén 2021a, b; Morin et al. 2022; Zitzmann et al. 2024)

Incorporating prior information is fundamental to Bayesian estimation. It can improve the accuracy of estimates, provided the prior is suitably specified. However, its effectiveness depends on factors such as sample sizes and intraclass correlation (ICC). This integration is especially critical in small samples with low ICCs—a metric indicating the proportion of variance attributable to group-level effects, a common issue in resource-constrained settings. In such cases, the careful selection of prior information plays a pivotal role in determining the accuracy of the estimates (Natarajan and Kass 2000; Zitzmann et al. 2015).

Recent studies have highlighted the importance of constructing "thoughtful" priors to improve estimation accuracy (Smid et al. 2020; Zitzmann et al. 2021b). While addressing small-sample bias is critical, it is equally essential to consider the estimator's variability. Methods that effectively reduce the Mean Squared Error (MSE) by balancing bias and variability yield more accurate estimates (Greenland 2000; Zitzmann et al. 2015).

One such approach was proposed by Dashuk et al. (2024), building on the model introduced by Lüdtke et al. (2008) and further developed by Zitzmann et al. (2021b). Dashuk et al. (2024) introduced a regularized Bayesian estimator for estimating between-group effects in multilevel latent variable models. This estimator automatically selects priors optimized for minimizing MSE, thereby reducing the risk of user misspecification. However, their approach was limited to models with a single regressor, which restricts its applicability to real-world hierarchical data.

In this article, we extend their estimator. The new estimator incorporates covariates for a more comprehensive and accurate estimation of between-group effects. By explicitly adjusting for the influence of the covariates, our method improves estimation efficiency while minimizing MSE. We demonstrate that the extended estimator substantially outperforms ML estimation, particularly in small samples and when ICCs are low.

It is important to clarify that the term "extended estimator" refers not to a change of the estimation procedure itself, but to the broader application of the previously proposed Bayesian regularized estimator (Dashuk et al. 2024). Specifically, we generalize the estimator to a more general model that includes measured covariates. The original model—without covariates—is a special case of the one presented here, and the estimator reduces exactly to the earlier version when no covariates are present. Thus, our approach presented here broadens the scope of the estimator to a much wider class of models.

By incorporating covariate effects, our method enables a more nuanced interpretation of contextual influences in multilevel models. The inclusion of covariates is standard practice in applied research across many disciplines. For example, covariate adjustment is widely applied in econometrics (Wooldridge 2009), causal inference (Angrist and Pischke 2008), multilevel structural equation modeling (Gelman and Hill 2007), and in educational and social research more broadly. We illustrate and highlight this relevance in the tutorial section of the paper, comparing models with and without covariates using real-world data.

To evaluate the estimator's performance, we conduct extensive simulations under various conditions, such as small sample sizes and low ICCs, demonstrating its practical advantages for multivariate multilevel data. Beyond simulations, we present a Step-by-Step Tutorial with a real-world dataset, in which we directly compare the estimator proposed by Dashuk et al. (2024) with our extended version. This empirical example illustrates how the proposed extension generalizes the original approach and confirms its applicability in real-world hierarchical data.

2 Extended Bayesian approach

To introduce the extension of the regularized Bayesian estimator, whose univariate form was proposed by Dashuk et al. (2024), we begin by summarizing the foundational work by Lüdtke et al. (2008). Their model, also known as the multilevel latent covariate model, provides an approach for unbiased estimation of between-group slopes in contextual studies. In this model, the dependent variable Y at the group level is predicted using a latent variable that represents the group's latent mean—a more robust and accurate alternative to the standard manifest mean. Moreover, this model also supports the integration of latent group means into the broader multilevel modeling framework, widely applied in psychological, educational, and related research.

The extended regularized estimator builds up on the univariate version introduced by Dashuk et al. (2024, see Zitzmann et al. 2021b, for a similar approach). This original estimator provides a solid foundation for estimating relationships between variables with a multilevel structure and is particularly efficient in estimating between-group effects. However, its applicability was initially limited to models with a single predictor. To address this limitation, we extend Dashuk et al.'s (2024) estimator in this study to incorporate multiple covariates, enabling more comprehensive and flexible analyses.

Our extended estimator retains the regularization techniques from Dashuk et al. (2024) but adapts them to address the additional complexity arising with multiple covariates. These theoretical and computational modifications enable the new Bayesian estimator to more accurately capture relationships between variables in the presence of covariates. Consequently, the extended estimator broadens its applicability to a wider range of multilevel studies.

2.1 Mathematical formulation

In extending the regularized Bayesian estimator, we retained the original model notation. Specifically, the model assumes that the individual-level predictor X is decomposed into two independent, normally distributed components: X_b, representing the latent group mean, and X_w, representing individual deviations from X_b. For an individual $i = 1, \ldots n$ within a group $j = 1, \ldots J$, this decomposition is expressed as:

$$\begin{aligned} X_{ij} &= X_{b,j} + X_{w,ij} \\ X_{b,j} &\sim N(\mu_X, \tau_X^2) \\ X_{w,ij} &\sim N(0, \sigma_X^2). \end{aligned} \tag{1}$$

The same assumption applies to the latent group mean and individual deviations for the dependent variable Y and a vector of k covariates $C = (C_1, C_2, ...C_k)$:

$$\begin{aligned} Y_{ij} &= Y_{b,j} + Y_{w,ij} & C_{ij} &= C_{b,j} + C_{w,ij} \\ Y_{b,j} &\sim N(\mu_Y, \tau_Y^2) & C_{b,j} &\sim N_{mult}(\mu_C, \tau_C^2) \\ Y_{w,ij} &\sim N(0, \sigma_Y^2) & C_{w,ij} &\sim N_{mult}(0, \sigma_C^2). \end{aligned} \tag{2}$$

In addition, we assume that all between-group components—namely, X_b, Y_b, and C_b—are independent of their corresponding within-group components—X_w, Y_w, and C_w.

We also assume that the sample is balanced, with each of the J groups containing n persons, thus resulting in an overall sample size of nJ.

Subsequently, we will refer to σ_X^2, σ_Y^2, σ_C^2 and τ_X^2, τ_Y^2, τ_C^2 as the within-group and between-group variances of X, Y, and C, respectively.

Note that, in contrast to the scalar variables X_{ij} and Y_{ij}, all covariates are combined into a vector C_{ij} from $\mathbb{R}^k$. The vector form of C_{ij} accounts for the multivariate nature of covariates, enabling simultaneous modeling of multiple effects. Consequently, the vector form of $\mu_C \in \mathbb{R}^k$ and the variance matrices $\tau_C^2 \in \mathbb{R}^{k \times k}$, $\sigma_C^2 \in \mathbb{R}^{k \times k}$ arise, leading to multivariate normal distributions for both C_b and C_w as well. In the special case where $k = 0$, the model reduces to the version without covariates from Lüdtke et al. (2008; see also Dashuk et al. (2024); Zitzmann et al. 2021b).

Following common notation and incorporating covariates, the individual-level and group-level regression equations are as follows:

$$\text{Level 1: } Y_{ij} = \beta_{0j} + \beta_w X_{w,ij} + C_{ij}\gamma + \varepsilon_{ij} \tag{3}$$

$$\text{Level 2: } \beta_{0j} = \alpha + \beta_b X_{b,j} + \delta_j. \tag{4}$$

By decomposing the covariates (C_{ij}) into their between-group ($C_{b,j}$) and within-group ($C_{w,ij}$) components, the model can be equivalently expressed as:

$$\text{Level 1: } Y_{ij} = \beta_{0j} + \beta_w X_{w,ij} + C_{w,ij}\gamma + \varepsilon_{ij} \tag{5}$$

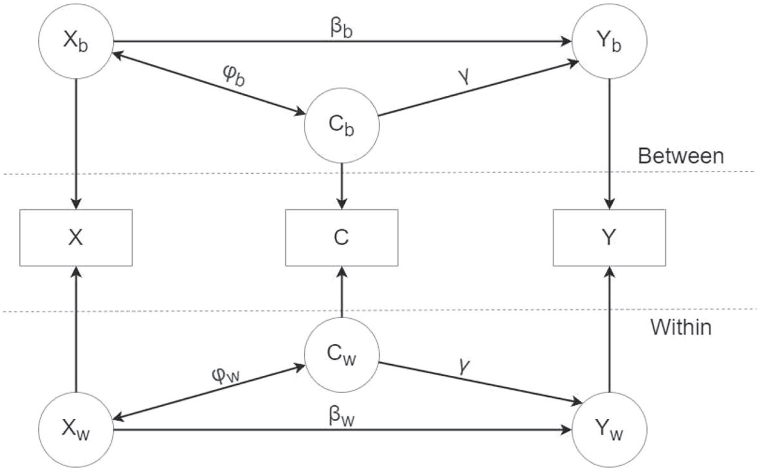

Fig. 1 A multilevel structural equation model using the within-between framework that decomposes the variables X, C, and Y into within-group and between-group components. ***Note.*** The within-group components are denoted by subscript w, and the between-group components are denoted by subscript b. The between-group components (X_b, C_b, and Y_b) are connected through a standard structural equation model, where Y_b serves as the dependent variable, X_b as the predictor, and C_b as covariate. Similarly, the within-group components (X_w, C_w, and Y_w) are related to each other in an analogous manner. The notation includes β_b for the between-group slope of X, β_w for the within-group slope of X, and γ for the slopes of C at both the between-group and within-group levels. Additionally, the relationship between the predictor and covariate is presented by ϕ_b and ϕ_w for the between-group and within-group components, respectively

$$\text{Level 2: } \beta_{0j} = \alpha + \beta_b X_{b,j} + C_{b,j}\gamma + \delta_j. \tag{6}$$

Both representations of the model are mathematically equivalent; however, the formulation in Eqs. 5 and 6 explicitly distinguishes between within-group and between-group components of the covariates, making it more suitable for explaining and showcasing the extension procedure. For the remainder of this article, we will refer to the representation in Eqs. 3 and 4, unless otherwise specified.

In Eqs. 3 and 5, β_w represents the within-group slope, describing the relationship between the predictor X and the dependent variable Y at the individual level, while β_{0j} is the intercept for each group j. The residuals, denoted as $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$, are assumed to be normally distributed.

The parameter vector $\gamma \in \mathbb{R}^k$ represents the relationship between the covariates C and the dependent variable Y. It is important to note that we assume identical between-group, within-group, and overall effects of the covariates; hence, γ remains the same in Eqs. 3–6. Therefore, the covariates C are used primarily to enhance the estimation of the between-group effect of the predictor X.

In Eqs. 4 and 6, α represents the overall intercept, β_b denotes the between-group slope between the predictor X and dependent variable Y, and $\delta_j \sim N(0, \tau_\delta^2)$ indicates the normally distributed residuals. A graphical representation of this model is shown in Fig. 1.

2.2 Transformation for covariate adjustment

To work with the model in Eqs. 3 and 4 in the same way as Dashuk et al. (2024), we transform the model into one without covariates. This transformation allows us to apply the regularized Bayesian estimation that minimizes the MSE of the between-group estimator for β_b. To make this transformation, we first regress the covariates on the between-group and within-group components of the predictor X. The regression model is:

$$C_{ij,k} = \phi_0 + \phi_b X_{b,j} + \phi_w X_{w,ij} + \psi_{1,ij}, \tag{7}$$

where the parameters ϕ_b and ϕ_w represent the relationships between the predictor and the covariate at the between-group and within-group levels, respectively, as shown in Fig. 1. Meanwhile, ϕ_0 represents the intercept, and $\psi_{1,ij}$ denotes the regression error term.

It is important to note that in Eq. 7, X_b represents a latent variable, meaning that it is not directly observed but can only be inferred from the observed X. On the one hand, X_w too could be viewed as latent, as it is the result of the decomposition of X into unobserved parts. On the other hand, however, the model does not employ any error correction when estimating the effect of X_w, and we know from our previous works that the result for this effect is identical to the result from the traditional manifest model without latent variables. In other words, X_w seems to be closely linked to X, raising doubts whether X_w can be considered truly latent. Irrespective of this ambiguity, both X_b and X_w can be used in regression analysis due to their well-defined covariances and variances (see Zitzmann et al. 2021b). For a single regressor, OLS estimation of the effects of X_b and X_w relies on their covariances with Y and their variances, and this principle extends to multiple independent regressors. Since X_b and X_w are constructed to be independent, the estimation in Eq. 7 can be computed by using the sample covariances and variances.

The estimated values of $C_{ij,k}$ are:

$$\hat{C}_{ij,k} = \hat{\phi}_0 + \hat{\phi}_b X_{b,j} + \hat{\phi}_w X_{w,ij}. \tag{8}$$

As the model in Eqs. 3 and 4 can include any number of covariates, we group all estimated values from Eq. 8, calculated for each of the k covariates, into a single vector $\hat{C}_{ij}$. The combined residuals are then:

$$\tilde{C}_{ij} = C_{ij} - \hat{C}_{ij}. \tag{9}$$

To obtain estimates of the parameters in Eq. 7, ML can be used, which simplifies to Ordinary Least Squares (OLS) under the assumption of normality and balanced sample.

The obtained residuals $\tilde{C}_{ij}$ in Eq. 9 serve as new covariates with the between-group and within-group effects of the predictor X eliminated. These residuals are then used to regress the dependent variable Y on them, thereby isolating the effect of the covariates on the dependent variable:

$$Y_{ij} = \gamma_0 + \tilde{C}_{ij}\gamma + \psi_{2,ij}, \tag{10}$$

where γ_0 represents the intercept, and γ reflects the same between-group and within-group slopes of C as in Equations 3–6 and Fig. 1 due to the removed effects of X_b and X_w. Additionally, $\psi_{2,ij}$ denotes the regression error term.

The parameter estimate $\hat{\gamma}$ is obtained from this model using ML estimation. Using $\hat{\gamma}$ (i.e., the ML estimate), we define $\hat{Y}_{ij}$ as a linear combination of the k covariates as:

$$\hat{Y}_{ij} = C_{ij}\hat{\gamma}. \tag{11}$$

We omitted the estimated intercept $\hat{\gamma}_0$ in Eq. 11, because it has no influence on the estimated slopes in the target model in Eqs. 3 and 4 and only affects the global intercept α in Eq. 4.

Using the estimated value of Y from Eq. 11, we define $\tilde{Y}_{ij}$ as the difference between Y_{ij} and its estimated value $\hat{Y}_{ij}$:

$$\tilde{Y}_{ij} = Y_{ij} - C_{ij}\hat{\gamma}. \tag{12}$$

In the new variable $\tilde{Y}_{ij}$ there is no variability associated with the covariates C_{ij}. As a result, $\tilde{Y}_{ij}$ can substitute for Y_{ij} in the two-level model from Eqs. 3 and 4, simplifying the model to one with a single predictor:

$$\text{Level 1: } \tilde{Y}_{ij} = \beta_{0j} + \beta_w X_{w,ij} + \varepsilon_{ij} \tag{13}$$

$$\text{Level 2: } \beta_{0j} = \beta_b X_{b,j} + \delta_j. \tag{14}$$

The resulting model allows for the estimation of the between-group effect β_b by using the procedure previously described by Dashuk et al. (2024).

However, to use $\tilde{Y}_{ij}$ in the same manner as Y_{ij}, we need the distribution of $\tilde{Y}_{ij}$. To obtain this distribution, we define the covariances between Y and C as:

$$\tau_{YC} = \mathrm{Cov}(Y_{b,j}, C_{b,j}) \tag{15}$$

$$\sigma_{YC} = \mathrm{Cov}(Y_{w,ij}, C_{w,ij}). \tag{16}$$

According to Eqs. 1 and 2, both the between-group and within-group components of Y and C are normally distributed. Additionally, since $\hat{\gamma}$ is a fixed value for each specific model, both the between-group and within-group components of $\tilde{Y}_{ij}$ are also normally distributed as they are linear combinations of normally distributed random variables. This enables us to state the distributions of the between-group and within-group parts of $\tilde{Y}_{ij}$ as:

$$\begin{aligned}\tilde{Y}_{ij} &= \tilde{Y}_{b,ij} + \tilde{Y}_{w,ij} \\ \tilde{Y}_{b,j} &\sim N(\mu_Y - \mu_C\hat{\gamma}, \tau_Y^2 - 2\tau_{YC}\hat{\gamma} + \hat{\gamma}'\tau_C^2\hat{\gamma}) \\ \tilde{Y}_{w,ij} &\sim N(0, \sigma_Y^2 - 2\sigma_{YC}\hat{\gamma} + \hat{\gamma}'\sigma_C^2\hat{\gamma}).\end{aligned} \tag{17}$$

See Appendix A for the derivations. To be able to apply the ML and the regularized Bayesian estimators from Dashuk et al. (2024), we also confirmed the independence of the between-group and within-group components of the new dependent variable $\tilde{Y}$ and the predictor X in Appendix A.

2.3 Regularized Bayesian estimation with covariates

To derive the target estimator for the between-group slope β_b in the model specified by Eqs. 13 and 14, we first outline the formulas for the ML estimator and the regularized Bayesian estimator for the model without covariates, as described in Dashuk et al. (2024). The ML estimator is simply:

$$\hat{\beta}_b = \frac{\hat{\tau}_{YX}}{\hat{\tau}_X^2}. \tag{18}$$

The regularized Bayesian estimator is:

$$\tilde{\beta}_b = \frac{\hat{\tau}_{YX}}{(1-\omega)\tau_0^2 + \omega\hat{\tau}_X^2}, \tag{19}$$

where ω and τ_0^2 are the prior parameters that are used to minimize the MSE of β_b and can be thought of as the optimal weighting parameter and the optimal prior guess, respectively (Zitzmann et al. 2021a, b; Dashuk et al. 2024). Zitzmann et al. (2021b) outlined this estimator, and Dashuk et al. (2024) have regularized it with respect to ω and τ_0^2 in order to reach the minimal MSE. This regularized Bayesian estimation procedure is provided in Appendix B.

Our two-level model (Eqs. 13 and 14) utilizes $\tilde{Y}$ instead of Y. Consequently, for estimation purposes, it is necessary to determine the between-group covariance between X and $\tilde{Y}$. We begin with the covariances between the between-group components of X with Y and C:

$$\begin{aligned}\mathrm{Cov}(Y_b, X_b) &= \tau_{YX} \\ \mathrm{Cov}(C_b, X_b) &= \tau_{CX}.\end{aligned} \tag{20}$$

Then, the between-group covariance of X and $\tilde{Y}$ is:

$$\tau_{\tilde{Y}X} = \mathrm{Cov}(\tilde{Y}_b, X_b) = \mathrm{Cov}(Y_b - C_b\hat{\gamma}, X_b) = \tau_{YX} - \hat{\gamma}'\tau_{CX}. \tag{21}$$

This result provides the covariance between the between-group components of $\tilde{Y}$ and X, using the given covariances of the original variables. With this, we can define the new estimators for β_b.

An ML estimator for the model with covariates is:

$$\hat{\beta}_b = \frac{\hat{\tau}_{\tilde{Y}X}}{\hat{\tau}_X^2} = \frac{\tau_{YX} - \hat{\gamma}'\tau_{CX}}{\hat{\tau}_X^2}. \tag{22}$$

To clarify, we will refer to ML applied to the original model in Eqs. 3 and 4 as the standard ML, and ML applied to the transformed model in Eqs. 13 and 14 as the transformed ML. Accordingly, the transformed ML estimator is defined by Eq. 22.

The regularized Bayesian estimator for this model is:

$$\tilde{\beta}_b = \frac{\hat{\tau}_{\tilde{Y}X}}{(1-\omega)\tau_0^2 + \omega\hat{\tau}_X^2} = \frac{\tau_{YX} - \hat{\gamma}'\tau_{CX}}{(1-\omega)\tau_0^2 + \omega\hat{\tau}_X^2}. \tag{23}$$

The transformed ML estimator in Eq. 22 is calculated directly after the transformation, whereas the regularized Bayesian estimator in Eq. 23 is implemented post-transformation to identify the optimal $\tilde{\beta}_b$ that minimizes MSE, as described by Dashuk et al. (2024). The complete procedure, including the computation of standard errors for the regularized Bayesian estimation is detailed in Appendix B. Note that the transformation procedure results in slightly larger standard errors because $\tilde{Y}$ is used instead of Y.

Following the Bayesian estimation procedure, the optimal values for the parameters ω and τ_0^2 are determined as those that minimize the MSE of $\tilde{\beta}_b$ as computed from its posterior distribution:

$$(\omega, \tau_0^2)_{\mathbf{optimal}} = \arg\min_{\omega, \tau_0^2}(MSE(\tilde{\beta}_b(\omega, \tau_0^2))). \tag{24}$$

The minimization algorithm is provided in Appendix B. The optimal parameter values define the extended optimally regularized Bayesian estimator $\tilde{\beta}_b$ in Eq. 23.

The key distinction between the formulas for the regularized Bayesian estimator from Eq. 19, introduced in Dashuk et al. (2024), and the extended regularized Bayesian estimator from Eq. 23 lies in the inclusion of the term $\hat{\gamma}'\tau_{CX}$ in the extended form. This term explicitly accounts for the effect of covariates in the between-group estimation by adjusting for the shared variance between X and C at both the between-group and within-group levels. By introducing $\hat{\gamma}'\tau_{CX}$, we extend the estimator to directly adjust for covariate effects, improving the accuracy of the regularized Bayesian estimator when covariates are present. Note that when the covariates do not correlate with the regressor X at the between-group level ($\tau_{CX} = 0$), or when they have no explanatory power for the dependent variable Y ($\gamma = 0$), the extended regularized Bayesian estimator from Eq. 23 simplifies to the regularized Bayesian estimator proposed by Dashuk et al. (2024) in Eq. 19.

This refinement enhances the applicability of the estimator in practical multilevel modeling scenarios where covariates play a crucial role in explaining between-group variability.

For computations and simulations, first part of Eqs. 22 and 23 can also be utilized, as $\hat{\tau}_{\tilde{Y}X}$ can be directly computed from $\tilde{Y}$ using the procedure outlined above (Eqs. 7–12). Since all properties of $\tilde{Y}$ are derived in Appendix A, this allows for straightforward calculations.

3 Simulation study

In this section we conducted an extensive simulation study in which the performance of an extended regularized Bayesian estimator is examined in comparison to the ML methods. That would allow us to establish how practically applicable and effective this estimator is in real-world multilevel modeling situations. We start by outlining the data generation process, including its constraints and parameters, such as group size (n), number of groups (J), ICC values, β_b, β_w, and γ. Using the generated data, we compute our extended Bayesian estimator as well as the standard ML and the transformed ML estimators. Simulations are conducted with 5,000 replications for each unique set of data-generating parameters. The results and performance measures are presented in both table form (see Appendix C and Supplement A) and graphical form, facilitating a comprehensive evaluation of the estimation accuracy across varying conditions.

The simulation procedure consists of three main steps. First, data are generated for 540 scenarios, each with 5,000 replications, following predefined parameter settings as described in the Data Generation subsection. Second, model estimation is conducted by applying the extended regularized Bayesian estimator, the standard ML estimator, and the transformed ML estimator, following the procedure detailed in Appendix B. Third, evaluation metrics are computed to access estimators' performance using MSE, relative bias, coverage rate, and standard error ratio, as described in the Evaluation Criteria subsection. The simulations were conducted using MATLAB R2023a on a system with an AMD Ryzen 5 7539U processor and 32 GB RAM. To enhance reproducibility, we provide a MATLAB script in Supplement B which includes all steps of our simulation framework. The script is available as supplementary material on the journal web page.

3.1 Data generation

The following subsection provides a detailed description of the data generation process, including a step-by-step outline of the simulation setup. Our data-generating process closely follows the approaches implemented by Zitzmann et al. (2021a) and Zitzmann et al. (2021b). Simulations were conducted for various combinations of parameters. The intraclass correlation of the predictor (ICC_X) took values of 0.05, 0.1, 0.3, and 0.5. The number of covariates (k) was fixed at 1, while the intraclass correlation of the covariate ICC_C was fixed at 0.4, and the intraclass correlation of the dependent variable (ICC_Y) was fixed at 0.3. The number of groups (J) varied across values of 5, 10, 20, 30, and 40, while the number of individuals per group (n) was set to 5, 15, and 30. The between-group parameter (β_b) took values of 0.2, 0.5, and 0.6, and the within-group parameter (β_w) took values of 0.2, 0.5, and 0.7. Lastly, the covariate parameter (γ) was fixed at 0.3.

In total, this setup resulted in $4 * 5 * 3 * 3 * 3 = 540$ conditions, with data generated 5,000 times for each condition. In our implementation, the covariance between the predictor and the covariate was preset to 0.4 within the code but can be modified as an input parameter along with the number of covariates, their parameters, and the ICCs. For the simulations, the number of covariates was fixed at 1 to avoid excessive model complexity and computational bulkiness.

While the proposed estimator extends previous work by incorporating multiple covariates, the simulations were conducted with a single additional covariate ($k = 1$) and fixed values for ICC_C, ICC_Y, and γ due to computational constraints. We recognize that different values and the number of covariates could influence the results. However, running simulations with multiple covariates and varying parameter values would significantly increase computational time, making large-scale validation infeasible. To give an example, simulations with just one covariate required approximately six weeks of continuous computation on an AMD Ryzen 5 7539U processor with 32 GB RAM. Expanding these simulations to multiple covariates and parameter variations would lead to an exponential increase in runtime. However, this restriction applies only to the simulated experiments and does not limit the method's applicability to multivariate settings. To demonstrate the method‘s capability beyond the specific simulation conditions, we provide an empirical validation using a real-world dataset, presented in the Step-by-Step Tutorial section.

3.2 Evaluation criteria

The performance of our new optimally regularized Bayesian estimator was evaluated using several accuracy and bias measures as detailed below. We compared this estimator to the standard ML estimator applied directly to the original model (Eqs. 3 and 4) and the transformed ML estimator applied to the model in Eqs. 13 and 14. While the primary target parameter was β_b, we also report results for the covariate parameter γ.

As performance criteria, we assessed Root Mean Squared Error (RMSE), Relative Bias, Coverage Rate (CR), and Standard Error Ratio (SER).

RMSE, as the square root of MSE, is easier to interpret in terms of magnitude, making it a more intuitive measure for comparing the accuracy of estimators.

Relative bias, on the other hand, quantifies the average deviation of the estimated parameters from their true values and is expressed as a proportion of the true value.

CR represents the proportion of replications in which the true parameter value fell within the 95% confidence interval estimated by the method. Confidence intervals were constructed using t-values with the corresponding degrees of freedom and making use of the normality assumption. A near-nominal CR indicates that the confidence intervals consistently captured the true parameter value at the expected frequency, suggesting that the estimator provides well-calibrated intervals.

The SER is defined as the ratio of the average standard error estimated by the method to the empirical standard deviation of the parameter estimates across replications. SER serves as a measure for comparing the variability of the estimates with the variability expected under the model assumptions. Ideally, an SER value close to 1 indicates that the estimator provides an accurate quantification of uncertainty. However, in our case, we expect SER values to exceed 1, as the methods are evaluated under small sample sizes, where uncertainty is typically not very well estimated.

The selected performance measures—RMSE, relative bias, CR, and SER—provide a comprehensive assessment of accuracy, bias, and uncertainty calibration, respectively. Together, these metrics ensure that the Bayesian and ML estimators are rigorously evaluated across a variety of conditions.

3.3 Results

We now present the results of our simulation study, focusing on two primary objectives: (1) to analyze the characteristics of the extended regularized Bayesian estimator, and (2) to compare these findings to both theoretical expectations and the performance of the ML estimators. To facilitate interpretation, Figures 2, 3, 4, 5 and 6 provide visual analyses of the behavior of the between-group parameter estimators and covariate estimators as functions of the group-level sample size (J). In Figs. 2, 3, 4 and 6, the regularized Bayesian estimator is depicted by a blue dashed line, the transformed ML by a black dot-dashed line, and the standard ML by a green solid line. For enhanced clarity and differentiation between methods, the logarithm of the RMSE is plotted in Fig. 2. Notably, while logged RMSE was used solely for improved visualization in Fig. 2, the differences in RMSE between the regularized Bayesian and ML estimators can span up to four orders of magnitude.

The results of our simulation study are detailed in Tables S1–S32 in Supplement A. These tables summarize the performance of both Bayesian and ML approaches in estimating the between-group parameter (β_b) and the covariance parameter (γ) across all estimation quality indicators: RMSE, RB, CR, and SER, as defined in the Evaluation Criteria subsection. Tables S1–S4 present the RMSE of the estimated between-group parameter without log transformations, providing a clear depiction of the performance differences between estimators. Tables S5–S8 report the RB of the estimated between-group parameter, while Tables S9–S12 summarize the CR, and Tables S13–S16 display the SER for β_b. Tables S17–S20 focus on the RMSE of the estimated γ, with Tables S21–S24 reporting the RB, Tables S25–S28 summarizing the CR, and Tables S29–S32 presenting the SER for γ.These measures provide comprehensive insights into the accuracy of each estimator across 540 conditions. To improve practical applicability, we additionally provide Tables 1, 2, 3 and 4 in Appendix C, which summarize the previously mentioned results by averaging across n and J. This aggregation can be particularly valuable for model selection in real-world data analysis, where key parameters (e.g., ICC_X and β_w) are often unknown.

Figure 2 illustrates the logarithm of RMSE for all three estimators. The y-axis displays log-transformed RMSE values to enhance visualization, given the substantial range of RMSE values. The four subplots, corresponding to different ICC levels,

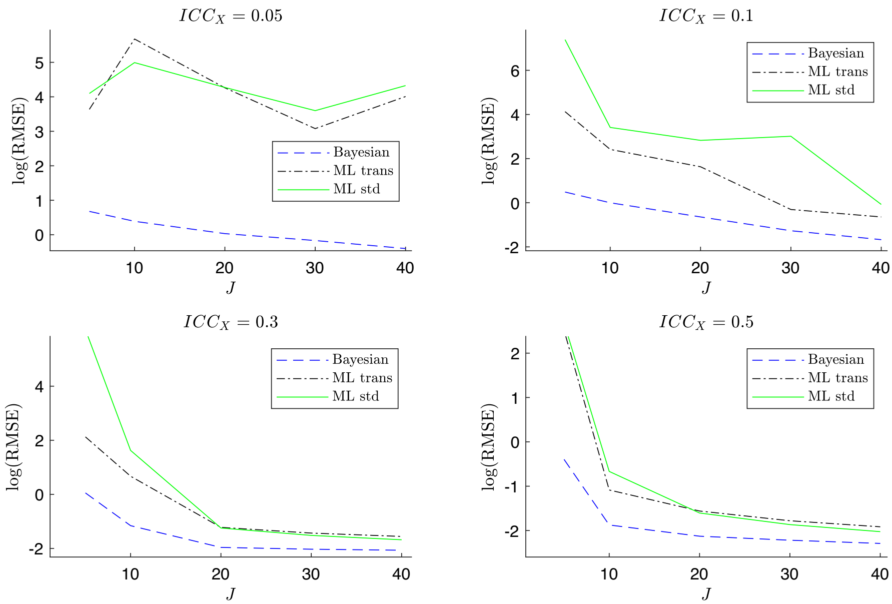

Fig. 2 Log of RMSE in estimating the between-group slope β_b for the regularized Bayesian, transformed, and standard ML estimators as a function of the group-level sample size (J). ***Note.*** Caution is advised, as the log transformation distorts certain relationships with respect to the original statistic (RMSE), allowing only ordinal (ranking) information to be inferred. It prevents accurate assessment of effect sizes between methods or along the x-axis. Additionally, note that the y-axis scale differs across the four subplots. Results are shown for $n = 15$ persons per group, with a within-group slope $\beta_w = 0.5$, a between-group slope $\beta_b = 0.2$, and varying levels of ICC_X

indicate that the regularized Bayesian estimator outperforms both ML estimators, particularly when the group-level sample size (J) is small (5 to 20) and ICC levels are low. Whereas the performance of all estimators converges as J increases, the Bayesian estimator consistently achieves the lowest RMSE. This underscores its robustness under conditions of limited group sizes and low ICC values, aligning with theoretical expectations: the regularized Bayesian approach improves RMSE more effectively than standard ML estimators by balancing bias and variance, especially in small sample scenarios.

Figure 3 presents the RB of the three estimation procedures across varying numbers of groups (J) and ICC levels. The subplots reveal that the regularized Bayesian estimator generally exhibits lower relative bias than the ML estimators, particularly when J is small and ICC levels are low. Furthermore, the relative bias of the regularized Bayesian estimator remains more stable compared to the ML estimators as ICC increases, especially in scenarios with small J. However, as J increases, the standard ML estimator approaches zero bias, consistent with its unbiased design. The regularized Bayesian estimator, due to its construction, introduces a small bias to reduce overall variance. Interestingly, despite this intentional bias, it exhibits a smaller relative bias than the ML estimators in small sample conditions, highlighting its advantage in

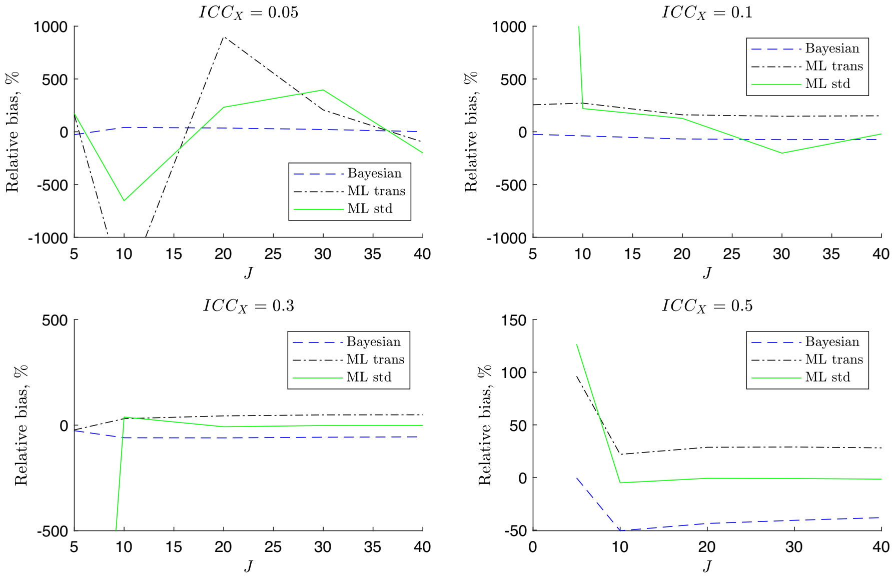

Fig. 3 Relative bias (RB) in % in estimating the between-group slope β_b for the regularized Bayesian, transformed, and standard ML estimators as a function of the group-level sample size (J). ***Note.*** The y-axis scale differs across the four subplots. Results are shown for $n = 15$ persons per group, with a within-group slope $\beta_w = 0.5$, a between-group slope $\beta_b = 0.2$, and varying levels of ICC_X

providing more accurate estimates of the between-group parameter. In larger samples, however, the ML estimator achieves smaller relative bias, as expected from its design.

Figure 4 illustrates the coverage rate (CR) for the three estimation methods. The regularized Bayesian estimator consistently demonstrated a high coverage rate, particularly when ICC values were relatively low (subplots 1 and 2 of Fig. 4). This indicates that the Bayesian estimator provides more conservative uncertainty quantification. Although the Bayesian estimator exhibited slight overcoverage, this may be advantageous in small-sample scenarios where undercoverage of the standard ML estimator can lead to severe issues. As ICC levels increase (subplots 3 and 4 of Fig. 4), the coverage rates for the ML estimators converge toward the nominal 95% level. In contrast, the Bayesian estimator does not converge to exactly 95%, but maintains higher coverage rates in small sample settings, especially at low ICC values. This suggests that the variance of the Bayesian estimator does not strictly adhere to the t-distribution as J increases.

To further investigate the issue of underestimated and overestimated confidence intervals, we compared the lengths of the true (blue solid line) and estimated (orange solid line) confidence intervals for both the regularized Bayesian and standard ML estimators, as shown in Fig. 5. The estimated confidence intervals were computed using the t test, while the theoretical confidence intervals were derived from the 2.5% and 97.5% quantiles of the sample distribution. Figure 5 illustrates the similarity in the lengths of the estimated confidence intervals for both estimators when compared to the theoretical ones. The results indicated no significant difference in the lengths of the

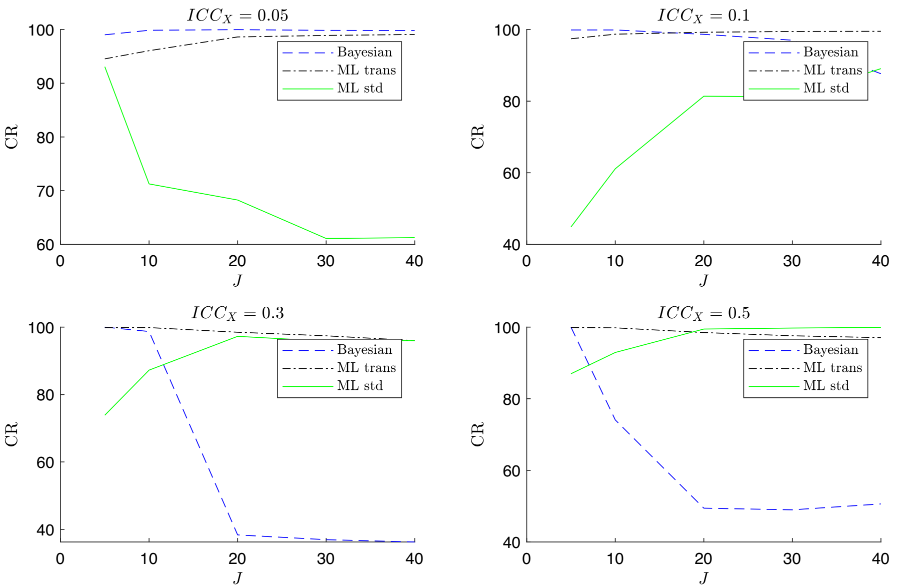

Fig. 4 Coverage rate (CR) in estimating the between-group slope β_b for the regularized Bayesian, transformed, and standard ML estimators as a function of the group level sample size (J). ***Note.*** The y-axis scale differs across the four subplots. Results are shown for $n = 15$ persons per group, with a within-group slope $\beta_w = 0.5$, a between-group slope $\beta_b = 0.2$, and varying levels of ICC_X

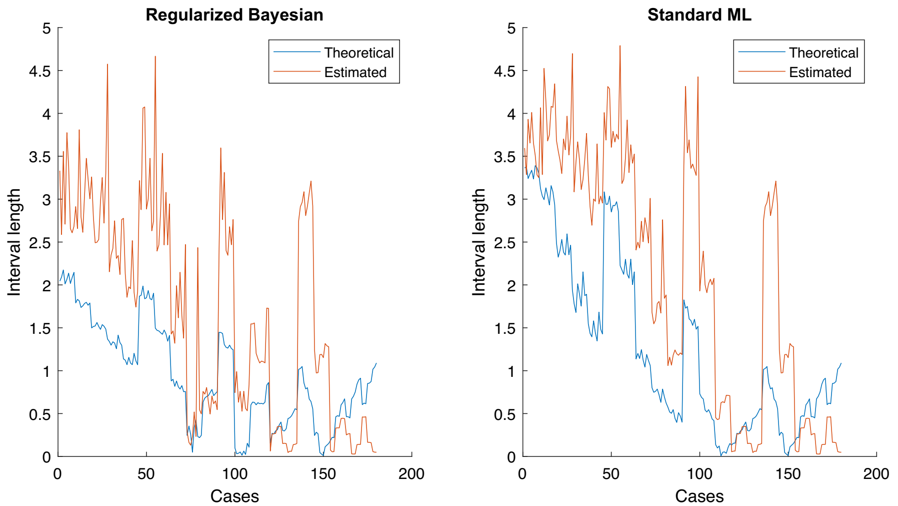

Fig. 5 Length of confidence intervals for the regularized Bayesian and standard ML estimators. ***Note.*** Results are shown for $n = 15$ persons per group, with a within-group slope $\beta_w = 0.5$, a between-group slope $\beta_b = 0.2$, and varying levels of ICC_X

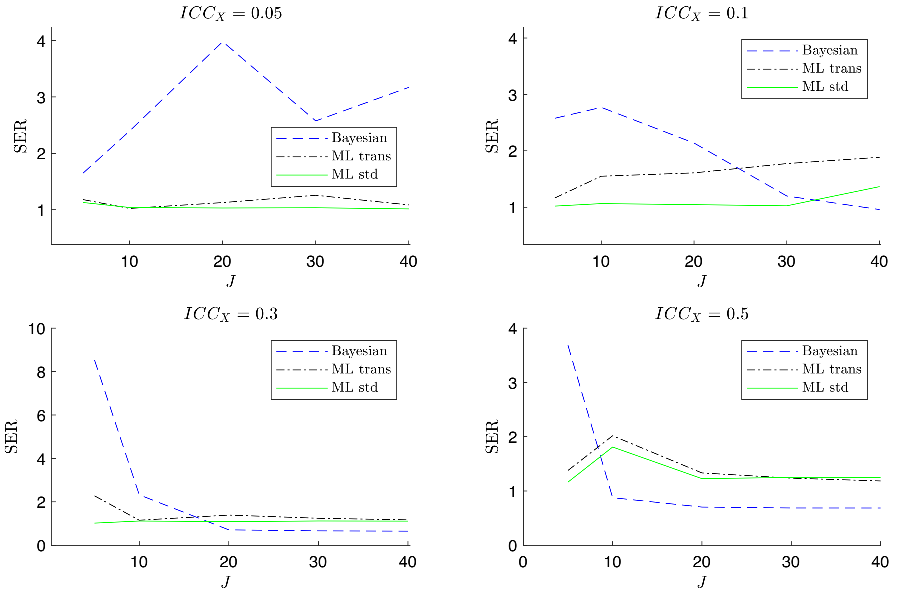

Fig. 6 Standard Error Ratio (SER) in estimating the between-group slope β_b for the regularized Bayesian, transformed, and standard ML estimators as a function of the group level sample size (J). ***Note.*** The y-axis scale differs across the four subplots. Results are shown for $n = 15$ persons per group, with a within-group slope $\beta_w = 0.5$, a between-group slope $\beta_b = 0.2$, and varying levels of ICC$_X$

theoretical and estimated confidence intervals between the two methods. This finding suggests that the regularized Bayesian estimator did not produce wider confidence intervals due to thicker tails. Instead, the bias inherent in the regularized Bayesian estimator resulted in skewed distributions. Therefore, to construct accurate confidence intervals for the regularized Bayesian estimator when J is large, a skewed t-distribution should be employed.

Combining the insights from Figs. 4 and 5, the results highlight the strength of the Bayesian estimator in producing more accurate parameter estimates and confidence intervals, especially in scenarios with limited data and low ICC values.

Following the previous discussion, Fig. 6 illustrates the standard error ratio (SER) for the three methods. The subplots indicate that the regularized Bayesian estimator tended to exhibit an SER greater than 1 when the ICC and sample size were small, suggesting that this estimator produced relatively large standard errors. In contrast, the ML estimators, particularly under small sample sizes and low ICC conditions, showed SER values closer to 1, aligning more closely with the empirical standard errors.

The SER for the ML estimator of the transformed model was slightly higher than that of the standard ML estimator in most scenarios. This increase can be attributed to the transformation procedure, where substituting unknown parameters with estimated ones introduces additional variability—similar to what is observed, for example, in instrumental variable models. As the number of groups (J) and ICC values increased, the SERs for all methods declined and converged toward 1, reflecting improved accuracy in standard error estimation.

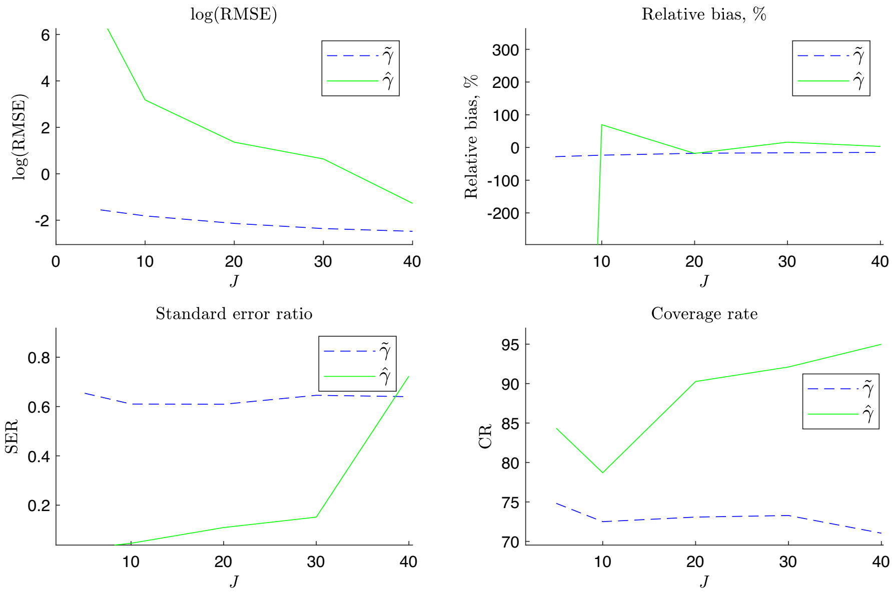

Fig. 7 Log(RMSE), RB, CR, and SER in estimating the covariate slope γ for the standard ML ($\hat{\gamma}$) and the transformed ML ($\tilde{\gamma}$) estimators as a function of the group level sample size (J). ***Note.*** The y-axis scale differs across the four subplots. Results are shown for $n = 15$ persons per group, with a within-group slope $\beta_w = 0.5$, a between-group slope $\beta_b = 0.2$, and varying levels of ICC_X

To address the overestimation of standard errors by the regularized Bayesian approach in small sample scenarios, resampling techniques such as the jackknife procedure proposed by Shao and Wu (1989) could be explored as a potential solution.

Figure 7 compares the performance of two ML estimators—standard ML (green solid line) and transformed ML (blue dashed line)—in estimating the covariate parameter γ. The transformation procedure used in the transformed ML is identical to that implemented in the regularized Bayesian estimation, and the estimation of γ is an integral part of this transformation. Consequently, the results for γ estimation are identical for the transformed ML and the regularized Bayesian estimator; thus, we report them only for the transformed ML. The complete results are provided in Tables S17–S32 of Supplement A.

Each subplot in Fig. 7 presents a different performance measure (logged RMSE, RB, CR, and SER) with ICC fixed to 0.1, $n = 15$, and varying group sizes (J). As in 2, logged RMSE was used for improved visualization. Note, however, that Tables S17–S20 report RMSE values without log transformations for clearer interpretation.

The first subplot demonstrates that RMSE for both estimators decreased as J increased, reflecting improved accuracy with larger group sizes. Notably, the transformed ML estimator consistently achieved a lower RMSE compared to the standard ML estimator, with differences reaching up to three orders of magnitude for small J. This indicates that the transformed ML estimator provided more accurate estimates of γ, particularly in scenarios with limited group sizes.

In the second subplot shows that the RB of the standard ML estimator fluctuated more at smaller J values and exhibited greater bias compared to the transformed ML estimator. As J increased, the bias of both estimators decreased; however, the transformed ML estimator approached near-zero bias more quickly, demonstrating greater robustness across all J values.

The third subplot indicates that the SER of the standard ML estimator was lower than that of the transformed ML estimator, particularly for smaller J, suggesting reduced variability in standard error estimates. The transformed ML estimator exhibited higher SERs, particularly at lower J, due to the added variability introduced by the transformation. Nonetheless, both ML estimators maintained SER values below 1 across all J values, indicating consistent reliability in standard error estimation.

The fourth subplot shows that the CR of the standard ML estimator is closer to the nominal 95% benchmark, particularly for larger J, consistently providing more reliable confidence intervals. This performance advantage can be attributed to the larger standard errors associated with the transformed ML estimator.

In summary, both ML methods demonstrated strong estimation quality for larger sample sizes. However, with limited data, a trade-off emerges between the higher accuracy of the transformed ML and the smaller standard errors of the standard ML estimator. Considering the substantial reduction in RMSE, we recommend using the transformed ML method for estimating the covariate parameter γ.

In conclusion, the simulations presented in this section highlight the relative advantages of the extended regularized Bayesian estimator over ML estimators, particularly in scenarios with small sample size (J) and low ICC levels. The regularized Bayesian estimator demonstrated better consistency, achieving lower RMSE and relative bias compared to ML estimators, resulting in more accurate estimates of the between-group parameter.

However, the simulations also revealed challenges for the regularized Bayesian estimator, including lower coverage rates and worse standard error ratios compared to both ML estimators in scenarios with small J and low ICC levels. These issues are primarily due to the inherent bias introduced by regularization.

To address these limitations, we recommend incorporating resampling techniques, such as the jackknife procedure proposed by Shao and Wu (1989), to adjust standard errors. Additionally, employing a skewed t-distribution for CR adjustments could improve the accuracy of confidence intervals and better align observed and expected variance in conditions with limited group sizes and low ICC levels.

For estimating the covariate parameter γ, the ML estimator of the transformed model proved more accurate than the standard ML estimator, albeit with slightly higher variance. This accuracy advantage justifies its role as an intermediate step in regularized Bayesian estimation procedure for γ.

In summary, our optimally regularized Bayesian estimator demonstrates its capacity to enhance accuracy for the between-group slope, particularly in small-sample scenarios. While recognizing the inherent bias and reliance on intermediate ML estimation, the regularized Bayesian estimator consistently outperforms the ML estimator in terms of accuracy under limited data conditions. By incorporating resampling techniques for standard error estimation and employing a skewed t-test, this approach is expected to yield more accurate estimates in challenging real-world applications.

4 Step-by-step tutorial using MLOB R package

To illustrate the practical application of the extended regularized Bayesian estimator, we use the MultiLevelOptimalBayes (MLOB) package, which includes the estimation function `mlob()`. In this section, we provide step-by-step instructions on applying this function. This tutorial demonstrates the performance of the extended estimator, comparing it with the estimator applied to a model without covariates.

The estimators are applied to the PASSNYC dataset—a real-world dataset on educational equity in New York City that includes data from 1272 schools across 32 districts. This dataset allows us to examine how economic need (ENI) affects math proficiency while accounting for the multilevel structure of the data. Additional covariates include English Language Arts (ELA) proficiency and its interaction with ENI.

4.1 Loading MLOB package

First, we install and load the MLOB package[1]:

```
install.packages("devtools")
library(devtools)
install_github("MLOB-dev/MLOB")
library(MultiLevelOptimalBayes)
```

4.2 Loading and preparing the dataset

To run the model, we use the PASSNYC dataset available on Kaggle.[2] In this step, we load, clean, and convert the necessary variables of the PASSNYC dataset to numeric values:

```
# Load data (set up the correct folder in R using setwd())
data <- read.table("2016 School Explorer.csv", sep = ',', header = TRUE)

# Create a subset excluding N/A values in Average.Math.Proficiency
data_subset <- data[data$Average.Math.Proficiency != 'N/A', ]

# Convert variable Average.Math.Proficiency to numeric
data_subset$math <- as.numeric(data_subset$Average.Math.Proficiency)

# Convert variable Economic.Need.Index to numeric variable ENI
data_subset$ENI = as.numeric(data_subset$Economic.Need.Index)

# Convert variable Average.ELA.Proficiency to numeric
data_subset$ELA = as.numeric(data_subset$Average.ELA.Proficiency)
```

4.3 Estimating the between-group effect

We estimate the contextual effect of economic need on average math proficiency by running two models. The first model incorporates additional covariates using the

[1] Also available on CRAN: https://cran.r-project.org/web/packages/MultiLevelOptimalBayes/.

[2] https://www.kaggle.com/datasets/passnyc/data-science-for-good/data.

extended regularized Bayesian estimator, whereas the second includes only the primary regressor and applies the non-extended estimator from Dashuk et al. (2024). We specify `District` as the grouping variable to account for the hierarchical structure of the data. This setup allows for a direct comparison between the two estimators.

To ensure reproducibility, we set a random seed before processing the dataset. Since the dataset is unbalanced (i.e., the number of individuals per group varies), our procedure randomly removes units from larger groups to equalize group sizes. Setting a seed ensures that the same entities are removed in each run, making the results fully replicable.

```
# Set seed for reproducibility
set.seed(123)

# Apply the mlob function without covariates
result_no_C <- mlob(math ˜ ENI, data = data_subset, group = 'District',
balancing.limit = 0.35)

# Apply the mlob function with covariates
result <- mlob(math ˜ ENI+ ELA + ENI:ELA,

 data = data_subset, group = 'District',
balancing.limit = 0.35)
```

To see the relationship between the R model and the theoretical framework introduced in Eqs. 3–4, we define: math as the dependent variable (Y), ENI as the regressor (X) whose between-group effect we aim to estimate, and ENI and ENI:ELA as the additional covariates (C). Note that the regressor must always be the first term on the right-hand side of the regression formula in `mlob()` to ensure correct model specification by the user.

If the data is unbalanced, a warning message will notify the user that a balancing procedure has been applied. Furthermore, if the imbalance is strong, the function will issue a warning that the resulting estimates may not be trustworthy.

4.4 Summary of results

The output of the customized `summary()` function provides the estimated between-group effect (β_b) and, if applicable, the estimated effects of the covariates (γ). For comparison purposes, the `summary()` function also includes ML estimation results.

Results of the model with covariates using the extended regularized Bayesian estimator read:

```
summary(result)
```

```
Call:
mlob(math ˜ ENI + ELA + ENI:ELA, data = data, group = 31, balancing.limit = 0.35)

Summary of Coefficients:
Estimate  Std. Error  Lower CI (95%)  Upper CI (95%)   Z value   Pr(>|z|)  Significance
beta_b        -0.3171          0.0051        -0.3272       -0.3072        -62.2151     0.00e+00        ***
gamma_ELA      1.2223          0.2581         0.7163        1.7282          4.7348     2.19e-06        ***
gamma_ENI:ELA 0.2299            0.5850        -0.9167        1.3765           0.3930     6.94e-01

For comparison, summary of coefficients from unoptimized analysis (ML):
```

```
Estimate  Std. Error  Lower CI (95%)  Upper CI (95%)   Z value   Pr(>|z|)  Significance
beta_b       -0.5344         0.2380             -1.0009      -0.0679       -2.2454    2.47e-02        *
gamma_ELA     1.2223         0.2581              0.7163        1.7282        4.7348    2.19e-06       ***
gamma_ENI:ELA 0.2299          0.5850             -0.9167        1.3765        0.3930    6.94e-01

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Results of the model without covariates using the regularized Bayesian estimator:

```
summary(result_no_C)

Call:
mlob(math ~ ENI, data = data_subset, group = "District", balancing.limit = 0.35)

Summary of Coefficients:
Estimate  Std. Error  Lower CI (95%)  Upper CI (95%)   Z value   Pr(>|z|)  Significance
beta_b   -1.0379       0.0183             -1.0737       -1.0020      -56.6769   0.00e+00       ***

For comparison, summary of coefficients from unoptimized analysis (ML):
Estimate  Std. Error  Lower CI (95%)  Upper CI (95%)   Z value   Pr(>|z|)  Significance
beta_b   -1.7415      0.7580       -3.2271        -0.2560      -2.2977   0.0216          *

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

4.5 Interpretation

The estimation results demonstrate a significant difference between the regularized Bayesian estimators. Both estimators show a highly significant negative between-group effect of economic need on math proficiency across districts, indicating that districts with higher average economic need tend to have lower average math proficiency. The effect estimated by the extended Bayesian estimator is nearly 70% smaller than that of the non-extended estimator (-0.3171 vs. -1.0379), highlighting the impact of incorporating additional covariates. The extended estimator also achieves substantially lower standard errors compared to the non-extended estimator (0.0051 vs. 0.0183). The between-group effect estimate ($\tilde{\beta}_b$) from the Bayesian extended estimator is more accurate, exhibiting reduced variability and better capturing the relationship between economic need (ENI) and math proficiency while controlling for ELA proficiency and its interaction with ENI.

Comparing the two Bayesian models, we observe that excluding covariates results less precise estimates, while their inclusion enhances accuracy. Notably, this pattern is also evident in the ML estimation: the ML-estimated between-group effect in the model with covariates is substantially smaller than in the model without covariates (-0.5344 vs. -1.7415), with correspondingly lower standard errors (0.2380 vs. 0.7580). However, both ML estimators seem to overestimate the magnitude of effects, underscoring their limitations in small-sample hierarchical data settings. Moreover, ML estimates are "less statistically significant" than Bayesian estimates, as indicated by the single-star significance level in contrast to the three-star significance of the Bayesian estimators.

5 Discussion

In this article, we extended the regularized Bayesian estimator to a multivariate setting by incorporating covariates to improve the estimation of the between-group parameter

β_b. Building on previous work (Dashuk et al. 2024; Zitzmann et al. 2021b), this approach leverages prior information to optimize the MSE of the Bayesian estimator within a two-level latent model with covariates. While rooted in Bayesian estimation, our approach draws parallels with frequentist techniques, such as regularized SEM and the Stein estimator, to enhance estimation accuracy. By introducing covariates, this extension adds complexity but allows for a more nuanced analysis of relationships between variables in multilevel structures, particularly in settings with small sample sizes and low ICC levels.

The proposed model differs from standard multilevel Bayesian models, because it incorporates latent variables structure that is not handled by conventional Bayesian hierarchical frameworks, such as those implemented in the `brms` package in R. In particular, standard Bayesian hierarchical models assume that all regressors are directly observed, whereas our approach allows for the decomposition of covariates into within-group and between-group components, introducing latent variables. Moreover, contrary to our proposed regularized estimator, a Bayesian model with non-informative priors is expected to perform similarly to the ML estimation. Note that our estimator is designed to automatically select priors that optimize performance in terms of MSE on finite samples by incorporating shrinkage. In most cases, this prior improves estimation efficiency and stability compared to conventional Bayesian hierarchical models. However, we acknowledge that in certain cases, alternative priors may provide better results, which we leave as a direction for future research.

Our simulations consistently demonstrated that the extended regularized Bayesian estimator outperforms both the standard and transformed ML estimators in estimating β_b, particularly in small-sample and low-ICC scenarios. The Bayesian estimator achieved lower RMSE and relative bias values compared to the ML estimators, emphasizing its accuracy under challenging conditions. While the regularized Bayesian method excelled in minimizing MSE, it exhibited lower coverage rates and higher standard error ratios. To address these limitations, we proposed incorporating resampling techniques such as the jackknife (Shao and Wu 1989), which has turned out to be advantageous in multilevel modeling (e.g., Zitzmann 2018), and employing a skewed t-test in order to improve both CR and SER.

Although in our simulations, we fixed ICC_C, ICC_Y, and γ, we acknowledge that different values could influence the results. However, the real-world application demonstrates that the proposed estimator may effectively generalize to hierarchical data beyond the specific simulation conditions. Future research could explore additional parameter variations.

To further encourage practitioners to adopt the extended regularized Bayesian estimation and the `MLOB` package, we provide Tables 1, 2, 3 and 4 in Appendix C, which summarize average evaluation criteria (RMSE, BR, CR, SER) for different combinations of group size n and number of groups J in order to equip users with expectations about performance when key parameters (e.g., ICC_X ad β_w) are unknown.

The transformed ML estimator proved valuable as an intermediate step for refining the Bayesian estimator, particularly in estimating the covariate parameter γ. By adjusting the dependent variable to account for covariate effects, the transformed ML estimator enhances the accuracy of γ parameter estimation in multilevel models compared to the standard ML estimator. However, the transformed ML estimator still exhibits significantly higher MSE when estimating the between-group parameter β_b compared to the extended regularized Bayesian estimator, particularly in scenarios with low ICCs and small sample sizes.

Despite its advantages, the benefits of the regularized Bayesian estimator diminish as sample sizes increase, with the performance of the ML estimators converging toward that of the Bayesian approach. This limitation highlights that the Bayesian estimator is most effective in small-sample scenarios.

Future research could explore the application of the Bayesian estimator to unbalanced data, where the number of individuals (n) varies across groups. This extension would enable the evaluation of more unstructured data, improving the estimator's alignment with real-world conditions. Currently, most existing solutions for balancing data are designed for large datasets and often introduce additional noise, which can compromise the accuracy of small-sample models. For small-samples, a balancing procedure should ideally avoid adding artificial entries or reducing the number of groups to preserve the structure and validity of the model.

Another avenue for future research is modeling measurement error for the predictors and specifying error distributions. Studies could investigate the trade-offs between increased model complexity and estimation accuracy, providing insights into the benefits and limitations of incorporating measurement error into multilevel models.

Other work could focus on refining the Bayesian estimator to handle models with more than two levels of grouping. Additionally, expanding the estimator to accommodate other model types and data structures, such as longitudinal data, could further enhance its applicability and practical utility.

In conclusion, this study advances the regularized Bayesian estimator, demonstrating its superiority over traditional ML methods, particularly in scenarios with small sample sizes and low ICC levels. By optimizing the balance between bias reduction and variance minimization, this estimator provides a powerful tool for improving parameter estimation in multilevel models. The possibility to include covariates enhances its applicability and robustness, while the transformed ML approach improves the accuracy of the covariate parameter γ. We believe that our approach holds substantial promise for researchers working with small samples, offering more reliable and informed modeling practices across diverse fields.

Appendix A: Derivations of the distribution of $\tilde{Y}$

In this section, we present the distribution of $\tilde{Y}_{ij}$ from Eq. 12, taking into account the covariance between Y and C.

As discussed in the main text, Eqs. 3 and 4 define the original two-level model, with the distribution parameters for the predictor, covariates, and dependent variable provided in Eqs. 1, 2, 15, and 16.

However, to consider the model in Eqs. 13 and 14, we need to define the distributions of $\tilde{Y}_{b,j}$ and $\tilde{Y}_{w,ij}$. To achieve this, we examined the between-group and within-group components of $\tilde{Y}_{ij}$ separately, treating them as complementary parts of $\tilde{Y}_{ij}$:

$$\tilde{Y}_{ij} = Y_{ij} - C_{ij}\hat{\gamma} = Y_{b,j} + Y_{w,ij} - (C_{b,j} + C_{w,ij})\hat{\gamma} \tag{25}$$

$$\tilde{Y}_{ij} = \underbrace{Y_{b,j} - C_{b,j}\hat{\gamma}}_{\tilde{Y}_{b,j}} + \underbrace{Y_{w,ij} - C_{w,ij}\hat{\gamma}}_{\tilde{Y}_{w,ij}} = \tilde{Y}_{b,j} + \tilde{Y}_{w,ij} \tag{26}$$

The between-group component, $\tilde{Y}_{b,j}$, follows a normal distribution as it is a linear combination of normally distributed random variables. Therefore, its distribution is fully characterized by its mean and variance. The mean is computed as:

$$E(\tilde{Y}_{b,j}) = E(Y_{b,j} - C_{b,j}\hat{\gamma}) = E(Y_{b,j}) - E(C_{b,j})\hat{\gamma} = \mu_Y - \mu_C\hat{\gamma} \tag{27}$$

The variance is computed as:

$$\mathrm{Var}(\tilde{Y}_{b,j}) = \mathrm{Var}(Y_{b,j} - C_{b,j}\hat{\gamma}) \tag{28}$$

$$\mathrm{Var}(\tilde{Y}_{b,j}) = \underbrace{\mathrm{Var}(Y_{b,j})}_{\tau_Y^2} - 2\underbrace{\mathrm{Cov}(Y_{b,j}, C_{b,j})}_{\tau_{YC}}\hat{\gamma} + \hat{\gamma}'\underbrace{\mathrm{Var}(C_{b,j})}_{\tau_C^2}\hat{\gamma} = \tag{29}$$

$$\mathrm{Var}(\tilde{Y}_{b,j}) = \tau_Y^2 - 2\tau_{YC}\hat{\gamma} + \hat{\gamma}'\tau_C^2\hat{\gamma} \tag{30}$$

Thus, the distribution of $\tilde{Y}_{b,j}$ is:

$$\tilde{Y}_{b,j} \sim N(\mu_Y - \mu_C\hat{\gamma}, \tau_Y^2 - 2\tau_{YC}\hat{\gamma} + \hat{\gamma}'\tau_C^2\hat{\gamma}) \tag{31}$$

The within-group component $\tilde{Y}_{w,ij}$ also follows a normal distribution, characterized by its mean and variance. The mean is calculated as:

$$E(\tilde{Y}_{w,ij}) = E(Y_{w,ij} - C_{w,ij}\hat{\gamma}) = E(Y_{w,ij}) - E(C_{w,ij})\hat{\gamma} = 0 - 0 \cdot \hat{\gamma} = 0 \tag{32}$$

The variance is calculated as:

$$\mathrm{Var}(\tilde{Y}_{w,ij}) = \mathrm{Var}(Y_{w,ij} - C_{w,ij}\hat{\gamma}) \tag{33}$$

$$\mathrm{Var}(\tilde{Y}_{w,ij}) = \underbrace{\mathrm{Var}(Y_{w,ij})}_{\sigma_Y^2} - 2\underbrace{\mathrm{Cov}(Y_{w,ij}, C_{w,ij})}_{\sigma_{YC}}\hat{\gamma} + \hat{\gamma}'\underbrace{\mathrm{Var}(C_{w,ij})}_{\sigma_C^2}\hat{\gamma} \tag{34}$$

$$\mathrm{Var}(\tilde{Y}_{w,ij}) = \sigma_Y^2 - 2\sigma_{YC}\hat{\gamma} + \hat{\gamma}'\sigma_C^2\hat{\gamma} \tag{35}$$

Thus, the distribution of $\tilde{Y}_{w,ij}$ is:

$$\tilde{Y}_{w,ij} \sim N(0, \sigma_Y^2 - 2\sigma_{YC}\hat{\gamma} + \hat{\gamma}'\sigma_C^2\hat{\gamma}) \tag{36}$$

The between-group and within-group components together define the overall distribution of $\tilde{Y}_{ij}$:

$$\tilde{Y}_{ij} \sim N(\mu_Y - \mu_C \hat{\gamma}, (\tau_Y^2 + \sigma_Y^2) - 2(\tau_{YC} + \sigma_{YC})\hat{\gamma} + \hat{\gamma}'(\tau_C^2 + \sigma_C^2)\hat{\gamma}) \tag{37}$$

To maintain model consistency, we further demonstrated the independence of $\tilde{Y}_b$ and $\tilde{Y}_w$, leveraging the fact that all between-group components are independent of all within-group components of the origin model (Eqs. 3 and 4). Since $\tilde{Y}_b$ and $\tilde{Y}_w$ are both normally distributed, it is sufficient to show that their covariance is zero:

$$\begin{aligned}\mathrm{Cov}(\tilde{Y}_b, \tilde{Y}_w) &= \mathrm{Cov}(Y_b - C_b\hat{\gamma}, Y_w - C_w\hat{\gamma}) \\ &= \underbrace{\mathrm{Cov}(Y_b, Y_w)}_{=0} - \underbrace{\mathrm{Cov}(Y_b, C_w)}_{=0}\hat{\gamma} \\ &\quad - \hat{\gamma}'\underbrace{\mathrm{Cov}(C_b, Y_w)}_{=0} + \hat{\gamma}'\underbrace{\mathrm{Cov}(C_b, C_w)}_{=0}\hat{\gamma} = 0\end{aligned} \tag{38}$$

Thus, we conclude that the between-group and within-group components of Y are independent.

In the same manner, the independence assumption must hold for the between-group and within-group components of $\tilde{Y}$ and X. Starting with the normal distribution of all components, $\tilde{Y}_b$, $\tilde{Y}_w$, X_b, and X_w, we computed the covariances for the two pairs: ($\tilde{Y}_b$ and X_w) and ($\tilde{Y}_w$ and X_b). The results are as follows:

$$\mathrm{Cov}(\tilde{Y}_b, X_w) = \mathrm{Cov}(Y_b - C_b\hat{\gamma}, X_w) = \underbrace{\mathrm{Cov}(Y_b, X_w)}_{=0} - \hat{\gamma}'\underbrace{\mathrm{Cov}(C_b, X_w)}_{=0} = 0 \tag{39}$$

$$\mathrm{Cov}(\tilde{Y}_w, X_b) = \mathrm{Cov}(Y_w - C_w\hat{\gamma}, X_b) = \underbrace{\mathrm{Cov}(Y_w, X_b)}_{=0} - \hat{\gamma}'\underbrace{\mathrm{Cov}(C_w, X_b)}_{=0} = 0 \tag{40}$$

Consequently, we derived the distribution of $\tilde{Y}$ and proved the independence of all between-group and within-group components in the adjusted two-level model from Eqs. 13 and 14. This key result enables the implementation of both the ML estimator and the regularized Bayesian estimator for the between-group slope β_b as described by Dashuk et al. (2024).

Appendix B: Regularized Bayesian estimation procedure

In this appendix, we introduce a revised step-by-step procedure based on the theoretical framework developed by Dashuk et al. (2024). This procedure offers an efficient and effective solution for the regularized Bayesian estimator of the between-group slope $\tilde{\beta}_b$ (Eq. 23) in a two-level model with covariates (Eqs. 3 and 4):

- Input data: n, J, X_{ij} and $\tilde{Y}_{ij}$ (Eq. 12)

- Define matrix A as:

$$A = \begin{array}{c} \begin{array}{ccccccc} 1 & \cdots & nJ & nJ+1 & \cdots & nJ+J & nJ+J+1 \end{array} \\ \begin{pmatrix} -\frac{1}{n(n-1)J} & \cdots & 0 & 0 & \cdots & 0 & 0 \\ 0 & \ddots & 0 & 0 & \cdots & 0 & 0 \\ 0 & \cdots & -\frac{1}{n(n-1)J} & 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & \frac{nJ-1}{(n-1)(J-1)J} & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 0 & \ddots & 0 & 0 \\ 0 & \cdots & 0 & 0 & \cdots & \frac{nJ-1}{(n-1)(J-1)J} & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 & -\frac{J}{J-1} \end{pmatrix} \end{array} \tag{41}$$

- Calculate the group manifest means: $\overline{X}_{\bullet j}$ of X, and $\overline{\tilde{Y}}_{\bullet j}$ of $\tilde{Y}$:

$$\overline{X}_{\bullet j} = \frac{1}{n}\sum_{i=1}^{n}(X_{b,j} + X_{w,ij}) = X_{b,j} + \frac{1}{n}\sum_{i=1}^{n} X_{w,ij} \tag{42}$$

$$\overline{\tilde{Y}}_{\bullet j} = \frac{1}{n}\sum_{i=1}^{n}(\tilde{Y}_{b,j} + \tilde{Y}_{w,ij}) = \tilde{Y}_{b,j} + \frac{1}{n}\sum_{i=1}^{n} \tilde{Y}_{w,ij} \tag{43}$$

- Calculate overall means: $\overline{X}_{\bullet\bullet}$ of X and $\overline{\tilde{Y}}_{\bullet\bullet}$ of $\tilde{Y}$:

$$\overline{X}_{\bullet\bullet} = \frac{1}{nJ}\sum_{j=1}^{J}\sum_{i=1}^{n}(X_{b,j} + X_{w,ij}) = \frac{1}{J}\sum_{j=1}^{J} X_{b,j} + \frac{1}{nJ}\sum_{j=1}^{J}\sum_{i=1}^{n} X_{w,ij} \tag{44}$$

$$\overline{\tilde{Y}}_{\bullet\bullet} = \frac{1}{nJ}\sum_{j=1}^{J}\sum_{i=1}^{n}(\tilde{Y}_{b,j} + \tilde{Y}_{w,ij}) = \frac{1}{J}\sum_{j=1}^{J} \tilde{Y}_{b,j} + \frac{1}{nJ}\sum_{j=1}^{J}\sum_{i=1}^{n} \tilde{Y}_{w,ij} \tag{45}$$

- Compute between-group and within-group covariances of X and $\tilde{Y}$: $\hat{\tau}_X^2$, $\hat{\tau}_{\tilde{Y}X}$, $\hat{\tau}_{\tilde{Y}}^2$, $\hat{\sigma}_X^2$, $\hat{\sigma}_{\tilde{Y}X}$, and $\hat{\sigma}_{\tilde{Y}}^2$:

$$\hat{\tau}_X^2 = -\frac{1}{n(n-1)J}\sum_{j=1}^{J}\sum_{i=1}^{n} X_{ij}^2 + \frac{nJ-1}{(n-1)(J-1)J}\sum_{j=1}^{J}\overline{X}_{\bullet j}^2 - \frac{J}{J-1}\overline{X}_{\bullet\bullet}^2 \tag{46}$$

$$\hat{\tau}_{\tilde{Y}X} = -\frac{1}{n(n-1)J}\sum_{j=1}^{J}\sum_{i=1}^{n} \tilde{Y}_{ij}X_{ij} + \frac{nJ-1}{(n-1)(J-1)J}\sum_{j=1}^{J}\overline{\tilde{Y}}_{\bullet j}\overline{X}_{\bullet j} - \frac{J}{J-1}\overline{\tilde{Y}}_{\bullet\bullet}\overline{X}_{\bullet\bullet} \tag{47}$$

$$\hat{\tau}_{\tilde{Y}}^2 = -\frac{1}{n(n-1)J}\sum_{j=1}^{J}\sum_{i=1}^{n}\tilde{Y}_{ij}^2 + \frac{nJ-1}{(n-1)(J-1)J}\sum_{j=1}^{J}\overline{\tilde{Y}}_{\bullet j}^2 - \frac{J}{J-1}\overline{\tilde{Y}}_{\bullet\bullet}^2 \tag{48}$$

$$\hat{\sigma}_X^2 = \frac{1}{(n-1)J}\sum_{j=1}^{J}\sum_{i=1}^{n}X_{ij}^2 - \frac{n}{(n-1)J}\sum_{j=1}^{J}\overline{X}_{\bullet j}^2 \tag{49}$$

$$\hat{\sigma}_{\tilde{Y}X} = \frac{1}{(n-1)J}\sum_{j=1}^{J}\sum_{i=1}^{n}X_{ij}\tilde{Y}_{ij} - \frac{n}{(n-1)J}\sum_{j=1}^{J}\overline{X}_{\bullet j}\overline{\tilde{Y}}_{\bullet j} \tag{50}$$

$$\hat{\sigma}_{\tilde{Y}}^2 = \frac{1}{(n-1)J}\sum_{j=1}^{J}\sum_{i=1}^{n}\tilde{Y}_{ij}^2 - \frac{n}{(n-1)J}\sum_{j=1}^{J}\overline{\tilde{Y}}_{\bullet j}^2 \tag{51}$$

- Find ML estimator $\hat{\beta}_b$:

$$\hat{\beta}_b = \frac{\hat{\tau}_{\tilde{Y}X}}{\hat{\tau}_X^2} \tag{52}$$

- Compute diagonal matrices of eigenvalues D_X, $D_{\tilde{Y}}$, $D_{\tilde{Y}X}$, and the matrix of eigenvectors V. Due to its complexity, the matrix V is provided for the case where $n = 3$ and $J = 4$, but it can be expanded as needed for larger dimensions. D_X:
  - $\lambda_i = 0$, $(J+1)$ eigenvalues
  - $\lambda_i = \sigma_X^2$, $((n-1)J)$ eigenvalues
  - $\lambda_i = (n+1)\left(\tau_X^2 + \frac{1}{n}\sigma_X^2\right)$, $(J-1)$ eigenvalues
  - $\lambda_{nJ+J+1} = \frac{nJ+J+1}{J}\left(\tau_X^2 + \frac{1}{n}\sigma_X^2\right)$

  $D_{\tilde{Y}}$:
  - $\lambda_i = 0$, $(J+1)$ eigenvalues
  - $\lambda_i = \sigma_{\tilde{Y}}^2$, $((n-1)J)$ eigenvalues
  - $\lambda_i = (n+1)\left(\tau_{\tilde{Y}}^2 + \frac{1}{n}\sigma_{\tilde{Y}}^2\right)$, $(J-1)$ eigenvalues
  - $\lambda_{nJ+J+1} = \frac{nJ+J+1}{J}\left(\tau_{\tilde{Y}}^2 + \frac{1}{n}\sigma_{\tilde{Y}}^2\right)$

  D_{YX}:
  - $\lambda_i = 0$, $(J+1)$ eigenvalues
  - $\lambda_i = \sigma_{\tilde{Y}X}$, $((n-1)J)$ eigenvalues
  - $\lambda_i = (n+1)\left(\tau_{\tilde{Y}X} + \frac{1}{n}\sigma_{\tilde{Y}X}\right)$, $(J-1)$ eigenvalues
  - $\lambda_{nJ+J+1} = \frac{nJ+J+1}{J}\left(\tau_{\tilde{Y}X} + \frac{1}{n}\sigma_{\tilde{Y}X}\right)$

$$V = \left(\begin{array}{ccccccccc}
-\frac{1}{\sqrt{n\,(n+1)}} & 0 & 0 & 0 & -\frac{1}{\sqrt{(J+J\,n)\,(J+J\,n+1)}} & -\frac{\sqrt{2}}{2} & -\frac{\sqrt{6}}{6} & 0 & \cdots \\
0 & 0 & 0 & 0 & 0 & 0 & -\frac{\sqrt{2}}{2\,\sqrt{n+1}} & -\frac{\sqrt{6}}{6\,\sqrt{n+1}} & \cdots \\
-\frac{1}{\sqrt{n\,(n+1)}} & 0 & 0 & 0 & -\frac{1}{\sqrt{(J+J\,n)\,(J+J\,n+1)}} & \frac{\sqrt{2}}{2} & -\frac{\sqrt{6}}{6} & 0 & \cdots \\
0 & -\frac{1}{\sqrt{n\,(n+1)}} & 0 & 0 & -\frac{1}{\sqrt{(J+J\,n)\,(J+J\,n+1)}} & 0 & 0 & -\frac{\sqrt{2}}{2} & \cdots \\
0 & -\frac{1}{\sqrt{n\,(n+1)}} & 0 & 0 & -\frac{1}{\sqrt{(J+J\,n)\,(J+J\,n+1)}} & 0 & 0 & \frac{\sqrt{2}}{2} & \cdots \\
0 & -\frac{1}{\sqrt{n\,(n+1)}} & 0 & 0 & -\frac{1}{\sqrt{(J+J\,n)\,(J+J\,n+1)}} & 0 & 0 & 0 & \cdots \\
0 & 0 & -\frac{1}{\sqrt{n\,(n+1)}} & 0 & -\frac{1}{\sqrt{(J+J\,n)\,(J+J\,n+1)}} & 0 & 0 & 0 & \cdots \\
0 & 0 & -\frac{1}{\sqrt{n\,(n+1)}} & 0 & -\frac{1}{\sqrt{(J+J\,n)\,(J+J\,n+1)}} & 0 & 0 & 0 & \cdots \\
0 & 0 & -\frac{1}{\sqrt{n\,(n+1)}} & 0 & -\frac{1}{\sqrt{(J+J\,n)\,(J+J\,n+1)}} & 0 & 0 & 0 & \cdots \\
0 & 0 & 0 & -\frac{1}{\sqrt{n\,(n+1)}} & -\frac{1}{\sqrt{(J+J\,n)\,(J+J\,n+1)}} & 0 & 0 & 0 & \cdots \\
0 & 0 & 0 & -\frac{1}{\sqrt{n\,(n+1)}} & -\frac{1}{\sqrt{(J+J\,n)\,(J+J\,n+1)}} & 0 & 0 & 0 & \cdots \\
0 & 0 & 0 & -\frac{1}{\sqrt{n\,(n+1)}} & -\frac{1}{\sqrt{(J+J\,n)\,(J+J\,n+1)}} & 0 & 0 & 0 & \cdots \\
\sqrt{\frac{n}{n+1}} & 0 & 0 & 0 & -\frac{1}{\sqrt{(J+J\,n)\,(J+J\,n+1)}} & 0 & 0 & 0 & \cdots \\
0 & \sqrt{\frac{n}{n+1}} & 0 & 0 & -\frac{1}{\sqrt{(J+J\,n)\,(J+J\,n+1)}} & 0 & 0 & 0 & \cdots \\
0 & 0 & \sqrt{\frac{n}{n+1}} & 0 & -\frac{1}{\sqrt{(J+J\,n)\,(J+J\,n+1)}} & 0 & 0 & 0 & \cdots \\
0 & 0 & 0 & \sqrt{\frac{n}{n+1}} & -\frac{1}{\sqrt{(J+J\,n)\,(J+J\,n+1)}} & 0 & 0 & 0 & \cdots \\
0 & 0 & 0 & 0 & \sqrt{\frac{J+J\,n}{J+J\,n+1}} & 0 & 0 & \cdots &
\end{array} \right.$$

$$\left. \begin{array}{ccccccccc}
0 & 0 & 0 & 0 & -\frac{\sqrt{2}}{2\,\sqrt{n+1}} & -\frac{\sqrt{6}}{6\,\sqrt{n+1}} & -\frac{\sqrt{3}}{6\,\sqrt{n+1}} & \frac{1}{\sqrt{J+J\,n+1}} & \\
0 & 0 & 0 & 0 & 0 & -\frac{\sqrt{2}}{2\,\sqrt{n+1}} & -\frac{\sqrt{6}}{6\,\sqrt{n+1}} & -\frac{\sqrt{3}}{6\,\sqrt{n+1}} & \frac{1}{\sqrt{J+J\,n+1}} \\
0 & 0 & 0 & 0 & 0 & -\frac{\sqrt{2}}{2\,\sqrt{n+1}} & -\frac{\sqrt{6}}{6\,\sqrt{n+1}} & -\frac{\sqrt{3}}{6\,\sqrt{n+1}} & \frac{1}{\sqrt{J+J\,n+1}} \\
-\frac{\sqrt{6}}{6} & 0 & 0 & 0 & 0 & \frac{\sqrt{2}}{2\,\sqrt{n+1}} & -\frac{\sqrt{6}}{6\,\sqrt{n+1}} & -\frac{\sqrt{3}}{6\,\sqrt{n+1}} & \frac{1}{\sqrt{J+J\,n+1}} \\
-\frac{\sqrt{6}}{6} & 0 & 0 & 0 & 0 & \frac{\sqrt{2}}{2\,\sqrt{n+1}} & -\frac{\sqrt{6}}{6\,\sqrt{n+1}} & -\frac{\sqrt{3}}{6\,\sqrt{n+1}} & \frac{1}{\sqrt{J+J\,n+1}} \\
\frac{\sqrt{2}\,\sqrt{3}}{3} & 0 & 0 & 0 & 0 & \frac{\sqrt{2}}{2\,\sqrt{n+1}} & -\frac{\sqrt{6}}{6\,\sqrt{n+1}} & -\frac{\sqrt{3}}{6\,\sqrt{n+1}} & \frac{1}{\sqrt{J+J\,n+1}} \\
0 & -\frac{\sqrt{2}}{2} & -\frac{\sqrt{6}}{6} & 0 & 0 & 0 & \frac{\sqrt{6}}{3\,\sqrt{n+1}} & -\frac{\sqrt{3}}{6\,\sqrt{n+1}} & \frac{1}{\sqrt{J+J\,n+1}} \\
0 & \frac{\sqrt{2}}{2} & -\frac{\sqrt{6}}{6} & 0 & 0 & 0 & \frac{\sqrt{6}}{3\,\sqrt{n+1}} & -\frac{\sqrt{3}}{6\,\sqrt{n+1}} & \frac{1}{\sqrt{J+J\,n+1}} \\
0 & 0 & \frac{\sqrt{2}\,\sqrt{3}}{3} & 0 & 0 & 0 & \frac{\sqrt{6}}{3\,\sqrt{n+1}} & -\frac{\sqrt{3}}{6\,\sqrt{n+1}} & \frac{1}{\sqrt{J+J\,n+1}} \\
0 & 0 & 0 & -\frac{\sqrt{2}}{2} & -\frac{\sqrt{6}}{6} & 0 & 0 & \frac{\sqrt{3}}{2\,\sqrt{n+1}} & \frac{1}{\sqrt{J+J\,n+1}} \\
0 & 0 & 0 & \frac{\sqrt{2}}{2} & -\frac{\sqrt{6}}{6} & 0 & 0 & \frac{\sqrt{3}}{2\,\sqrt{n+1}} & \frac{1}{\sqrt{J+J\,n+1}} \\
0 & 0 & 0 & 0 & \frac{\sqrt{2}\,\sqrt{3}}{3} & 0 & 0 & \frac{\sqrt{3}}{2\,\sqrt{n+1}} & \frac{1}{\sqrt{J+J\,n+1}} \\
0 & 0 & 0 & 0 & 0 & -\frac{\sqrt{2}}{2\,\sqrt{n+1}} & -\frac{\sqrt{6}}{6\,\sqrt{n+1}} & -\frac{\sqrt{3}}{6\,\sqrt{n+1}} & \frac{1}{\sqrt{J+J\,n+1}} \\
0 & 0 & 0 & 0 & 0 & \frac{\sqrt{2}}{2\,\sqrt{n+1}} & -\frac{\sqrt{6}}{6\,\sqrt{n+1}} & -\frac{\sqrt{3}}{6\,\sqrt{n+1}} & \frac{1}{\sqrt{J+J\,n+1}} \\
0 & 0 & 0 & 0 & 0 & 0 & \frac{\sqrt{6}}{3\,\sqrt{n+1}} & -\frac{\sqrt{3}}{6\,\sqrt{n+1}} & \frac{1}{\sqrt{J+J\,n+1}} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{\sqrt{3}}{2\,\sqrt{n+1}} & \frac{1}{\sqrt{J+J\,n+1}} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{\sqrt{J+J\,n+1}}
\end{array} \right) \tag{53}$$

- Calculate square root matrices $S_X = \sqrt{D_X}$ and $S_{\tilde{Y}} = \sqrt{D_{\tilde{Y}}}$
- Compute diagonal matrix of coefficients $L_1 = S_X V' A V S_X$
- Calculate matrix Q:

$$Q = \begin{pmatrix} 0 & \frac{1}{2} S_X V' A V S_{\tilde{Y}} \\ \frac{1}{2} S_X V' A V S_{\tilde{Y}} & 0 \end{pmatrix} \tag{54}$$

- Compute the diagonal matrix of eigenvalues D_H and the matrix of eigenvectors V_H. Note that the generalized inverse of matrices S_X and $S_{\tilde{Y}}$ is used, as described by Penrose (1955), since they contain zero eigenvalues and are not invertible in the classical sense. These generalized inverses, denoted as S_X^+ and $S_{\tilde{Y}}^+$, include the inverse of non-zero diagonal elements and zeros otherwise.

$$D_H = \begin{pmatrix} \mathbf{I} + S_X^+ D_{\tilde{Y}X} S_{\tilde{Y}}^+ & 0 \\ 0 & \mathbf{I} - S_X^+ D_{\tilde{Y}X} S_{\tilde{Y}}^+ \end{pmatrix} \tag{55}$$

$$V_H = \frac{\sqrt{2}}{2} \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 1 & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots & \ddots & \dots & \dots \\ 1 & -1 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 1 & -1 \end{pmatrix} \tag{56}$$

- Calculate square root matrix $S_H = \sqrt{D_H}$
- Compute diagonal matrix of coefficients $L_2 = S_H V_H' Q V_H S_H$
- Compute coefficients k_{sum1}, θ_{sum1}, k_{sum2} and θ_{sum2} (note that $\mathbb{1}$ is a vector of ones):

$$k_{sum1} = \frac{\left(\mathbb{1}'_{nJ+J+1} L_1\right)^2}{2L_1' L_1} \tag{57}$$

$$\theta_{sum1} = \frac{L_1' L_1}{\mathbb{1}'_{nJ+J+1} L_1} \tag{58}$$

$$k_{sum2} = \frac{\left(\mathbb{1}'_{2(nJ+J+1)} L_2\right)^2}{2L_2' L_2} \tag{59}$$

$$\theta_{sum2} = \frac{L_2' L_2}{\mathbb{1}'_{2(nJ+J+1)} L_2} \tag{60}$$

- Define vectors W and T_{02} with the values of ω and τ_0^2 that specify grid search region. For example: W goes from 0 to 1 with step 0.01 and T_{02} goes from 0.1 to 10 with step 0.1

- Compute MSE for each value of W and T_{02}. β_b should be substituted with $\hat{\beta}_b$ as well. The formula is delineated as:

$$
\begin{aligned}
MSE_{ML}(i,j) = & \left\{ k_{sum2}\theta_{sum2}^2(k_{sum2}+1)\left((1-W(i))T_{02}(j)+W(i)\mathbb{1}'_{nJ+J+1}L_1\right)\right\} \\
& / \left\{ \left(((1-W(i))T_{02}(j)+W(i)\mathbb{1}'_{nJ+J+1}L_1)^2 - 2W(i)^2(L1'L1)\right) * \right. \\
& \left. \left(((1-W(i))T_{02}(j)+W(i)\mathbb{1}'_{nJ+J+1}L_1)^2 - 4W(i)^2(L1'L1)\right)\right\} \\
& - \frac{2\hat{\beta}_b k_{sum2}\theta_{sum2}\left((1-W(i))T_{02}(j)+W(i)\mathbb{1}'_{nJ+J+1}L_1\right)}{\left(((1-W(i))T_{02}(j)+W(i)\mathbb{1}'_{nJ+J+1}L_1)^2 - W(i)^2\cdot(L1'L1)\right)} + \hat{\beta}_b
\end{aligned}
\tag{61}
$$

- Find the minimum MSE and indexes i^* and j^* that produce this minimum
- Define optimal parameters $\omega^* = W(i^*)$ and $\tau_0^{2*} = T_{02}(j^*)$
- Compute regularized multivariate Bayesian estimator $\tilde{\beta}_b$ as:

$$
\tilde{\beta}_b = \frac{\hat{\tau}_{\tilde{Y}X}}{(1-\omega^*)\tau_0^{2*} + \omega^*\hat{\tau}_X^2}
\tag{62}
$$

- The distribution of the regularized multivariate Bayesian estimator $\tilde{\beta}_b$ with optimal parameters ω^* and τ_0^{2*}:

$$
\frac{k_B\theta_B}{k_{sum2}\theta_{sum2}}\tilde{\beta}_b \sim F(2k_{sum2}, 2k_B)
\tag{63}
$$

where:

$$
k_B = \frac{(\omega^*\theta_{sum1}k_{sum1} + (1-\omega^*)\tau_0^{2*})^2}{\omega^{*2}\theta_{sum1}^2 k_{sum1}}
\tag{64}
$$

$$
\theta_B = \frac{\omega^{*2}\theta_{sum1}^2 k_{sum1}}{\omega^*\theta_{sum1}k_{sum1} + (1-\omega^*)\tau_0^{2*}}
\tag{65}
$$

- Compute standard error of $\tilde{\beta}_b$ from its distribution as:

$$
SE\left(\tilde{\beta}_b\right) = \frac{\theta_{sum2}}{\theta_B\,(k_B-1)}\sqrt{\frac{k_{sum2}\,(k_B+k_{sum2}-1)}{k_B-2}}
\tag{66}
$$

The standard error performs well in relatively large samples. For smaller samples, we suggest using resampling procedures to obtain standard errors, such as the delete-d jackknife approach (Shao and Wu 1989, see also Zitzmann 2018).

■

Appendix C: Tables

See Tables 1, 2, 3 and 4.

Table 1 Average RMSE and relative bias values of the regularized Bayesian estimator ($RMSE_{Bay}$ and RB_{Bay}, respectively), the standard ML ($RMSE_{ML_{std}}$ and $RB_{ML_{std}}$, respectively), and the transformed ML ($RMSE_{ML_{tr}}$ and $RB_{ML_{tr}}$, respectively) for estimation of the between-group parameter β_b across different values of n and J

n	J	$RMSE_{Bay}$	$RMSE_{ML_{std}}$	$RMSE_{ML_{tr}}$	RB_{Bay}	$RB_{ML_{std}}$	$RB_{ML_{tr}}$
5	5	2.253	2364.9044	71.5921	−53.0282	−2222.0905	105.4727
5	10	1.298	663.6092	56.8486	−46.115	3530.4663	21.0517
5	20	0.8143	97.7846	161.6247	−30.9842	479.2251	574.417
5	30	0.6448	61.2553	51.9519	−22.2578	−36.3986	−81.1739
5	40	0.5611	104.2087	27.9186	−19.7638	96.6046	73.1663
15	5	1.2614	165.2694	112.4208	−2.0487	224.2394	209.4308
15	10	0.7623	80.7334	82.3766	−19.7352	8.8636	−56.3894
15	20	0.5146	58.5046	22.6432	−26.2599	−253.4795	27.5943
15	30	0.4166	19.6364	6.8662	−30.9451	6.1411	80.9494
15	40	0.3541	20.3744	11.7295	−35.9619	6.8211	87.3609
30	5	0.9853	409.7473	149.1664	3.1077	−3377.0196	−1.2029
30	10	0.6578	86.6451	39.8985	−0.7899	−2992.0016	31.4906
30	20	0.3701	43.8525	7.8371	3.9369	−165.6842	−254.5301
30	30	0.2551	24.3741	0.7218	3.1644	−0.0824	92.5981
30	40	0.2314	5.4566	0.639	−0.5986	182.6504	27.2748

Table 2 Average coverage rate and standard error ratio values of the regularized Bayesian estimator (CR_{Bay} and SER_{Bay}, respectively), the standard ML ($CR_{ML_{std}}$ and $SER_{ML_{std}}$, respectively), and the transformed ML ($CR_{ML_{tr}}$ and $SER_{ML_{tr}}$, respectively) for estimation of the between-group parameter β_b across different values of n and J

n	J	CR_{Bay}	$CR_{ML_{std}}$	$CR_{ML_{tr}}$	SER_{Bay}	$SER_{ML_{std}}$	$SER_{ML_{tr}}$
5	5	98.07	87.3611	97.0056	3.3536	0.0795	0.4198
5	10	98.55	80.2706	97.6278	2.8084	0.05	0.762
5	20	90.8806	76.0417	98.3306	2.9753	0.1747	1.277
5	30	85.1683	74.5033	98.4606	2.8589	0.3781	1.1529
5	40	75.3561	76.1217	98.4906	1.8661	0.4984	1.8871
15	5	99.8189	87.3939	98.2639	4.4267	1.097	1.4124
15	10	95.085	81.6411	99.1044	3.3746	1.1921	1.8179

Table 2 continued

n	J	CR_{Bay}	$CR_{ML_{std}}$	$CR_{ML_{tr}}$	SER_{Bay}	$SER_{ML_{std}}$	$SER_{ML_{tr}}$
15	20	72.2817	86.8789	99.1856	2.1452	1.1855	1.7335
15	30	69.6222	85.3122	99.1344	1.593	1.214	1.7714
15	40	52.9456	86.1056	98.915	1.2601	1.2505	1.6569
30	5	94.9165	83.9644	93.6456	2.1917	0.0749	0.3921
30	10	95.0476	77.2767	93.8058	2.6469	0.0467	0.701
30	20	87.9862	73.3851	95.1125	2.7611	0.1632	1.1926
30	30	82.8528	72.4898	95.3796	1.9016	0.3414	1.0703
30	40	72.89	73.2047	95.2292	1.7563	0.4691	1.6835

Table 3 Average RMSE and relative bias values of the standard ML ($RMSE_{ML_{std}}$ and $RB_{ML_{std}}$, respectively), and the transformed ML ($RMSE_{ML_{tr}}$ and $RB_{ML_{tr}}$, respectively) for estimation of the covariate parameter γ across different values of n and J

n	J	$RMSE_{ML_{std}}$	$RMSE_{ML_{tr}}$	$RB_{ML_{std}}$	$RB_{ML_{tr}}$
5	5	2364.9045	71.5921	2472.214	−25.2863
5	10	663.6091	56.8486	−1438.9018	−22.108
5	20	97.7846	161.6247	−98.8264	−22.1302
5	30	61.2553	51.9519	−34.3401	−23.4084
5	40	104.2087	27.9186	−15.865	−24.5338
15	5	71.9913	0.1961	−94.8047	−18.7221
15	10	23.5566	0.1528	3.6916	−16.8186
15	20	22.7572	0.1206	78.9697	−15.9531
15	30	5.1038	0.1047	−6.0068	−15.4905
15	40	5.2125	0.0952	3.1951	−15.1146
30	5	19.3898	12.3358	10.2772	5.6034
30	10	12.3239	4.969	−231.8906	−4.2588
30	20	2.7321	11.1771	−11.5875	−0.1379
30	30	4.2404	3.6634	30.9181	−1.718
30	40	4.3897	1.8202	−3.5585	2.3588

Table 4 Average coverage rate and standard error ratio values of the standard ML ($CR_{ML_{std}}$ and $SER_{ML_{std}}$, respectively), and the transformed ML ($CR_{ML_{tr}}$ and $SER_{ML_{tr}}$, respectively) for estimation of the covariate parameter γ across different values of n and J

n	J	$CR_{ML_{std}}$	$CR_{ML_{tr}}$	$SER_{ML_{std}}$	$SER_{ML_{tr}}$
5	5	87.9722	93.2122	0.0756	0.8947
5	10	85.8333	84.5972	0.0578	0.8188
5	20	86.2433	77.125	0.193	0.7677
5	30	86.1894	70.4217	0.3861	0.7449
5	40	88.4128	67.8422	0.5438	0.73
15	5	90.88	89.8061	0.1205	0.7486
15	10	89.3344	78.9544	0.2532	0.6931
15	20	91.7683	74.1622	0.6957	0.6681
15	30	90.4217	72.0894	0.7793	0.6636
15	40	92.4472	69.5989	0.8469	0.6632
30	5	85.2558	89.8141	0.0754	0.8944
30	10	82.5054	81.2512	0.0581	0.8178
30	20	82.9987	74.4928	0.1962	0.7597
30	30	82.7329	68.0252	0.3893	0.7476
30	40	85.8832	65.5986	0.5448	0.7353

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s41237-025-00264-7.

Funding This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 471861494. Open Access funding enabled and organized by Projekt DEAL.

Data Availibility All data used in this study are provided in Supplement A. Simulated data to estimate the regularized Bayesian, standard and transformed ML estimators were generated in MATLAB. The resulting files ran using seeds from 1 to 540, which corresponds exactly to the cases shown in Supplement A in Tables (grouped by 4): S1–S4, S5–S8, S9–S12, S13–S16, S17–S20, S21–S24, S25–S28, S29–S32. Each group of Tables covers computed evaluation measures for all 540 cases. Tables 1, 2, 3 and 4 in Appendix C reformulate Tables S1–S32 across number of groups and group sizes. Twister random number generator has been used. For reproducibility, the command in MATLAB to regenerate all data will be `rng(c,'twister');` where c runs from 1 to 540 (the seed number). A MATLAB script to replicate the simulation process and generate the data for all 540 cases is available in Supplement B and will be made available on the journal web page.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Ethical Approval This study does not require ethical approval as it relies solely on simulated data.

References

Angrist JD, Pischke J-S (2008) Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press, New Jersey
Asparouhov T, Muthén B (2021) Bayesian Analysis of Latent Variable Models using Mplus (Version 5). Technical report, Mplus
Asparouhov T, Muthén B (2021) Bayesian estimation of single and multilevel models with latent variable interactions. Struct Equ Modeling 28(2):314–328
Dashuk V, Hecht M, Lüdtke O, Robitzsch A, Zitzmann S (2024) An optimally regularized estimator of multilevel latent variable models with improved MSE performance. Manuscript submitted for publication
Gelman A, Hill J (2007) Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press, Cambridge
Greenland S (2000) Principles of multilevel modelling. Int J Epidemiol 29(1):158–167
Hamaker E, Klugkist I (2011) Bayesian estimation of multilevel models. In: Hox J, Roberts JK (eds) Handbook of Advanced Multilevel Analysis. Routledge, pp 137–161
Lüdtke O, Marsh HW, Robitzsch A, Trautwein U, Asparouhov T, Muthén B (2008) The multilevel latent covariate model: a new, more reliable approach to group-level effects in contextual studies. Psychol Methods 13(3):203–229
Lüdtke O, Robitzsch A, Kenny DA, Trautwein U (2013) A general and flexible approach to estimating the social relations model using Bayesian methods. Psychol Methods 18(1):101–119
Morin A, Blais A-R, Chénard-Poirier LA (2022) Doubly latent multilevel procedures for organizational assessment and prediction. J Bus Psychol 37(1):47–72
Muthén B, Asparouhov T (2012) Bayesian structural equation modeling: a more flexible representation of substantive theory. Psychol Methods 17(3):313–335
Natarajan R, Kass RE (2000) Reference Bayesian methods for generalized linear mixed models. J Am Stat Assoc 95(449):227–237
Penrose R (1955) A generalized inverse for matrices. Math Proc Cambridge Philos Soc 51(3):406–413
Shao J, Wu CFJ (1989) A general theory for jackknife variance estimation. The Annals of Statistics, 17(3)
Smid SC, McNeish D, Miočević M, van de Schoot R (2020) Bayesian versus frequentist estimation for structural equation models in small sample contexts: A systematic review. Struct Equ Modeling 27(1):131–161
Wooldridge JM (2009) Introductory Econometrics: A Modern Approach. South-Western Cengage Learning, 4th edition
Zitzmann S (2018) A computationally more efficient and more accurate stepwise approach for correcting for sampling error and measurement error. Multivar Behav Res 53:612–632
Zitzmann S, Helm C, Hecht M (2021) Prior specification for more stable Bayesian estimation of multilevel latent variable models in small samples: A comparative investigation of two different approaches. Front Psychol 11:1–11
Zitzmann S, Lüdtke O, Robitzsch A (2015) A Bayesian approach to more stable estimates of group-level effects in contextual studies. Multivar Behav Res 50(6):688–705
Zitzmann S, Lüdtke O, Robitzsch A, Hecht M (2021) On the performance of Bayesian approaches in small samples: A comment on Smid, McNeish, Miocevic, and van de Schoot (2020). Struct Equ Modeling 28(1):40–50
Zitzmann S, Lüdtke O, Robitzsch A, Marsh HW (2016) A Bayesian approach for estimating multilevel latent contextual models. Struct Equ Modeling 23(5):661–679
Zitzmann S, Nagengast B, Hübner N, Hecht M (2024) A simple solution to heteroscedasticity in multilevel nonlinear structural equation modeling. Manuscript submitted for publication