AIMS Mathematics, 11(3): 6955–6988.
DOI: 10.3934/math.2026286
Received: 14 December 2025
Revised: 28 February 2026
Accepted: 02 March 2026
Published: 17 March 2026

https://www.aimspress.com/journal/Math

*Research article*

# Robust Bayesian empirical likelihood estimation for linear mixed-effects models

**Youxi Luo, Shiqi Zhou, Chaozhu Hu and Hanfang Li***

School of Science, Hubei University of Technology, Wuhan, China, 430068

* **Correspondence:** Email: 20020015@hbut.edu.cn.

**Abstract:** In this study, we propose a Bayesian Empirical Likelihood (BEL) method for linear mixed-effects models by integrating Huber-type influence functions and random effects annihilation matrices. At the group-level, we constructed several interpretable moment conditions, including robust residual balance moments and slope-related moments obtained after eliminating random intercepts through the annihilation matrix. We employed a block-diagonal weighting matrix to harmonize scaling across moment blocks. The BEL posterior distribution was derived by combining empirical likelihood with the prior distribution, and a Metropolis-Hastings algorithm for practical computation was designed. In simulation studies under adverse conditions, the finite-sample behavior of the proposed BEL method was compared with that of standard empirical likelihood (EL), restricted maximum likelihood (REML), maximum likelihood (ML), fully parametric Bayesian estimation (Bayes), and generalized moment estimation (GMM). The results showed that under correctly specified Gaussian LMMs, BEL attains efficiency comparable to parametric methods; whereas under adverse conditions, BEL generally achieves smaller mean squared errors than other methods, and its robustness is comparable to GMM. The application of real panel data further demonstrated that BEL can mitigate the impact of observations in extreme regions while preserving the major trend relationships, thus obtaining a more stable and reliable fixed effects estimate in real-world situations where model misspecification and atypical observations are difficult to avoid.

**Keywords:** linear mixed-effects model; Huber-type influence function; Bayesian empirical likelihood; annihilation matrix
**Mathematics Subject Classification:** 62J12, 65C60

## 1. Introduction

Linear mixed-effects models (LMMs) are a core tool for analyzing longitudinal and clustered data. Laird & Ware [1] first proposed a framework for longitudinal data analysis based on random effects and formalizing the decomposition into fixed and random effects. Pinheiro & Bates [2] developed the theoretical foundations and practical construction of mixed-effects and nonlinear mixed-effects models, which has shaped contemporary mixed-effects modeling. Mixed-effects models have been widely applied across psychology and related fields, including education [3], medical clinical trials [4], and linguistics [5].

In the conventional LMMs inference, parameter estimation and inference typically rely on likelihood-based procedures such as maximum likelihood (ML) and restricted maximum likelihood (REML) procedures. These methods are efficient when the model is correctly specified with high-quality data, but they are sensitive to misspecification and outliers. With the growing availability of complex hierarchical, nested, and unbalanced data in population health monitoring, education quality tracking, and financial risk supervision, there is increasing demand for the efficient estimation and robust inference of LMMs. This has motivated extensions of LMMs parameter estimation methods from robustness and semi-parametric and non-parametric directions.

From the perspective of robust estimation, one approach is to modify the likelihood distribution model by introducing t-distributions, skewed normal distributions, or skewed heavy-tailed distributions to attenuate the influence of outliers and heavy-tailed distributions. Some scholars have introduced robust linear mixed-effects models into the random effects and prediction terms and develop high-likelihood-based inference procedures [6]; Wang [7] extended the model to mixed normal distributions, improving the ability to accommodate heavy tails at the random effects level. These methods mitigate the impact of outliers on estimation to some extent, but they typically require explicit specification of the heavy-tailed distribution and covariance structure. When the random-effects dimension is large or the covariance structure is complex, model specification and numerical implementation become difficult. Another approach is robust regression based on M-estimation and Huber loss. Huber [8] proposed the Huber loss function, which combines the advantages of least squares and minimum absolute bias, and the associated $\psi$ function maintains linearity in small residual intervals and truncates in large residual intervals, thereby balancing between efficiency and robustness. Building upon this, Mohammadi & Kazemi [9] introduced robust regression methods into a parameterized environment, constructing robust models based on multivariate skew-Huber distributions. This framework is suitable for longitudinal data with heavy tails and skewed structures. Wu [10] introduced Huber loss into a mixed-effects model of longitudinal data. Overall, studies indicate that the Huber function is effective in mitigating heavy tails and anomalous perturbations. Nevertheless, these approaches largely remain within the parametric paradigm: On the one hand, a complete data generation distribution needs to be provided in the likelihood; on the other hand, robustness is usually achieved through weight functions or the shape of the distribution tails, and its performance may be limited when the model is severely misspecified or the distribution of random effects is unknown. In large-scale longitudinal and stratified data, outliers, heavy-tailed errors, and stratified outliers are rare in practice. They may appear at the observation level or be reflected in the skewness and heavy tails of the stratified random effects distribution. ML and REML inferences for classical LMMs are prone to fixed-effects bias, variance component distortion, and overconfident interval estimation when the distribution setting deviates. Furthermore, robust inference of mixed-effects models also faces the following structural challenges: (1) Random effects and error

distributions are often difficult to parameterize reliably, and a misconfiguration of the covariance structure will amplify the inference error; (2) when the dimension of random effects and the dimension of covariates increase, the cost of likelihood integration and numerical optimization increases significantly; and (3) when using high-dimensional moment conditions to characterize robust constraints, empirical likelihood may show "feasibility failure" where the convex hull does not contain zeros, requiring an operable diagnostic and correction mechanism; otherwise, it is difficult to determine when the method fails. The above factors together illustrate that it is practically necessary to build a robust inference system with weak distribution dependence, computability, and diagnostics under the LMMs framework.

Against the backdrop of the increasing limitations of robust parameterized models, empirical likelihood (EL) avoids the requirement for complete parameterization of random effects and error distributions based on moment conditions. Combined with robust estimating equations, it can improve performance in heavy-tailed and anomalous scenarios and is widely used for nonparametric and semiparametric inference based on estimation equations. Wang [11] reviewed the foundations of empirical likelihood. Qin & Lawless [12] combined empirical likelihood with generalized estimating equations (GEE) and revealed the effectiveness of empirical likelihood in estimation equation scenarios. Subsequently, Tang et al. [13] and Leng [14] further extended empirical likelihood to more complex scenarios, such as missing data and high-dimensional estimation equations, establishing improved and penalized empirical likelihood methods. In the context of mixed-effects and longitudinal data, Chen et al. [15] constructed an effective and robust empirical likelihood statistic for linear mixed-effects models, enabling distribution-free inference for fixed-effects parameters through appropriate moment conditions. The empirical likelihood method achieves an inference framework that combines flexibility and efficiency through moment conditions, providing a natural basis for its subsequent combination with Bayesian methods. Liu [16] discussed the hierarchical structure and inference strategy of mixed-effects models from a Bayesian perspective. Ouyang & Bondell [17] used empirical likelihood to perform Bayesian analysis on longitudinal data and developed a fully Bayesian method to analyze longitudinal data based on a set of moment equations parallel to the form of the generalized estimation equation.

To incorporate prior information while retaining the advantages of nonparametric and moment-conditional empirical likelihood, Lazar [18] first proposed BEL and established its theoretical validity; Lancaster & Jae Jun [19] discussed the statistical inference problem of Bayesian quantile regression from the perspective of empirical likelihood; Porter et al. [20] introduced a general hierarchical Bayesian framework that integrates flexible nonparametric data model specifications using empirical likelihood methods. Vexler et al. [21] developed BEL in quantile regression; Wang et al. [22] extended the Bayesian empirical likelihood (BEL) method to composite quantile regression models; Dong et al. [23], to avoid the arduous task of directly optimizing empirical likelihood, constructed a BEL function under B-J estimation and used the M-H algorithm to obtain point estimates and confidence regions. Furthermore, BEL has been explored in complex structures such as spatial small-area estimation [24], stochastic fluctuation models [25], variable selection, and partially linear regression. Studies show that Empirical likelihood (EL) can achieve distribution-free inference at the estimation equation level and has been incorporated with mixed-effects and longitudinal data background to avoid full likelihood specification. BEL, by embedding EL into a Bayesian framework to introduce prior information, can stabilize posterior uncertainty quantification in small samples and complex structures. However, there are important gaps that remain when extending BEL to classical LMMs: The high-dimensional integrals and covariance structures brought about by random effects significantly increase the

computational burden of "direct Bayesianization"; robust moment conditions, if not addressed for the structure of random effects, are easily driven by group outliers and become unbalanced; furthermore, the convex hull feasibility, dual solution stability, and failure diagnosis of EL in high-dimensional moment systems can degrade. Therefore, there is an urgent need for a robust inference scheme for BEL-LMMs that can eliminate the influence of random effects at the moment condition level, stabilize high-dimensional moment systems at the numerical level, and provide a diagnostic mechanism at the inference level.

Based on the above motivations and research gaps, we propose a robust BEL framework for LMMs: The residuals are robustly processed by a Huber-type influence function to down-weight of outlier observations and heavy-tailed distributions at the estimation equation level; further, the random effects annihilation matrix is used to construct group-level moment conditions that do not explicitly depend on the form of random effects distribution, thereby avoiding strong specification of high-dimensional random effects integrals and covariance structures; and block-diagonal scaling is introduced to standardize different moment blocks to alleviate the instability of dual solutions caused by heterogeneous numerical scales in high-dimensional moment systems. Finally, the posterior distribution of fixed effects parameters and uncertainty quantification are obtained under the BEL framework.

Introducing robust BEL inference into linear mixed-effects models presents key challenges in the following four aspects: (1) Data anomalies in LMMs not only originate from the observation layer but also often appear in the form of "group-level clustering". If the complete parameterization of random effects and error distributions is relied upon, fixed-effects inference is easily dominated by a few extreme groups and produces systematic bias. (2) Under the mixed-effects structure, directly using residual moments often introduces random effects terms, making moment conditions sensitive to the distribution of random effects. Furthermore, when fixed effects include intercepts and multidimensional slopes, it is not straightforward to construct an over-identification system that can identify each component while avoiding moment condition degradation and collinearity. (3) In high-dimensional moment systems, empirical likelihood depends on convex hull conditions. After multiple types of moment conditions are combined, the dimension increases. EL may lead to the convex hull not containing zeros, leading to infeasibility or numerical instability. Therefore, a feasibility diagnosis and robust fallback mechanism are needed. (4) BEL posterior usually has no analytical form and requires MCMC simulation. However, each candidate parameter must be solved repeatedly with Lagrange multipliers and feasibility checked. The computational cost and convergence parameter tuning difficulty increase significantly with the number of groups and the moment dimension.

To address the aforementioned challenges, we introduce Huber-type transformations, robust centering, and scale standardization at the robust identification level to suppress the influence of extreme residuals. An annihilation matrix is used to remove random effect terms at the estimation equation level, thereby reducing dependence on the distribution and variance structure of random effects. At the moment condition level, we construct a joint system comprising intercept constraints, covariate-residual orthogonality, and annihilated slope moment conditions to ensure sufficient identification of each component of the fixed effects. At the numerical level, robust scale standardization is performed on different moment blocks through block-diagonal scaling, and dual feasibility judgment, Broyden-type quasi-Newton solvers, and adjusted empirical likelihood (AEL) fallback are combined to alleviate convex-hull boundary issues. Feasibility ratios, KKT residuals, minimum dual denominator, and Hessian eigenvalues are reported for diagnostic purposes. At the

posterior computation level, a pilot-adaptive M-H scheme is adopted, and we freeze the proposal covariance during the formal sampling phase to preserve ergodicity and ensure valid posterior inference. Convergence diagnostics such as acceptance rates, ESS, and multi-chain assessments are provided to ensure that the posterior summary is verifiable and reproducible. Based on the above ideas, the major contributions of this paper are: (1) A robust BEL inference framework for LMMs is proposed, and three types of group-level moment conditions with clear statistical meaning are constructed to achieve robust identification of fixed effects; (2) a random effect annihilation mechanism is introduced at the estimation equation level to reduce the interference of random effect distribution misconfiguration and group stratification outliers on fixed effect inference; and (3) for the numerical instability of high-dimensional moment systems, a default setting for block scaling and a sensitivity analysis approach are given to enhance the reproducibility of the method.

## 2. Materials and methods

### *2.1. Mixed-effects model*

A mixed-effects model is a statistical model that incorporates fixed and random effects. The "fixed effects" capture population-level effects of covariates in the study, while the "random effects" account for within-cluster dependence. It is widely used for hierarchical and repeated-measures data. Assuming there are $n$ observations, and the $i$ th individual has $m$ observations, then the total sample size has $N = nm$ observations. Let $Y_i$ represent the response variable for the $i$ th individual.

For linear mixed-effects models:

$$Y_i = X_i^T \beta + Z_i^T \alpha_i + \varepsilon_i, i = 1, \cdots, n$$

Assuming random effects $\alpha_i \sim N\left(0, \sigma_\alpha^2 I\right)$ and residuals $\varepsilon_i \sim N\left(0, \sigma_\varepsilon^2 I\right)$, $\alpha_i$ is an unobserved random variable, which can be eliminated by integration to obtain the marginal distribution of $Y_i \sim N\left(X_i^T \beta, V_i\right)$, where $V_i = \sigma_\alpha^2 Z_i^T Z_i + \sigma_\varepsilon^2 I$ is the marginal variance matrix and $Z$ is the random effects design matrix. The linear mixed-effects model estimates the parameters by maximizing the following log-likelihood function:

$$l\left(\beta, \sigma_\alpha^2, \sigma_\varepsilon^2\right) = -\frac{1}{2}\sum_{i=1}^{n}\left[\log\left|V_i\right| + \left(Y_i - X_i\beta\right)^T V_i^{-1}\left(Y_i - X_i\beta\right) + m\log\left(2\pi\right)\right]$$

Assuming the observed data are divided into "groups" $i = 1, 2, \cdots, n$, and the observations within each group are $j = 1, 2, \cdots, m$, consider a linear mixed-effects model:

$$Y_{ij} = X_{ij}^T \beta + Z_{ij}^T \alpha_i + \varepsilon_{ij}, i = 1, \cdots, n$$

where $\beta = \left(\beta_0, \beta_1, \cdots, \beta_p\right)^T$ is the $p+1$ -dimensional fixed effects vector, $X_{ij}$ is the covariate corresponding to the fixed effects, $\alpha_i = \left(\alpha_{i1}, \alpha_{i2}, \cdots, \alpha_{iq}\right)^T$ is the $q$ -dimensional random effects vector of the $i$ -th individual, $Z_{ij}$ is the covariate corresponding to the random effects, and $\varepsilon_{ij}$ is the independent and identically distributed random error term with zero mean.

The model satisfies the following general assumptions: 1) The random effects $\alpha_i$ are independent and identically distributed, $E(\alpha_i)=0$ and $Var(\alpha_i)=\Sigma_\alpha$, where $\Sigma_\alpha$ is a finite positive definite matrix; 2) the errors $\varepsilon_{ij}$ are independent between different groups $i$, $E(\varepsilon_{ij}|x_{ij},z_{ij},\alpha_i)=0$ and $Var(\varepsilon_{ij})<\infty$; 3) the random effects $\alpha_i$, errors $\varepsilon_{ij}$, and covariates $x_{ij}$ and $z_{ij}$ satisfy orthogonality, $E(\alpha_i|x_{ij},z_{ij})=0$, $E(\varepsilon_{ij}x_{ij})=0$, and $E(\varepsilon_{ij}z_{ij})=0$; 4) for each $i$, the random effects design matrix $Z_i$ is full column rank and $m>q$, so $Z_i^TZ_i$ is invertible, which ensures the existence of the annihilation matrix.

*2.2. Bayesian empirical likelihood estimation*

Empirical likelihood estimation is a nonparametric statistical method that does not require specifying a parametric form for the data-generating distribution. Let $X_1,X_2,\cdots,X_n\in R^P$ be independent entities with a common cumulative distribution $F$. Then the nonparametric likelihood of $F$ is $L(F)=\prod_{i=1}^{n}F(X_i)$, where $F(X_i)$ is the probability mass of distribution $F$ at point $X_i$. Under the constraint of the moment condition $E(g(X_i,\beta))=0$, empirical likelihood estimation is performed by maximizing the weighted likelihood function. The empirical likelihood of parameter $\beta$ is defined as:

$$L(\beta)=\max\left\{\prod_{i=1}^{n}p_i\,\middle|\,p_i\geq 0,\sum_{i=1}^{n}p_i=1,\sum_{i=1}^{n}p_ig(X_i,\beta)=0\right\}$$

The problem of maximizing the empirical likelihood can be formulated as a constrained optimization problem. By calculating $p_i=\frac{1}{n(1+\lambda^Tg_i(\beta))}$ using the Lagrange multiplier method, and given that equation $\lambda^T(\beta)$ satisfies $\sum_{i=1}^{n}\frac{g(X_i,\beta)}{1+\lambda^T(\beta)g(X_i,\beta)}=0$, the empirical likelihood function can be obtained as $L(\beta)=\prod_{i=1}^{n}\frac{1}{n(1+\lambda^Tg_i(\beta))}$.

Using the empirical likelihood function as the likelihood function within the BEL framework, the relevant information of the parameters to be estimated is given in the form of a prior distribution $\pi(\beta)$. The same likelihood is used to derive the posterior of the parameters for a given observation. The posterior can be defined as:

$$\pi(\beta|x)=\frac{L(\beta)\pi(\beta)}{\int L(\beta)\pi(\beta)d\theta}\propto L(\beta)\pi(\beta)$$

*2.3. Bayesian empirical likelihood estimation of linear mixed-effects models*

To enhance robustness, based on Huber's proposed Huber loss function $\rho_c$ and its derivative Huber transform function $\psi_c$, which satisfy the properties of "linearity in the middle, information retention in small residuals, and truncation in large residuals," Shuangyue Wu [10] constructed a weighted residual squared using the Huber loss function that combines the advantages of $L_1$ and $L_2$ loss functions. The difference is that we introduce the Huber transform function into the moment condition. Given $\beta$, the residual vector of the $i$ th group is defined as:

$$r_i(\beta) = Y_i - X_i^T\beta = \left(r_{i1}(\beta), \cdots, r_{im}(\beta)\right)^T \tag{1}$$

For any $r$, define the Huber function:

$$\psi(r) = \begin{cases} r, |r| \le c \\ csign(r), |r| > c \end{cases} \tag{2}$$

where $c > 0$ is a given constant. Perform robust centering $\mu_r$ and scaling $s_r$ standardization on the residuals, and denote the standardized residuals $u_{ij}(\beta) = \dfrac{r_{ij}(\beta) - \mu_r}{s_r}$ and $\psi_i(\beta) = \left[\psi\left(u_{i1}(\beta)\right), \cdots, \psi\left(u_{im}(\beta)\right)\right]^T$.

The fixed-effects design matrix is decomposed into $X_i = (1_m, W_i)$, where $1_m$ is an $m$ -dimensional vector of all 1s, and $W_i$ is the covariate matrix after removing the intercept column corresponding to the slope parameter $\beta = \left(\beta_1, \cdots, \beta_p\right)^T$. To extend the scalar residual moments into vector moments that match the $\beta$ dimension, we need to ensure that the $(\beta_0, \beta_1)$ pair has sufficient identification information and avoid moment condition degradation or repetition. Using the residual vector $r_i(\beta)$ and the covariate matrix $X_i$, two types of moment conditions independent of the specific distribution of random effects are constructed. If only conditions without $W_i$ are used, the conditions mainly constrain the intercept but provide insufficient information about the slope $\beta_1$. Furthermore, under covariate centering or symmetric designs, a mismatch between the moment conditions and parameter directions may occur, leading to insufficient rank of $E\left(\dfrac{\partial g(\beta)}{\partial \beta}\right)$. Introducing $W_i$ makes the moment conditions equivalent to orthogonalizing the robust residuals to the covariate directions: $E\left\{W_i^T\psi_i\left(\beta^0\right)\right\} = 0$. This provides independent estimation equations for $\beta_0$ and $\beta_1$, improving finite sample efficiency and numerical stability. The first type of moment condition is defined as follows:

$$g_{i1}(\beta) = \frac{1}{m} 1_m^T \psi_i(\beta) \tag{3}$$

This moment condition describes that "the mean of the within-group residuals is 0 in a robust sense,"

primarily providing information for a fixed intercept $\beta_0$. The second kind of moment condition is defined to provide constraints on each slope parameter $\beta_1,\cdots,\beta_p$ as follows:

$$g_{i2}(\beta)=\frac{1}{m}W_i^T\psi_i(\beta) \tag{4}$$

To mitigate the impact of random effects distribution errors on slope estimation, an annihilation matrix $A_i=I_m-Z_i\left(Z_i^TZ_i\right)^{-1}Z_i^T$ is used. By the definition of an annihilation matrix, $A_iZ_i=0$. Multiplying both sides of the model by $A_i$ on the left yields:

$$A_iy_i=A_iX_i\beta+A_iZ_i\alpha_i+A_i\varepsilon_i=A_iX_i\beta+A_i\varepsilon_i \tag{5}$$

At this point, the random effect term $Z_i\alpha_i$ is eliminated. Let $\tilde{y}_i=A_iy_i$ and $\tilde{X}_i=A_iX_i=\left(A_i1_m,\tilde{W}_i\right)$, where $\tilde{W}_i=A_iW_i$ is the transformation of the slope covariate under the annihilation matrix. Then, at the truth value $\beta^0$, $\tilde{y}_i=\tilde{X}_i\beta^0+A_i\varepsilon_i$. Based on this, the residual vector after annihilation is defined as follows:

$$r_i^A(\beta)=\tilde{y}_i-\tilde{X}_i\beta=A_iy_i-A_iX_i\beta \tag{6}$$

Let $\tilde{u}_{ij}(\beta)=\frac{r_{ij}^A(\beta)-\tilde{\mu}_r}{\tilde{s}_r}$ and $\psi_i^A(\beta)=\left[\psi\left(\tilde{u}_{i1}(\beta)\right),\cdots,\psi\left(\tilde{u}_{im}(\beta)\right)\right]^T$ be the conditions for constructing the third kind of moments:

$$g_{i3}(\beta)=\tilde{W}_i^T\psi_i^A(\beta) \tag{7}$$

Similarly, left to multiply the covariate matrix after transformation under the annihilation matrix is retaining the identification information of the slope parameter after eliminating the random intercept or group layer translation. When $Z_i=1_m$ corresponds to a purely random intercept model, $A_i$ degenerates into a within-group centered matrix $A_i=I_m-\frac{1}{m}1_m1_m^T$. In this case, the above moment condition is the within-group mean-reducing slope moment condition commonly used in the special case of the random intercept.

To ensure the validity of the moment conditions under the empirical likelihood framework, the following conditional centralization assumption is adopted: At the truth value $\beta^0$, A has a robust center and scales $(\mu_r,s_r)$ and $(\tilde{\mu}_r,\tilde{s}_r)$ such that, given $(X_i,Z_i)$, $E\left[\psi_i\left(\beta^0\right)|X_i,Z_i\right]=0$, $E\left[\psi_i^A\left(\beta^0\right)|X_i,Z_i\right]=0$, and $E\left\|W_i^T\psi_i\left(\beta^0\right)\right\|<\infty$, $E\left\|\tilde{W}_i^T\psi_i^A\left(\beta^0\right)\right\|<\infty$, we have:

$$E\left(g_{i1}\left(\beta^0\right)|X_i,Z_i\right)=\frac{1}{m}1_m^TE\left(\psi_i\left(\beta^0\right)|X_i,Z_i\right)=0$$

$$E\left(g_{i2}\left(\beta^0\right)|X_i,Z_i\right)=\frac{1}{m}W_i^TE\left(\psi_i\left(\beta^0\right)|X_i,Z_i\right)=0$$

$$E\left(g_{i3}\left(\beta^0\right)|X_i,Z_i\right)=\frac{1}{m}\tilde{W}_i^TE\left(\psi_i^A\left(\beta^0\right)|X_i,Z_i\right)=0$$

Taking the full expectation of the conditional expectation yields: $E\left(g_{i1}\left(\beta^0\right)\right)=0$, $E\left(g_{i2}\left(\beta^0\right)\right)=0$, $E\left(g_{i3}\left(\beta^0\right)\right)=0$. In actual calculations, the sample median and MAD are used for plug-in centering and scaling estimation, respectively.

From the perspective of construction ideas, the three types of Huber-type group-level moment conditions proposed in this paper can be regarded as a unified generalization of the empirical likelihood moment conditions of the classical linear mixed-effects model and the BEL moment conditions under the GEE framework. When the Huber cutoff constant tends to infinity, the residuals obtained by selecting the scale parameter are standardized and approach ordinary residuals, and the annihilation matrix is ignored, that is, when the annihilation matrix is taken, the robust residual moment conditions in Eqs (3) and (4) degenerate into linear moments based on conditional mean residuals. Its form is consistent with the residual-balance moments and covariate-residual orthogonality moments used in EL work under the linear mixed-effects model. It corresponds to the estimation equation of fixed effects characterized by group-level residuals when random effects have not been annihilated. When random effects are not explicitly modeled and only the covariate-residual orthogonal information at the cluster level is retained, the working variance structure can be replaced by the inverse working-correlation matrix in the GEE framework. In that case, the second and third types of moment conditions have the same structure at the cluster level as the robust estimation equation used by GEE-BEL. Both use covariate and residual orthogonality to characterize marginal fixed effects. Relative to moment conditions, we introduce a Huber-type transformation function within each group to truncate and balance heavy-tailed errors, in-group outliers, and skewness in random effect. Furthermore, it explicitly constructs an annihilation matrix, identifying slope effects and eliminating interference from random intercepts through in-group variations without making specific parameterization assumptions about the random effect distribution. Therefore, the three types of Huber-type moment conditions presented in this paper are compatible with the scoring structures of existing LMM-EL and GEE-BEL estimation equations in the limiting cases of ordinary residuals and without annihilation matrices. In the general case, the Huber transformation and annihilation matrix enhance robustness against heavy-tailed errors, misspecified random effects, and group outliers at the moment condition level.

Combining the three types of moment conditions mentioned above, a basic moment condition vector $g_i\left(\beta\right)=\begin{pmatrix} g_{i1}\left(\beta\right) \\ g_{i2}\left(\beta\right) \\ g_{i3}\left(\beta\right) \end{pmatrix}$ is defined for each group $i$, where $g_{i1}\left(\beta\right)\in\mathbb{R}$, $g_{i2}\left(\beta\right)\in\mathbb{R}^p$, and $g_{i3}\left(\beta\right)\in\mathbb{R}^p$, and the dimension of the fixed effects parameter is $p+1$. Therefore, when $p\geq 1$, $\dim\left\{g_i\left(\beta\right)\right\}=1+2p>p+1=\dim\left(\beta\right)$, forming an over-identified moment condition system. The truth value $\beta^0$ satisfies $E\left(g_i\left(\beta^0\right)\right)=0, i=1,\cdots,n$; thus, $g_i\left(\beta\right)$ can be used to construct empirical likelihood and BEL.

In practical applications, the numerical dimensions and variances of different moment conditions can differ significantly. For example, $g_{i1}\left(\beta\right)\in\mathbb{R}$ based on the residual mean usually exhibits less fluctuation, while the slope moment condition $g_{i3}\left(\beta\right)\in\mathbb{R}^p$ based on the annihilation matrix may be more sensitive when the random effects distribution has heavy tails or the error distribution is misspecified. If the original $g_i\left(\beta\right)$ is used directly, the corresponding Lagrange multiplier $\lambda\left(\beta\right)$

in solving the empirical likelihood may overly favor a certain type of moment condition, thus exhibiting a bias towards intercept or slope estimation under finite sample conditions.

To balance the numerical effects of different moment conditions and improve the numerical solution stability of Lagrange multipliers in empirical likelihood dual problems, for each group $i$, a non-singular weight matrix $W \in \mathbb{R}^{(1+2p)\times(1+2p)}$ is introduced, and the weighted moment condition is defined as $\tilde{g}_i(\beta) = Wg_i(\beta)$. Obviously, $E\left\{\tilde{g}_i\left(\beta^0\right)\right\} = WE\left\{g_i\left(\beta^0\right)\right\} = 0$ still holds true under the true value $\beta^0$. Therefore, the weighted moment conditions are equivalent to the original moment conditions in identifying the target parameter $\beta^0$, and will not change the objective of the BEL. It affects only the utilization of information in different directions and numerical stability under finite samples. Although fixed weights can theoretically be absorbed by the dual variable, the magnitudes of different moment conditional blocks may differ significantly in a finite sample, leading to ill-conditioned KKT equations for $\lambda(\beta)$ and step size sensitivity, manifesting as dual solution failure or the need for frequent use of AEL as a fallback. Therefore, we will use $W$ as a preconditioning parameter to ensure that each moment block is numerically similar in scale, thereby improving the stability of $\lambda$ solution and the mixing efficiency of MCMC.

In this paper, we use a block-diagonal form of the weight matrix $W = diag\left(w_1, w_2 I_p, w_3 I_p\right)$ to maintain the interpretability of the structure. First, we calculate the moment conditions $g_i\left(\beta^\dagger\right)$ for all groups at the maximum likelihood estimation point $\beta_{ML}$. $\beta_{ML}$ is used only to estimate the relative scale of each moment block. This scaling does not change the recognition structure of the unbiased moment conditions. Define a robust scale for each moment block: $s_1 = MAD_i\left(g_{i1}\left(\beta^\dagger\right)\right)$, $s_2 = MAD_i\left(\left\|g_{i2}\left(\beta^\dagger\right)\right\|_2\right)$, $s_3 = MAD_i\left(\left\|g_{i3}\left(\beta^\dagger\right)\right\|_2\right)$. If the corresponding MAD is too small or unstable, it degenerates into the sample standard deviation. Construct the block-diagonal matrix $S = diag\left(s_1, s_2, s_3\right)$, and by default, take the scalar weight matrix $W = diag\left(w_1, w_2, w_3\right) = diag\left(\frac{1}{s_1}, \frac{1}{s_2}, \frac{1}{s_3}\right)$. Here, $w_1$、$w_2$、$w_3 > 0$ represents scalar weights, corresponding to the intercept moment condition $g_{i1}(\beta)$, the covariate-residual moment condition $g_{i2}(\beta)$, and the slope moment condition $g_{i3}(\beta)$ after annihilating random effects, respectively. The weighted moment condition is $\tilde{g}_i(\beta) = \begin{pmatrix} w_1 g_{i1}(\beta) \\ w_2 g_{i2}(\beta) \\ w_3 g_{i3}(\beta) \end{pmatrix}$. Furthermore, the sensitivity of the estimation results to block-diagonal scaling is examined by randomly perturbing $W$ based on the default settings. For each simulation, the baseline scaling matrix W and the EL point estimate $\hat{\beta}_{EL}$ are obtained at the empirical likelihood maximum point. Subsequently, 20 sets of perturbation weights $\tilde{w}_k^{(s)} = w_k \exp\left(Z_k^{(s)}\right)$ and $Z_k^{(s)} \sim N\left(0, \sigma_w^2\right)$ are independently generated. Geometric mean normalization is performed using $\prod_{k=1}^{3} \tilde{w}_k^{(s)} = 1$ to keep the overall scale constant. For each set of

perturbation weights $\tilde{W}^{(s)} = diag\left(\tilde{w}_1^{(s)}, \tilde{w}_2^{(s)}, \tilde{w}_3^{(s)}\right)$, the EL point estimate $\hat{\beta}_{EL}^{(s)}$ is recalculated, and its absolute deviation $\Delta^{(s)} = \left|\hat{\beta}_{EL}^{(s)} - \hat{\beta}_{EL}\right|$ relative to the baseline estimate is recorded.

BEL relies on the standard convex hull condition, which states that given $\beta$, the convex hull of the sample moment vector $\left(\tilde{g}_1(\beta), \cdots, \tilde{g}_n(\beta)\right)$ must contain the origin 0. If this condition is not met, there are no positive weights satisfying the moment and probability constraints, and the equation $\sum_{i=1}^{n} \frac{\tilde{g}_i(\beta)}{1+\lambda^T(\beta)\tilde{g}_i(\beta)} = 0$ no longer has a feasible solution; the empirical likelihood degenerates to $-\infty$ at this point. When the model is correct, the number of samples is sufficiently large, and the Huber truncation and scaling parameters are reasonably selected. Moreover, the probability that the convex hull condition holds within the neighborhood of the true value approaches 1 as the sample size increases; therefore, EL is feasible and has a unique solution for large samples. However, in finite samples, if the dimension of the moment condition is relatively high or the proposed parameter $\beta$ is far from the true value, the sample moment vectors may almost all fall within a certain half of the space, causing 0 to not be within the convex hull. In this case, the EL method is considered "ineffective" for that parameter value. In simulation experiments, the diagnostic results can be demonstrated through indicators such as the feasible ratio, the AEL usage ratio, KKT residuals, and dual Hessian eigenvalues.

Assuming there is a zero within the convex hull of $\left(\tilde{g}_1(\beta), \cdots, \tilde{g}_n(\beta)\right)$, the empirical likelihood function $L(\beta)$ has a unique value. Given $\beta$ and $\tilde{g}_i(\beta)$, we solve for the Lagrange multipliers $\lambda^T(\beta)$ such that $\sum_{i=1}^{n} \frac{\tilde{g}_i(\beta)}{1+\lambda^T(\beta)\tilde{g}_i(\beta)} = 0$, and calculate the empirical likelihood function $L(\beta) = \prod_{i=1}^{n} \frac{1}{n\left(1+\lambda^T\tilde{g}_i(\beta)\right)}$. Based on the empirical likelihood function, we assign a multivariate normal prior $\pi(\beta) \sim N(\mu_0, \Sigma_0)$ to the fixed effect parameter $\beta$, with the density function being:

$$\pi(\beta) = \frac{1}{\sqrt{(2\pi)^p |\Sigma_0|}} \exp\left\{-\frac{1}{2}(\beta-\mu_0)^T \Sigma_0^{-1}(\beta-\mu_0)\right\}$$

Combining the prior density function $\beta$ with the empirical likelihood function constructed based on the Huber moment condition, the posterior distribution is expressed as:

$$\pi(\beta|x) = \frac{1}{\sqrt{(2\pi)^p |\Sigma_0|}} \exp\left[-\frac{1}{2}(\beta-\mu_0)^T \Sigma_0^{-1}(\beta-\mu_0)\right] \prod_{i=1}^{n} \frac{1}{n\left(1+\lambda^T\tilde{g}_i(\beta)\right)}$$

Assume that the group-level observations are independent, and the moment condition $E\left\{g_i\left(\beta^0\right)\right\} = 0$ holds only at the true value $\beta^0$; $g_i(\beta)$ is differentiable in the neighborhood of $\beta^0$, and $E\left\{\frac{\partial g_i\left(\beta^0\right)}{\partial \beta^T}\right\}$ is full rank; the moment condition has finite second moments. Then, when the number of groups is $n \to \infty$, the empirical likelihood maximization estimate of $\hat{\beta}_{BEL}$ is

consistent, $\hat{\beta}_{BEL} \xrightarrow{P} \beta^0$ ; $\sqrt{n}\left(\hat{\beta}_{BEL} - \beta^0\right)$ asymptotically follows a normal distribution, $\sqrt{n}\left(\hat{\beta}_{BEL} - \beta^0\right) \xrightarrow{d} N\left(0, \left(G^T \Omega^{-1} G\right)^{-1}\right)$, where $G = E\left\{\frac{\partial g_i\left(\beta^0\right)}{\partial \beta^T}\right\}$ and $\Omega = E\left\{g_i\left(\beta^0\right)^T g_i\left(\beta^0\right)\right\}$ are constants.

It should be noted that the Bayesian empirical apparent posterior distribution constructed in this paper is defined on fixed-effects parameters. Random effects and their covariance matrices are not explicitly incorporated into the posterior hierarchy, but are eliminated at the estimation equation level through annihilation matrices and orthogonality moment conditions, thus avoiding specific parameterization assumptions about the form and variance structure of the random effects distribution. Therefore, the inference results obtained by the method in this paper are mainly applicable to the interpretation of average effects and overall trend analysis at the fixed-effects level. Within the BEL framework, obtaining the analytical form of the posterior distribution is often very difficult. Therefore, it is necessary to use the Markov Chain Monte Carlo (MCMC) method to simulate the posterior distribution.

### *2.4. The Metropolis-Hastings algorithm flow for bayesian empirical likelihood estimation under Huber-type moment conditions.*

The Metropolis-Hastings (MH) algorithm, one of the most commonly used MCMC algorithms, can efficiently generate samples while ensuring a stationary distribution as the target posterior. Its core idea is to continuously generate candidate points using a "proposal distribution" and combine it with an acceptance-rejection mechanism to construct a Markov chain that satisfies a detailed balance with respect to the target posterior.

Let $\beta^{(0)}$ be the initial value. For each state $\beta^{(t)}$, $t = 0,1,\cdots,T-1$, and calculate the logarithmic posterior $l_{BEL}\left(\beta^{(t)}\right)$ according to the following steps.

(1) Use $\beta^{(t)}$ to fit the linear part to calculate the residual $r_{ij}\left(\beta^{(t)}\right)$, and estimate the scale parameter $s\left(\beta^{(t)}\right)$ by the median absolute deviation (MAD) of the OLS residuals. If the MAD is too small or unstable, it degenerates into the sample standard deviation SD to obtain the standardized residual $u_{ij}\left(\beta^{(t)}\right) = \frac{r_{ij}\left(\beta^{(t)}\right)}{s\left(\beta^{(t)}\right)}$. Given the Huber cutoff constant $c$, which is usually taken as 1.345, it can be adjusted appropriately according to the situation. Obtain the Huber residual vector $\psi_{ij}\left(u_{ij}\left(\beta^{(t)}\right)\right)$ by the Huber influence function, and construct the weight moment condition $\tilde{g}_i\left(\beta^{(t)}\right) = W g_i\left(\beta^{(t)}\right)$ according to formulas (3), (4), and (7). The block-diagonal weight matrix $W$ can be adjusted according to experience. At this time, the random effect has been eliminated. When the moment conditions are significantly skewed, consider adding a robust centering correction to the residuals: In each iteration t, first calculate the full sample median $\hat{\mu}_r^{(t)} = median\left\{r_{ij}\left(\beta^{(t)}\right)\right\}$ of the

current residuals, and use $\psi\left(\frac{r_{ij}\left(\beta^{(t)}\right)-\hat{\mu}_r^{(t)}}{s\left(\beta^{(t)}\right)}\right)$ instead of $\psi\left(\frac{r_{ij}\left(\beta^{(t)}\right)}{s\left(\beta^{(t)}\right)}\right)$ to construct the Huber moments in the moment conditions. This correction can reduce the impact of random effects and error skew on the centering moment conditions without changing the overall trend.

(2) Under conditions $\beta^{(t)}$ and $\tilde{g}_i\left(\beta^{(t)}\right)$, $\lambda^T\left(\beta^{(t)}\right)$ is a solution to the nonlinear system of equations $F\left(\lambda;\beta\right)=\sum_{i=1}^{n}\frac{\tilde{g}_i\left(\beta^{(t)}\right)}{1+\lambda^T\left(\beta^{(t)}\right)\tilde{g}_i\left(\beta^{(t)}\right)}=0$, which typically has no analytical solution. We employ a quasi-Newton method to numerically solve it, constructing a linear correction $J\left(\lambda;\beta\right)=\frac{\partial F\left(\lambda;\beta\right)}{\partial\lambda^T}=-\sum_{i=1}^{n}\frac{\tilde{g}_i\left(\beta\right)\tilde{g}_i\left(\beta\right)^T}{\left(1+\lambda^T\tilde{g}_i\left(\beta\right)\right)^2}$ based on the Jacobian matrix at each step. The Newton update can be written as:

$$\lambda^{(k+1)}=\lambda^{(k)}-\alpha_k J\left(\lambda^k;\beta\right)^{-1}F\left(\lambda^k;\beta\right)$$

where $\alpha_k\in\left(0,1\right]$ is the step size. Directly calculating and inverting $J\left(\lambda;\beta\right)$ may be costly; therefore, a Broyden-type update $J\left(\lambda;\beta\right)$ is used to approximate the Jacobian matrix, thus avoiding explicit differentiation and matrix inversion. $\left\|F\left(\lambda^{(k)};\beta\right)\right\|_2\le\varepsilon$ is used as the convergence criterion; if convergence is not achieved within a given number of iterations, it is considered that the empirical likelihood under the current $\beta^{(t)}$ is infeasible. If the original empirical likelihood is infeasible, an adjusted empirical likelihood (AEL) is used to add a spurious observation in the mean moment direction to expand the convex hull. If AEL is still infeasible, $l_{BEL}\left(\beta\right)$ is set to $-\infty$, and this value is always rejected in the MH step, thus ensuring that the chain always remains within the empirically feasible region, and we obtain the empirical likelihood $L\left(\beta^{(t)}\right)=\prod_{i=1}^{n}\frac{1}{n\left(1+\lambda^T\left(\beta^{(t)}\right)\tilde{g}_i\left(\beta^{(t)}\right)\right)}$. Based on the empirical likelihood, given the prior parameters $\mu_0$ and $\Sigma_0$, and combined with the multivariate normal prior density function, the initial logarithmic posterior $l_{BEL}\left(\beta^{(t)}\right)$ is obtained.

(3) A multivariate normal distribution $N\left(\beta^{(t)},\Sigma_{prop}\right)$ is chosen as the proposal distribution. A new parameter $\beta^*$ is drawn from the proposal distribution, and $l_{BEL}\left(\beta^*\right)$ is calculated. We propose to first run a short pilot test on the covariance near the empirical likelihood maximum. We then construct an initial matrix using the sample covariance obtained from the pilot test and adjust the overall step size using a Robbins-Monro update to keep the acceptance rate within a reasonable range during the pilot test. In the formal sampling phase, we fix the covariance and freeze the proposal covariance during the main sampling phase to preserve ergodicity and ensure validity of posterior inference.

For any target density $\pi\left(\beta^{(t)}\right)$ and proposal distribution $q\left(\beta^*\middle|\beta^{(t)}\right)$, the standard acceptance

probability of MH is $\alpha\left(\beta^{(t)},\beta^{*}\right)=\min\left\{1,\frac{\pi\left(\beta^{*}\right)q\left(\beta^{(t)}\middle|\beta^{*}\right)}{\pi\left(\beta^{(t)}\right)q\left(\beta^{*}\middle|\beta^{(t)}\right)}\right\}=\min\left\{1,\exp\left(l_{BEL}\left(\beta^{*}\right)-l_{BEL}\left(\beta^{(t)}\right)\right)\frac{q\left(\beta^{(t)}\middle|\beta^{*}\right)}{q\left(\beta^{*}\middle|\beta^{(t)}\right)}\right\}$. The chosen proposal distribution satisfies the symmetry $q\left(\beta^{*}\middle|\beta^{(t)}\right)=q\left(\beta^{(t)}\middle|\beta^{*}\right)$; therefore, the proposal density ratio is 1. The MH acceptance probability simplifies to $\alpha\left(\beta^{(t)},\beta^{*}\right)=\min\left\{1,\exp\frac{l_{BEL}\left(\beta^{*}\right)}{l_{BEL}\left(\beta^{(t)}\right)}\right\}$, and $\beta^{*}$ is accepted with this probability. Let $\beta_{t+1}=\beta^{*}$, and conversely, $\beta_{t+1}=\beta_{t}$. Then the posterior distribution $\pi_{BEL}\left(\beta^{*}\right)$ is the stationary distribution of the MH chain. Repeat the above steps until a stable state is reached.

We run the Metropolis-Hastings chain for T iterations. The first $b$ iterations, used to eliminate the influence of the initial value on the sampling distribution, are discarded. Only the samples from the last $T-b$ iterations are used to estimate parameter $\beta$, obtaining the posterior sample $\left\{\beta_{k}\right\}_{k=1}^{T-b}$. The BEL estimate of B is expressed as $\beta=\frac{1}{T-k}\sum_{t=k+1}^{T}\beta_{t}$. In each simulation scenario, we perform single-chain diagnostics on each chain and multi-chain (three-chain) diagnostics on five randomly selected chains and several of the worst-performing chains. We record the average acceptance rate, the median effective sample size (ESS), and the maximum value of the Geweke diagnostic statistic to check convergence. The following explains that the posterior distribution is stationary:

Let the target density be $\pi_{BEL}\left(\beta\right)$ and the proposal kernel be $P\left(\beta,d\beta^{*}\right)$. Then the transition kernel of MH can be written as:

$$P\left(\beta,d\beta^{*}\right)=q\left(\beta^{*}\middle|\beta\right)\alpha\left(\beta,\beta^{*}\right)d\beta^{*}+r\left(\beta\right)\delta_{\beta}\left(d\beta^{*}\right)$$

where $r\left(\beta\right)=1-\int q\left(u\middle|\beta\right)\alpha\left(\beta,u\right)du$ is the probability mass of staying in place after rejection, and $\delta_{\beta}\left(d\beta^{*}\right)$ is the Dirac measure of the mass of point $\beta$. For all $\beta\neq\beta^{*}$, define the ratio $R\left(\beta,\beta^{*}\right)=\frac{\pi\left(\beta^{*}\right)q\left(\beta\middle|\beta^{*}\right)}{\pi\left(\beta\right)q\left(\beta^{*}\middle|\beta\right)}$, defined by the MH acceptance rate: $\alpha\left(\beta,\beta^{*}\right)=\min\left\{1,R\left(\beta,\beta^{*}\right)\right\}$, we have:

$$\pi\left(\beta\right)q\left(\beta^{*}\middle|\beta\right)\alpha\left(\beta,\beta^{*}\right)=\min\left\{\pi\left(\beta^{*}\right)q\left(\beta\middle|\beta^{*}\right),\pi\left(\beta\right)q\left(\beta^{*}\middle|\beta\right)\right\}$$

Swapping $\beta$ and $\beta^{*}$, we get:

$$\pi\left(\beta^{*}\right)q\left(\beta\middle|\beta^{*}\right)\alpha\left(\beta^{*},\beta\right)=\min\left\{\pi\left(\beta\right)q\left(\beta^{*}\middle|\beta\right),\pi\left(\beta^{*}\right)q\left(\beta\middle|\beta^{*}\right)\right\}$$

When $\beta \neq \beta^*$, we have:

$$\pi(\beta)q(\beta^*|\beta)\alpha(\beta,\beta^*) = \pi(\beta^*)q(\beta|\beta^*)\alpha(\beta^*,\beta)$$

This is the detailed stationarity condition of the transition kernel under weighted $\pi$. For any measurable set $A$, using the stationarity condition, we have:

$$\int \pi(\beta)P(\beta,A)d\beta = \iint 1_A(\beta^*)\pi(\beta)P(\beta,d\beta^*)d\beta = \int 1_A(\beta^*)\pi(\beta^*)\left[\int P(\beta^*,d\beta)\right]d\beta^*$$

Since $P(\beta^*,\bullet)$ is a probability measure for any fixed $\beta^*$, $\int P(\beta^*,d\beta)=1$. Therefore, $\int \pi(\beta)P(\beta,A)d\beta = \int_A \pi(\beta^*)d\beta^* = \pi(A)$, that is, $\pi_{BEL}(\beta)$ is the stationary distribution of the transition kernel $P(\beta,d\beta^*)$.

In the above algorithm, each M-H update requires calculating the Huber-type moment condition $g_i(\beta)$ across all groups and constructing the empirical likelihood $L(\beta)$ by solving the Lagrange multipliers $\lambda(\beta)$. Therefore, the computational cost per iteration increases linearly with the number of groups. For medium-sized clustering data, directly using the full-sample BEL is feasible; however, for large-scale clustering data, how to further reduce computational costs while maintaining robustness is an important issue.

Some scholars have proposed an adaptive subsampling algorithm based on residuals. This algorithm iteratively selects influential observations based on the interpolation residuals. The core idea is to prioritize retaining observations with larger residuals and perform probability sampling on the remaining observations, then adjust the weights appropriately to reduce approximation errors. Inspired by this, in the case of clustered data in this paper, the following approach can be considered to accelerate BEL: First, calculate the Huber moment vector $g_i(\hat{\beta})$ for each group at an initial estimate $\hat{\beta}$, and use its norm $\|g_i(\hat{\beta})\|$ or robust transformation as a measure of "influence"; then, retain all groups with high influence, and subsample from the remaining groups according to probabilities related to influence, attaching appropriate sampling weights to the sampled groups in the empirical likelihood constraint, thereby obtaining an approximate BEL based on subsamples. This type of residual-driven subsampling scheme is expected to significantly reduce the computational cost of each iteration when the number of groups is large while controlling the loss of statistical efficiency as much as possible. In this paper, we mostly focus on the robustness and coverage performance of full-sample BEL under medium-sized clustered data; system construction and theoretical analysis can be left for future research.

## 3. Simulation

In this chapter, we compare the performance of BEL of linear mixed-effects models in finite samples under different scenarios and with varying variances of random effects through numerical simulations. For each simulation of BEL, the MCMC program executes the algorithm from Chapter 3, performing T=6000 iterations and 2000 annealing cycles.

**Simulation 1:** Comparison of estimation bias and MSE results between BEL and EL, REML, ML, ordinary Bayesian, and generalized moment estimation in different scenarios.

The data is generated by the following model: $Y_i = X_i^T\beta + Z_i^T\alpha_i + \varepsilon_i$, where $\beta = (1, 0.5)^T$. Covariates are generated based on $X_i = (X_{1i}, X_{2i})$, where $X_{1i}, X_{2i} \sim U(-1,1)$. The response variable is generated by this model, and the random effects and error distribution are perturbed under different scenarios to characterize "bad scenarios" such as heavy-tailed distributions, heteroscedasticity, and outliers. The following six scenarios are considered:

1）Baseline scenario: The error and random effects follow a normal distribution, $\varepsilon_i \sim N(0,1)$, $\alpha_i \sim N(0,1)$;

2）Thick tail error： $\varepsilon_i \sim 0.85N(0,1) + 0.15N(0,6^2)$ and $\alpha_i \sim N(0,1)$;

3）Heavy-tailed random effects: The random effects follow a t-distribution with 3 degrees of freedom, $T_i \sim t(3)$, $\sigma_\alpha$ is set to 1, and let $\alpha_i = \sigma_\alpha \frac{T_i}{\sqrt{3}}$, thus keeping $Var(\alpha_i) = \sigma_\alpha^2$ unchanged, but the distribution of the random intercept has the heavy-tailed characteristics of the t-distribution, and the error term $\varepsilon_i \sim N(0,1)$;

4）High-leverage covariates and heteroscedasticity: $\alpha_i \sim N(0,1)$, where covariates are randomly divided into a 20% "high-leverage group" and an 80% "normal group," with the high-leverage group assigned a larger variance, $\varepsilon_i \sim \begin{cases} N(0,1), & \text{Normal Group} \\ N(0,3^2), & \text{High leverage group} \end{cases}$;

5）Outlier points exist: $\alpha_i \sim N(0,1)$, where the error is based on a normal distribution, with an additional $N(0,8^2)$ noise added to 8% of the observations;

6）Outlier group: $\varepsilon_i \sim N(0,1)$ and $\alpha_i \sim N(0,1)$, where the true random intercept of 4% of the groups is set to $\alpha_i + 2$, forming a clear "outlier group".

For the Bayes method, it refers to the ordinary Bayesian inference of a fully parameterized linear mixed-effects model. $\alpha_i \sim N(0, \sigma_\alpha^2)$ and $\varepsilon_i^2 \sim N(0, \sigma_\varepsilon^2)$ are used, and relatively relaxed conjugate priors $\beta \sim N(\mu_0, V_0)$, $\sigma_\varepsilon^2 \sim Inv - Gamma(a_0, b_0)$, and $\sigma_\alpha^2 \sim Inv - Gamma(c_0, d_0)$ are taken, where $\mu_0 = (0,0)^T$ and $V_0 = diag(3^2, 3^2)$ are in the range of $a_0 = b_0 = c_0 = d_0 = 2$. The posterior distribution is obtained by Gibbs sampling. This method is a correctly configured model in the baseline scenario, but an incorrectly configured model in other scenarios.

Under each setting, the independent simulation is repeated 100 times, with each simulation generating 25 groups and 6 observations per group. The result of each simulation is denoted as $\beta_t$. The mean of the 100 simulation results is used as the final estimate. The estimation results are compared by calculating the absolute value of the estimation bias, MSE, average length of the 95% confidence interval, and coverage, where:

$$\hat{\beta} = \frac{1}{100}\sum_{t=1}^{100}\hat{\beta}_t,\ Bias(\hat{\beta}) = |\bar{\beta} - \beta|,\ MSE(\hat{\beta}) = \frac{1}{nsim}\sum_{t=1}^{nsim}(\hat{\beta}_t - \beta)^2$$

(1) Numerical stability, algorithm convergence and weight perturbation sensitivity analysis

Before discussing the estimation performance of each method, such as bias, mean square error, and interval coverage, it is necessary to examine the reliability of the proposed BEL method at the numerical computation level. The diagnosis is carried out from three levels: Feasibility and stability of the empirical likelihood dual problem, convergence of the MCMC algorithm, and evaluation of the sensitivity of the block-diagonal weight setting to the estimation results. The results are given with 50 groups as representative samples.

As shown in Table 1, the feasible proportion is 1 for all 6 scenarios, indicating that the original empirical likelihood dual equation can be solved within the feasible region in most repeated samples. The overall usage proportion of AEL is relatively low, only slightly higher in the "high leverage covariates and heteroscedasticity" scenario, reflecting that the convex hull boundary is more easily reached in this scenario, requiring limited convex hull expansion through AEL. The maximum value of the KKT residual is close to 0 in all scenarios, indicating that the first-order conditions are satisfied. The 5th percentile of the minimum dual denominator is significantly greater than 0, and the 5th percentile of the minimum eigenvalue of the dual Hessian is also at a medium to high level, indicating that the dual problem is far from numerically singular and has good condition number and stability overall.

After confirming the numerical stability of the empirical likelihood, we further evaluate the convergence of BEL posterior sampling from the perspective of MCMC, and the relevant diagnostic results are shown in Table 2.

**Table 1.** Stability index of empirical likelihood values in various scenarios when the number of groups is 50.

| Baseline | Feasible proportion | AEL usage proportion | KKT residual maximum value | minimum dual denominator 5th percentile | dual Hessian minimum eigenvalue 5th percentile |
|---|---|---|---|---|---|
| Thick tail error | 1 | 0.03 | 0 | 0.3219 | 7.1185 |
| Heavy-tailed random effects | 1 | 0.04 | 0 | 0.3497 | 6.7063 |
| High-leverage covariates and heteroscedasticity | 1 | 0.06 | 0 | 0.4061 | 7.7080 |
| Outliers exist | 1 | 0.21 | 0 | 0.1199 | 18.4066 |
| Group outliers | 1 | 0.05 | 0 | 0.2430 | 7.4121 |
| Baseline | 1 | 0.02 | 0 | 0.3831 | 10.4648 |

From the perspective of single-chain diagnostics, the average acceptance rate for each scenario is generally within the reasonable range of random walk. Except for the scenario of "high leverage covariates and heteroscedasticity" where the median ESS of $\beta_0$ is slightly lower, the median ESS of $\beta_0$ and $\beta_1$ in other scenarios are all above 300, indicating that the effective sample size of single chains is sufficient and the overall chain mixture is good. After performing Gelman-Rubin diagnostics on three independent chains for random sampling and extreme datasets, the upper bound of $\hat{R}$ for

$\beta_0$ and $\beta_1$ does not exceed 1.02, and the ESS of multi-chains is usually in the hundreds or thousands, indicating that the posterior sampling of BEL exhibits stable convergence characteristics under different error and random effect conditions.

**Table 2.** Algorithm convergence index when the number of groups is 50.

| Baseline | Single-chain diagnostics | | | Multichain diagnostics | | | |
|---|---|---|---|---|---|---|---|
| | Average acceptance rate | ESS median | | upper bound of $\hat{R}$ | | ESS | |
| | | $\beta_0$ | $\beta_1$ | $\beta_0$ | $\beta_1$ | $\beta_0$ | $\beta_1$ |
| Thick tail error | 0.3488 | 379.8 | 548.4 | 1.018 | 1.007 | 856.9 | 1489.9 |
| Heavy-tailed random effects | 0.4244 | 420.5 | 443.7 | 1.000 | 1.000 | 1081.7 | 1486.6 |
| High-leverage covariates and heteroscedasticity | 0.3168 | 430.3 | 513.7 | 1.001 | 1.001 | 1253.6 | 1586.5 |
| Outliers exist | 0.2541 | 141.5 | 360.7 | 1.001 | 1.002 | 443.8 | 1103.3 |
| Group outliers | 0.4030 | 426.5 | 474.5 | 1.001 | 1.002 | 1319.3 | 1224.0 |
| Baseline | 0.3539 | 360.9 | 591.9 | 1.004 | 1.001 | 744.2 | 1742.0 |

Finally, the impact of the block-diagonal weight setting on EL point estimation is examined to test the robustness of the numerical results to the weight selection. The relevant sensitivity analysis results are shown in Table 3.

**Table 3.** Sensitivity analysis of weight perturbation when the number of groups is 50.

| Baseline | $\beta_0$ | $\beta_1$ |
|---|---|---|
| | Mean absolute deviation | Mean absolute deviation |
| Thick tail error | 0.0307 | 0.0394 |
| Heavy-tailed random effects | 0.0405 | 0.0279 |
| High-leverage covariates and heteroscedasticity | 0.1025 | 0.1639 |
| Outliers exist | 0.2467 | 0.3419 |
| Group outliers | 0.0750 | 0.0929 |
| Baseline | 0.0159 | 0.0358 |

As can be seen, the mean absolute deviations of $\beta_0$ and $\beta_1$ are relatively small in the baseline scenario, heavy-tailed error, and group outlier scenarios. However, the deviations increase in the heavy-tailed random effects and high-leverage covariates and heteroscedasticity scenarios. This indicates that after using robust scaling for block-diagonal scaling, EL and BEL inferences are generally robust to the weight settings within a reasonable range, and are only slightly sensitive under extreme high leverage and heteroscedasticity conditions.

(2) Parameter estimation of each method

The numerical stability, convergence diagnosis, and weight sensitivity analyses described above demonstrate that the proposed BEL framework exhibits good numerical properties across the scenarios considered. In the following section, we compare the differences in estimation performance among different methods in terms of bias, mean squared error, and interval coverage. Tables 4 and 5 show the

parameter estimation results for each method under different scenarios when the number of groups is 50 and the number of observations per group is 6.

**Table 4.** Estimation results of methods for $\beta_0$ and $\beta_1$ under different scenarios when the number of groups is 50.

| Method | Baseline | | | | Thick tail error | | | | Heavy-tailed random effects | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\beta_0$ | | $\beta_1$ | | $\beta_0$ | | $\beta_1$ | | $\beta_0$ | | $\beta_1$ | |
| | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE |
| BEL | 0.0105 | 0.0222 | 0.0108 | 0.0123 | 0.0169 | 0.0284 | 0.0124 | 0.0278 | 0.0035 | 0.0161 | 0.0025 | 0.0109 |
| EL | 0.0266 | 0.0493 | 0.0387 | 0.0769 | 0.0032 | 0.0619 | 0.0097 | 0.0331 | 0.0105 | 0.0461 | 0.0128 | 0.0127 |
| REML | 0.0124 | 0.0215 | 0.0132 | 0.0124 | 0.0201 | 0.0451 | 0.0220 | 0.0638 | 0.0213 | 0.0291 | 0.0081 | 0.0099 |
| ML | 0.0124 | 0.0215 | 0.0132 | 0.0124 | 0.0201 | 0.0451 | 0.0224 | 0.0639 | 0.0213 | 0.0291 | 0.0082 | 0.0099 |
| Bayes | 0.0160 | 0.0217 | 0.0123 | 0.0123 | 0.0253 | 0.0454 | 0.0177 | 0.0622 | 0.0173 | 0.0282 | 0.0090 | 0.0099 |
| GMM | 0.0130 | 0.0218 | 0.0119 | 0.0127 | 0.0118 | 0.0276 | 0.0022 | 0.0307 | 0.0070 | 0.0154 | 0.0000 | 0.0100 |

**Table 5.** Estimation results of methods for $\beta_0$ and $\beta_1$ under different scenarios when the number of groups is 50.

| Method | High-leverage covariates and heteroscedasticity | | | | Outliers exist | | | | Group outliers | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\beta_0$ | | $\beta_1$ | | $\beta_0$ | | $\beta_1$ | | $\beta_0$ | | $\beta_1$ | |
| | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE |
| BEL | 0.0090 | 0.0344 | 0.0010 | 0.0088 | 0.0195 | 0.0340 | 0.0200 | 0.0184 | 0.0766 | 0.0341 | 0.0031 | 0.0099 |
| EL | 0.0013 | 0.1048 | 0.0149 | 0.1052 | 0.0339 | 0.0745 | 0.0052 | 0.0984 | 0.0773 | 0.0351 | 0.0349 | 0.0638 |
| REML | 0.0156 | 0.0388 | 0.0019 | 0.0093 | 0.0215 | 0.0551 | 0.0110 | 0.0625 | 0.0943 | 0.0372 | 0.0067 | 0.0101 |
| ML | 0.0155 | 0.0388 | 0.0019 | 0.0093 | 0.0215 | 0.0551 | 0.0110 | 0.0627 | 0.0943 | 0.0372 | 0.0067 | 0.0101 |
| Bayes | 0.0121 | 0.0387 | 0.0020 | 0.0092 | 0.0273 | 0.0553 | 0.0161 | 0.0615 | 0.0896 | 0.0361 | 0.0059 | 0.0101 |
| GMM | 0.0134 | 0.0523 | 0.0040 | 0.0093 | 0.0133 | 0.0331 | 0.0075 | 0.0714 | 0.0783 | 0.0342 | 0.0053 | 0.0102 |

As shown in Tables 4 and 5, under the baseline scenario, the parameter methods based on normal distribution, such as REML, ML, and Bayes, have high efficiency, with MSE slightly lower than BEL, but the difference is not significant. For parameter $\beta_0$, in bad scenarios such as heavy-tailed error, heavy-tailed random effects, high-leverage covariates with heteroscedasticity, presence of point outliers, and group outliers, the MSE of traditional methods REML, ML, and Bayes increases significantly. In contrast, the increase in MSE of BEL in these scenarios is relatively moderate. For parameter $\beta_1$, the MSE of traditional methods increases significantly under heavy-tailed error and outlier scenarios, while the MSE of BEL is generally at a lower level among the methods.

As can be seen more intuitively from Figures 1 and 2, the height of each column in BEL changes relatively smoothly across scenarios. From the baseline scenario to the scenarios with heavy tails, high leverage heteroscedasticity, point outliers, and group outliers, the MSE increases, but overall, it remains within a relatively narrow range. In contrast, the MSE of other methods increases significantly under the conditions of heavy tails and outliers, especially the MSE of $\beta_1$, indicating that these methods lack suppression of the impact of extreme residuals and outliers.

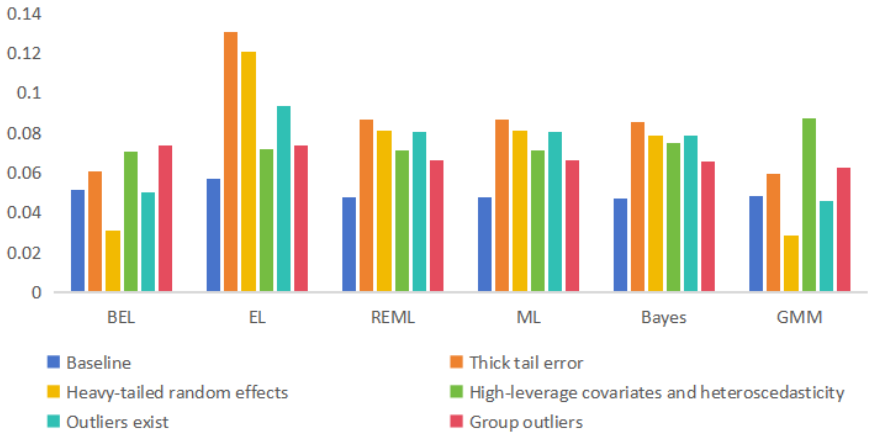


**Figure 1.** Bar chart of MSE estimates for $\beta_0$ under 6 scenarios with 50 groups.

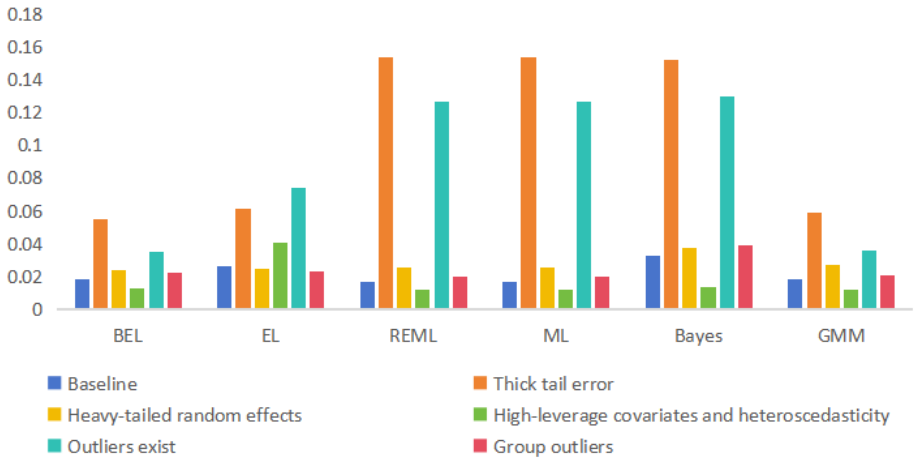


**Figure 2.** Bar chart of MSE estimates for $\beta_1$ under 6 scenarios with 50 groups.

These results show that, under ideal settings of normal random effects and normal errors, the estimation accuracy of BEL is comparable to that of REML, ML, and Bayes. The three-dimensional moment conditions constructed based on Huber transform and annihilation matrix do not compromise the consistency of the estimation and do not result in significant efficiency loss. However, in "bad scenarios", such as heavy-tailed errors, heavy-tailed random effects, high leverage covariates and heteroscedasticity, the presence of outliers, and group outliers, BEL's MSE is generally lower than that of traditional empirical likelihood and generally outperforms several classic estimation methods based on the normal assumption, demonstrating robustness. This indicates that the chosen moment conditions can effectively mitigate the impact of heavy-tailed errors and misconfigured random effects on the estimation.

To examine the impact of the number of groups on estimation performance, we change the number of groups to 25 and 100 in the same six scenarios. The parameter estimation results are shown in Tables 6 and 7 and Tables 8 and 9, respectively.

The numerical results from different numbers of groups in various scenarios show that when the number of observations remains constant and only the number of groups is increased, in the baseline scenario, as the number of groups increases from 25 to 100, the bias and MSE of each method decrease to varying degrees in most scenarios, reflecting the convergence trend of the estimator as the effective

sample size increases. The BEL is generally at a low level in different numbers of groups and different scenarios. In the high leverage covariates and heteroscedasticity scenarios, the MSE of the traditional EL is abnormally amplified when the number of groups is 100. In contrast, the BEL can alleviate the numerical instability problem.

**Table 6.** Estimation results of methods for $\beta_0$ and $\beta_1$ under different scenarios when the number of groups is 25.

| Method | Baseline | | | | Thick tail error | | | | Heavy-tailed random effects | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\beta_0$ | | $\beta_1$ | | $\beta_0$ | | $\beta_1$ | | $\beta_0$ | | $\beta_1$ | |
| | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE |
| BEL | 0.0279 | 0.0492 | 0.0181 | 0.0284 | 0.0258 | 0.0428 | 0.0013 | 0.0555 | 0.0256 | 0.0307 | 0.0012 | 0.0309 |
| EL | 0.0274 | 0.0514 | 0.0227 | 0.0302 | 0.0087 | 0.0437 | 0.0280 | 0.1865 | 0.0200 | 0.0608 | 0.0031 | 0.0310 |
| REML | 0.0277 | 0.0456 | 0.0117 | 0.0275 | 0.0047 | 0.0612 | 0.0160 | 0.0946 | 0.0338 | 0.0418 | 0.0070 | 0.0270 |
| ML | 0.0277 | 0.0456 | 0.0118 | 0.0276 | 0.0047 | 0.0612 | 0.0151 | 0.0948 | 0.0338 | 0.0418 | 0.0071 | 0.0270 |
| Bayes | 0.0218 | 0.0452 | 0.0127 | 0.0273 | 0.0051 | 0.0604 | 0.0071 | 0.0902 | 0.0303 | 0.0400 | 0.0053 | 0.0268 |
| GMM | 0.0289 | 0.0466 | 0.0103 | 0.0263 | 0.0212 | 0.0415 | 0.0059 | 0.0510 | 0.0265 | 0.0261 | 0.0036 | 0.0299 |

**Table 7.** Estimation results of methods for $\beta_0$ and $\beta_1$ under different scenarios when the number of groups is 25.

| Method | High-leverage covariates and heteroscedasticity | | | | Outliers exist | | | | Group outliers | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\beta_0$ | | $\beta_1$ | | $\beta_0$ | | $\beta_1$ | | $\beta_0$ | | $\beta_1$ | |
| | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE |
| BEL | 0.0329 | 0.0968 | 0.0019 | 0.0122 | 0.0185 | 0.0658 | 0.0008 | 0.0508 | 0.0788 | 0.0588 | 0.0098 | 0.0193 |
| EL | 0.0394 | 0.0893 | 0.0134 | 0.0351 | 0.0072 | 0.0691 | 0.0010 | 0.0479 | 0.0692 | 0.0600 | 0.0081 | 0.0200 |
| REML | 0.0386 | 0.0899 | 0.0014 | 0.0121 | 0.0107 | 0.0968 | 0.0557 | 0.1361 | 0.0856 | 0.0556 | 0.0115 | 0.0188 |
| ML | 0.0388 | 0.0897 | 0.0016 | 0.0119 | 0.0107 | 0.0968 | 0.0554 | 0.1361 | 0.0856 | 0.0556 | 0.0116 | 0.0188 |
| Bayes | 0.0456 | 0.0887 | 0.0018 | 0.0119 | 0.0201 | 0.0947 | 0.0475 | 0.1303 | 0.0796 | 0.0533 | 0.0125 | 0.0186 |
| GMM | 0.0225 | 0.1411 | 0.0040 | 0.0161 | 0.0079 | 0.0597 | 0.0007 | 0.0565 | 0.0693 | 0.0543 | 0.0087 | 0.0191 |

**Table 8.** Estimation results of methods for $\beta_0$ and $\beta_1$ under different scenarios when the number of groups is 100.

| Method | Baseline | | | | Thick tail error | | | | Heavy-tailed random effects | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\beta_0$ | | $\beta_1$ | | $\beta_0$ | | $\beta_1$ | | $\beta_0$ | | $\beta_1$ | |
| | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE |
| BEL | 0.0062 | 0.0120 | 0.0071 | 0.0061 | 0.0119 | 0.0128 | 0.0015 | 0.0162 | 0.0040 | 0.0080 | 0.0124 | 0.0053 |
| EL | 0.0105 | 0.0315 | 0.0089 | 0.0067 | 0.0121 | 0.0139 | 0.0002 | 0.0253 | 0.0257 | 0.0626 | 0.0403 | 0.0292 |
| REML | 0.0088 | 0.0111 | 0.0067 | 0.0057 | 0.0078 | 0.0195 | 0.0038 | 0.0423 | 0.0058 | 0.0127 | 0.0130 | 0.0053 |
| ML | 0.0088 | 0.0111 | 0.0067 | 0.0057 | 0.0078 | 0.0195 | 0.0037 | 0.0423 | 0.0058 | 0.0127 | 0.0130 | 0.0053 |
| Bayes | 0.0104 | 0.0113 | 0.0062 | 0.0057 | 0.0106 | 0.0195 | 0.0057 | 0.0418 | 0.0084 | 0.0127 | 0.0134 | 0.0053 |
| GMM | 0.0070 | 0.0114 | 0.0078 | 0.0060 | 0.0093 | 0.0126 | 0.0077 | 0.0168 | 0.0011 | 0.0079 | 0.0106 | 0.0053 |

**Table 9.** Estimation results of methods for $\beta_0$ and $\beta_1$ under different scenarios when the number of groups is 100.

| Method | High-leverage covariates and heteroscedasticity | | | | Outliers exist | | | | Group outliers | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\beta_0$ | | $\beta_1$ | | $\beta_0$ | | $\beta_1$ | | $\beta_0$ | | $\beta_1$ | |
| | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE |
| BEL | 0.0028 | 0.0173 | 0.0023 | 0.0025 | 0.0114 | 0.0128 | 0.0123 | 0.0085 | 0.0591 | 0.0178 | 0.0018 | 0.0048 |
| EL | 0.0562 | 0.6952 | 0.1337 | 0.7824 | 0.0134 | 0.0150 | 0.0338 | 0.0522 | 0.0412 | 0.0464 | 0.0120 | 0.0302 |
| REML | 0.0055 | 0.0171 | 0.0011 | 0.0028 | 0.0079 | 0.0213 | 0.0272 | 0.0348 | 0.0722 | 0.0192 | 0.0029 | 0.0047 |
| ML | 0.0055 | 0.0171 | 0.0011 | 0.0028 | 0.0079 | 0.0213 | 0.0272 | 0.0348 | 0.0722 | 0.0192 | 0.0029 | 0.0047 |
| Bayes | 0.0035 | 0.0168 | 0.0012 | 0.0028 | 0.0102 | 0.0210 | 0.0259 | 0.0345 | 0.0698 | 0.0190 | 0.0031 | 0.0047 |
| GMM | 0.0022 | 0.0289 | 0.0030 | 0.0028 | 0.0095 | 0.0127 | 0.0101 | 0.0088 | 0.0603 | 0.0181 | 0.0032 | 0.0050 |

Figures 3 and 4 present the numerical results as bar charts, which more intuitively confirm this. Under different scenarios, as the number of groups increases, the overall bar height decreases for most methods, the MSE of parameters $\beta_0$ and $\beta_1$ decreases significantly, and the BEL remains at a low level overall.

From the perspective of sample size, the Huber transform weakens the influence of fat tails and outliers on the moment condition, and the annihilation matrix reduces the dependence on the setting of the random effects distribution, so that the empirical mean of the moment condition converges to 0 as the number of groups increases. Thus, it achieves the effect of slightly losing efficiency in the correct scenario and gradually amplifying the robustness advantage in the complex scenario as the sample size increases under limited sample size.

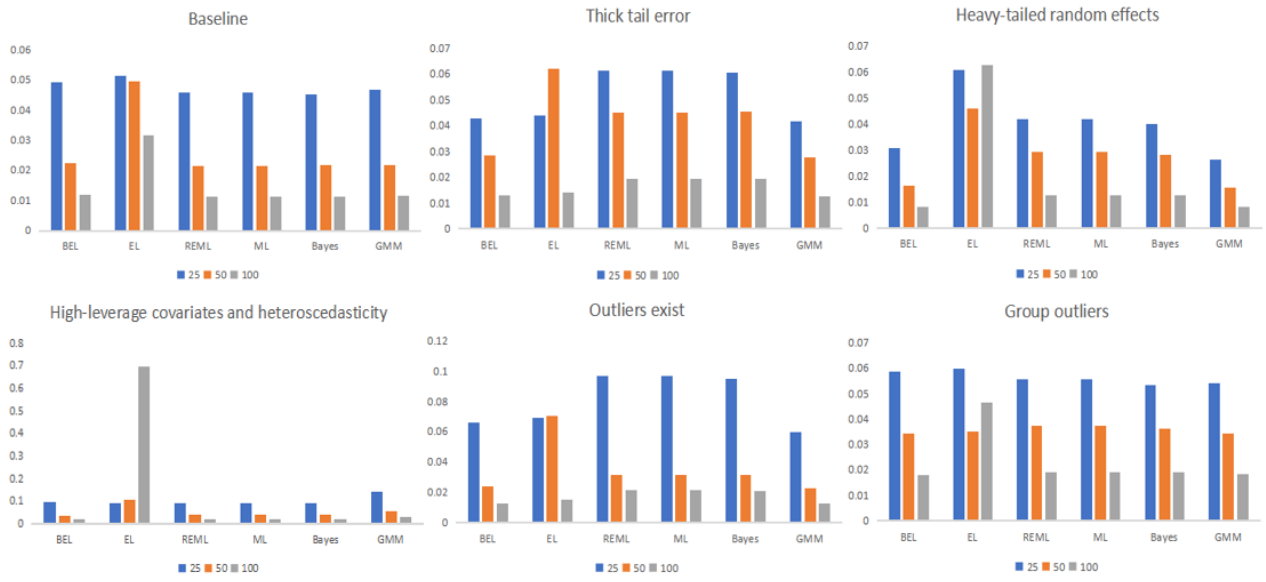


**Figure 3.** MSE histograms for $\beta_0$ under different number of groups.

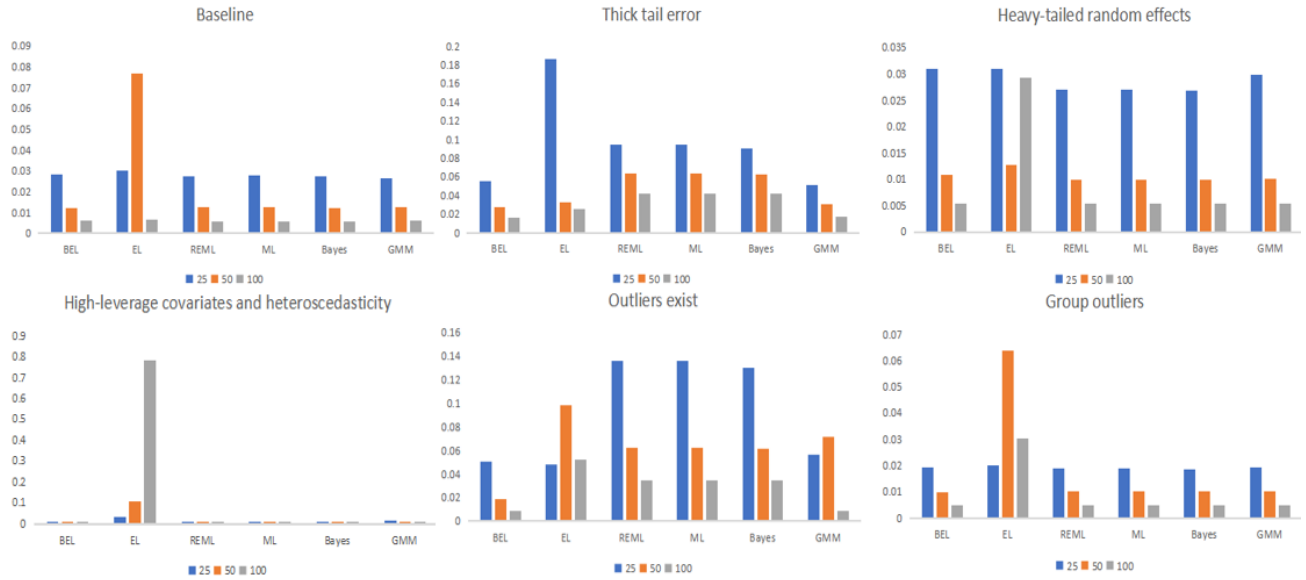


**Figure 4.** MSE histograms for $\beta_1$ under different number of groups.

As shown in Tables 10 and 11, the overall coverage of each scenario is mostly concentrated between 0.90 and 0.97, with the coverage in most scenarios approaching the nominal level of 0.95. When the number of groups is 25, the coverage of scenarios with high leverage covariates and heteroscedasticity, outliers, and stratified outliers is low. However, when the number of groups increases to 50 and 100, the coverage increases. The interval length decreases significantly with the increase of the number of groups. Combined with the results of bias and MSE mentioned above, BEL can maintain good frequency accuracy in complex environments such as heavy-tailed error, heavy-tailed random effects, and outliers, thus demonstrating its robustness and effectiveness.

**Simulation 2:** Finite sample performance of BEL and comparison methods under different random effects variances.

**Table 10.** $\beta_0$'s 95% posterior interval coverage and 95% posterior interval length.

| Scenario | 25 | | 50 | | 100 | |
|---|---|---|---|---|---|---|
| | Coverage | Interval Length | Coverage | Interval Length | Coverage | Interval Length |
| Baseline | 0.94 | 0.8409 | 0.93 | 0.5980 | 0.93 | 0.4327 |
| Thick tail error | 0.94 | 0.8557 | 0.93 | 0.6335 | 0.94 | 0.4499 |
| Heavy-tailed random effects | 0.92 | 0.6861 | 0.93 | 0.5053 | 0.93 | 0.3573 |
| High-leverage covariates and heteroscedasticity | 0.83 | 0.9489 | 0.95 | 0.7025 | 0.91 | 0.4931 |
| Outliers exist | 0.89 | 0.8605 | 0.90 | 0.6118 | 0.97 | 0.4338 |
| Group outliers | 0.88 | 0.8376 | 0.92 | 0.6369 | 0.89 | 0.4469 |

**Table 11.** $\beta_1$'s 95% posterior interval coverage and 95% posterior interval length.

| Scenario | 25 | | 50 | | 100 | |
|---|---|---|---|---|---|---|
| | Coverage | Interval Length | Coverage | Interval Length | Coverage | Interval Length |
| Baseline | 0.93 | 0.6063 | 0.95 | 0.4301 | 0.95 | 0.3069 |
| Thick tail error | 0.92 | 0.8420 | 0.95 | 0.6261 | 0.92 | 0.4400 |
| Heavy-tailed random effects | 0.90 | 0.5807 | 0.94 | 0.4265 | 0.96 | 0.3066 |
| High-leverage covariates and heteroscedasticity | 0.83 | 0.3651 | 0.88 | 0.2799 | 0.94 | 0.1991 |
| Outliers exist | 0.92 | 0.7988 | 0.95 | 0.5605 | 0.96 | 0.3958 |
| Group outliers | 0.93 | 0.5863 | 0.95 | 0.4212 | 0.95 | 0.3045 |

The bad scenarios in Simulation 1 represent deliberate deviations from the normality assumption. This simulation, however, assumes that the random intercept and error follow a normal distribution, but changes only the variance of the random effects to examine the finite sample performance of each method under the correct model and its sensitivity to the variance of the random effects. Consider the random intercept model:

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + \alpha_i + \varepsilon_{ij}$$

where $\beta_0 = 1$, $\beta_1 = 1$, $\alpha_i \sim N\left(0, \tau^2\right)$, and $\varepsilon_{ij} \sim N\left(0,1\right)$, and the covariate $x_{ij} \sim N\left(0,1\right)$ is independent and identically distributed and independent of the random effect $\alpha_i$. Considering five variance levels of $\tau^2$ (0.1, 0.5, 1, 2, and 4), 50 groups with 6 observations per group are generated under each setting, and the simulation is repeated 100 times. Estimates for BEL, EL, REML, ML, Ordinary Bayesian, and GMM are provided for each simulated sample, and the bias and MSE are calculated. Results are shown in Tables 7 and 8.

As can be seen from Tables 7 and 8, the estimation biases of all 6 methods are close to zero at all variance levels. Therefore, it can be considered that there is essentially no systematic bias at this simulation scale.

**Table 12.** Estimation results of $\beta_0$ under different variances of random effects.

| Method | 0.1 | | 0.5 | | 1 | | 2 | | 4 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE |
| BEL | 0.0030 | 0.0048 | 0.0100 | 0.0122 | 0.0164 | 0.0226 | 0.0285 | 0.0426 | 0.0412 | 0.0853 |
| EL | 0.0029 | 0.0049 | 0.0097 | 0.0125 | 0.0167 | 0.0227 | 0.0270 | 0.0434 | 0.0413 | 0.0847 |
| REML | 0.0035 | 0.0046 | 0.0110 | 0.0113 | 0.0167 | 0.0205 | 0.0247 | 0.0396 | 0.0360 | 0.0787 |
| ML | 0.0035 | 0.0046 | 0.0110 | 0.0113 | 0.0167 | 0.0205 | 0.0247 | 0.0396 | 0.0360 | 0.0787 |
| Bayes | 0.0034 | 0.0047 | 0.0116 | 0.0117 | 0.0160 | 0.0215 | 0.0230 | 0.0410 | 0.0382 | 0.0819 |
| GMM | 0.0029 | 0.0046 | 0.0111 | 0.0116 | 0.0184 | 0.0211 | 0.0293 | 0.0409 | 0.0450 | 0.0820 |

**Table 13.** Estimation results of $\beta_1$ under different variances of random effects.

| Method | 0.1 | | 0.5 | | 1 | | 2 | | 4 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE | Bias | MSE |
| BEL | 0.0020 | 0.0033 | 0.0020 | 0.0034 | 0.0017 | 0.0035 | 0.0013 | 0.0036 | 0.0010 | 0.0038 |
| EL | 0.0024 | 0.0034 | 0.0025 | 0.0035 | 0.0022 | 0.0036 | 0.0017 | 0.0037 | 0.0012 | 0.0038 |
| REML | 0.0002 | 0.0034 | 0.0001 | 0.0035 | 0.0004 | 0.0036 | 0.0007 | 0.0037 | 0.0009 | 0.0037 |
| ML | 0.0003 | 0.0034 | 0.0000 | 0.0035 | 0.0003 | 0.0036 | 0.0006 | 0.0036 | 0.0009 | 0.0037 |
| Bayes | 0.0006 | 0.0033 | 0.0009 | 0.0035 | 0.0010 | 0.0036 | 0.0012 | 0.0037 | 0.0011 | 0.0037 |
| GMM | 0.0025 | 0.0032 | 0.0017 | 0.0035 | 0.0010 | 0.0038 | 0.0002 | 0.0046 | 0.0008 | 0.0063 |

From the perspective of MSE, the impact of variance $\tau^2$ on the intercept and slope varies significantly. In Table 7, as the variance of random effects increases, the MSE of each method $\beta_1$ pair increases significantly, with EL consistently showing the highest MSE. In Table 8, the MSE of $\beta_1$ changes relatively smoothly with $\tau^2$, and the differences among the methods are also small, indicating that the variance of the random intercept mainly affects the intercept term while having a limited impact on the slope estimation.

Under a correct model where the random intercept and error are normally distributed, BEL, EL, and GMM exhibit comparable finite-sample efficiency to REML, ML, and parametric Bayesian methods in slope estimation, with only a slight efficiency loss in the intercept parameter. To more intuitively compare the performance of each method under different variances of random effects, a bar chart of the MSE indices in Tables 7 and 8 is presented in Figure 5.

As can be seen from Figure 5 and Figure 6, the MSE of $\beta_0$ estimated by each method shows a significant upward trend as $\tau^2$ increases, while the MSE of $\beta_1$ remains stable within a small range at all variance levels, without any obvious monotonic changes.

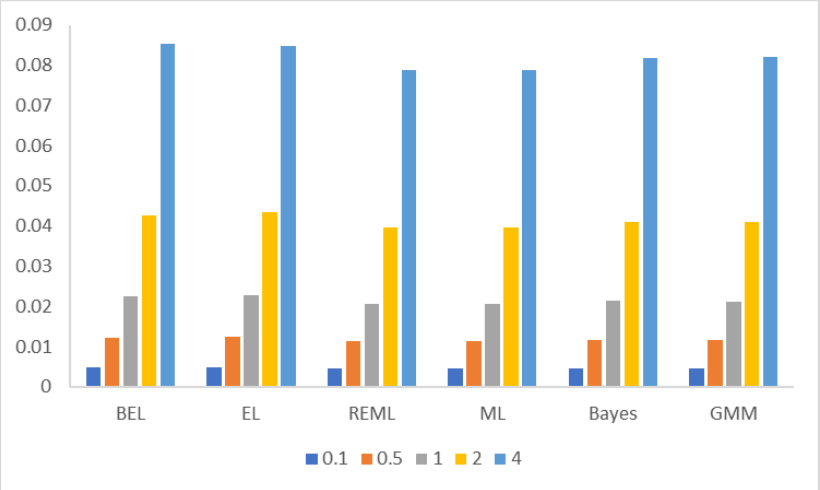


**Figure 5.** MSE histograms for $\beta_0$ under different variances of random effects.

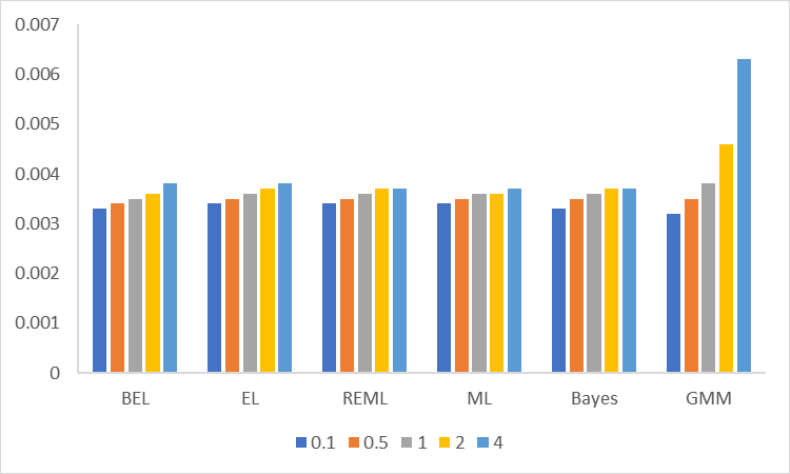


**Figure 6.** MSE histograms for $\beta_1$ under different variances of random effects.

This phenomenon is consistent with the theoretical properties of the stochastic intercept model: The variance of the stochastic intercept mainly affects the overall level of each group, thus significantly increasing the uncertainty of the intercept estimate, while the slope $\beta_1$ is mainly provided by the changes in the covariates within the group. When the covariates are centered and independent of the stochastic intercept, its estimated variance is less dependent on $\tau^2$, and therefore exhibits relative stability.

Based on the results of Simulation 1 under the conditions of heavy-tailed error, misconfiguration of random effects distribution, and outlier, it can be concluded that when the model is correct, BEL trades only robustness at a small efficiency cost, while when the model deviates from the normality assumption, it can significantly improve the estimation bias and mean square error, thus verifying the theoretical advantages of constructing moment conditions based on the Huber transform function and introducing the BEL method in this paper.

## 4. Empirical analysis

As China's population ages, the relationship between healthcare expenditure and economic development levels across regions has become a core concern in public finance and health policy. A reasonable characterization of the linkages between per capita healthcare expenditure, per capita GDP, and the degree of aging can help identify differences in healthcare input across provinces and their potential influencing factors. However, significant differences in development levels between regions, and the existence of extreme values or structural shifts in a few provinces, mean that traditional linear mixed-effects models based on the normality assumption may face interference from heavy tails and outliers in parameter estimation and interval inference. Using the determined BEL framework of the mixed-effects model constructed above, we examine the applicability and robustness of this method in real-world data, taking panel data from 31 provinces in China as an example.

In this chapter, we use panel data from 31 provinces across China from 2014 to 2022, as published in the China Statistical Yearbook. Per capita GDP and aging rate are used as covariates, and per capita healthcare expenditure is used as the response variable. Provinces are treated as a random effects layer to characterize the differences in healthcare infrastructure among regions. Because economic level and aging rate indicators differ significantly in scale and fluctuation range, direct modeling can easily lead to significant differences in the numerical scale of the estimators and unstable algorithm convergence.

Therefore, we first standardize the covariates and then construct a linear mixed-effects model based on this standardization before conducting subsequent analysis.

As shown in Figure 7, there is a significant positive correlation between standardized per capita healthcare expenditure and both per capita GDP and aging population level. The correlation between the two covariates is moderate, indicating that both have a significant positive impact on healthcare expenditure and there is no severe collinearity. Moreover, the diagonal histogram shows that per capita healthcare expenditure and per capita GDP are slightly right-skewed, with higher dispersion in high-income regions. To reduce right-tail fatness and heteroscedasticity, and to mitigate the leverage effect of high-income provinces, we first take the natural logarithm of per capita healthcare expenditure and per capita GDP before standardization in subsequent modeling while standardizing the aging population level on the original proportional scale. Figure 8 presents the correlation coefficients among the three variables after logarithmic transformation for per capita healthcare expenditure and GDP.

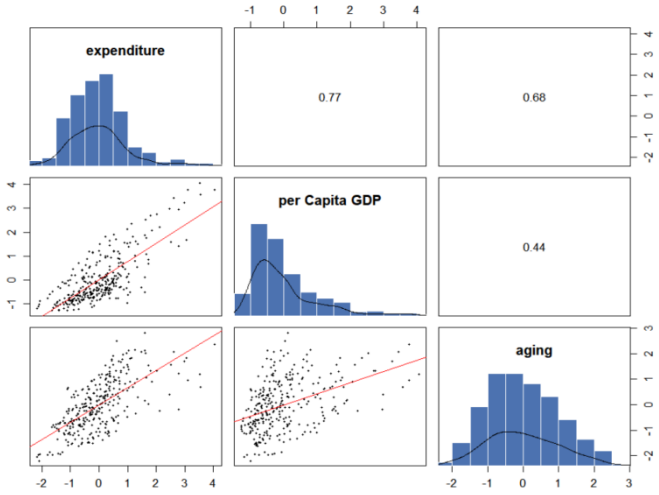


**Figure 7.** Correlation analysis of standardized variables.

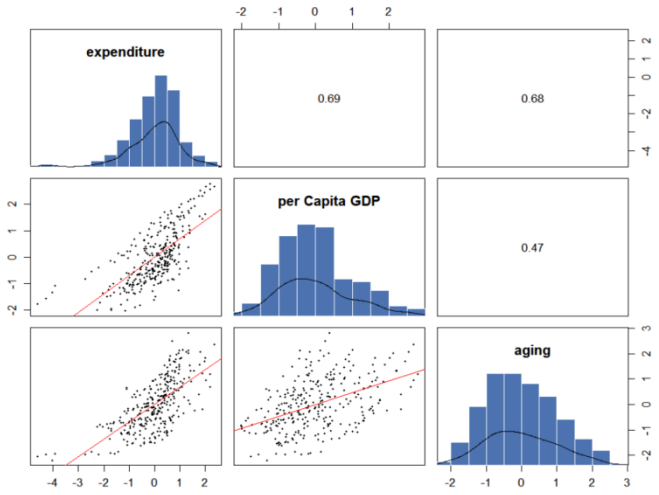


**Figure 8.** Correlation analysis of log-standardized variables.

After logarithmic and standardization processing, the variable distribution is closer to symmetry and scale comparability, while per capita healthcare expenditure retains some heavy-tailed characteristics. Based on this, a linear mixed-effects model is first fitted to the log-standardized data, and diagnostic analysis is performed on the residuals.

As can be seen from the residual distribution plot in Figure 9 and the normal QQ plot in Figure 10, the residuals are generally unimodal and approximately symmetrically distributed. This indicates that after linearly adjusting for per capita GDP and aging, the systematic structure of per capita healthcare expenditure has been well characterized. A small number of extreme observations can be seen at both ends, and the right tail is slightly longer. The residual distribution has a certain thick tail characteristic compared to the normal distribution.

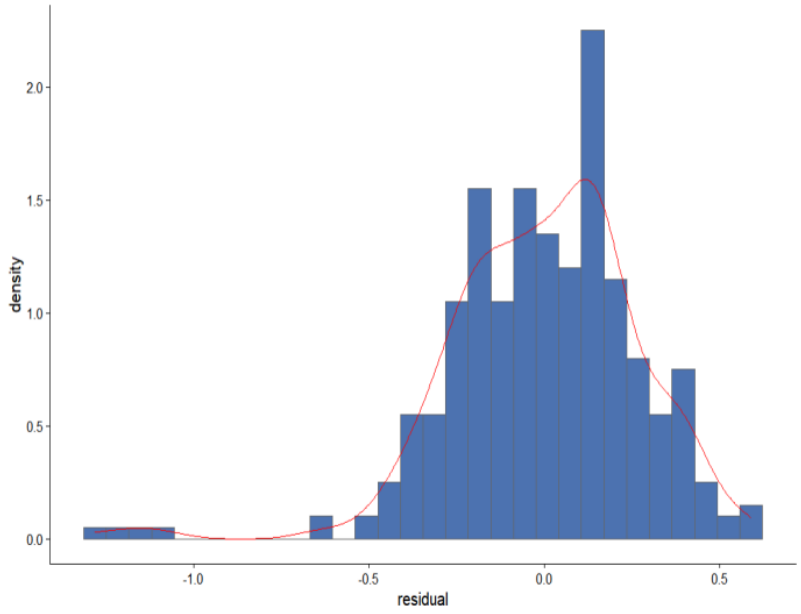


**Figure 9.** Residual distribution map.

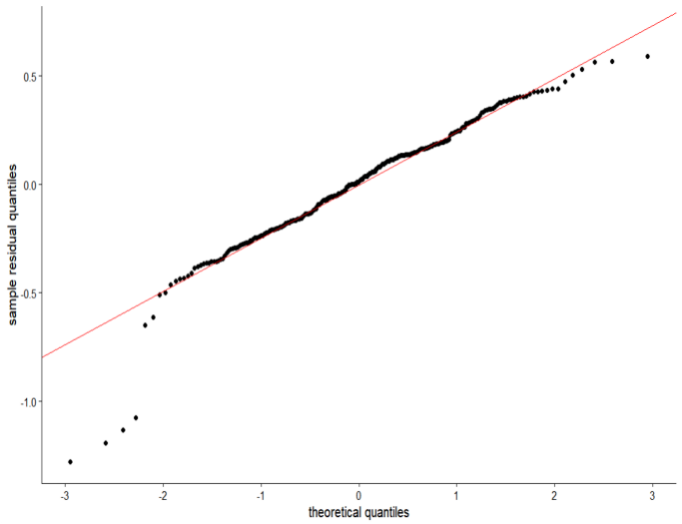


**Figure 10.** Residual Normal QQ Chart.

The Huber transformation is applied to the residuals. The Huber residual Figure 11 shows that the transformed residuals are concentrated near the cutoff, with flattened tails, while the overall shape remains approximately symmetric. This indicates that the transform attenuates the influence of extreme observations while preserving the major structural features of the residual distribution. Figure 12 further shows that, in the case of real data with certain heavy tails and outliers, the Huber transform effectively suppresses the influence of extreme observations and, empirically, satisfies the moment condition requirement that "Huber residuals are approximately orthogonal to covariates".

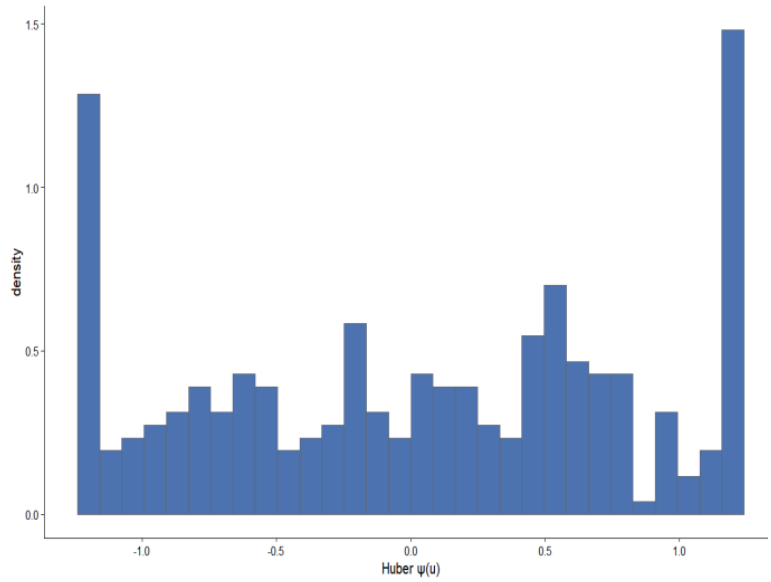


**Figure 11.** Huber residual distribution map.

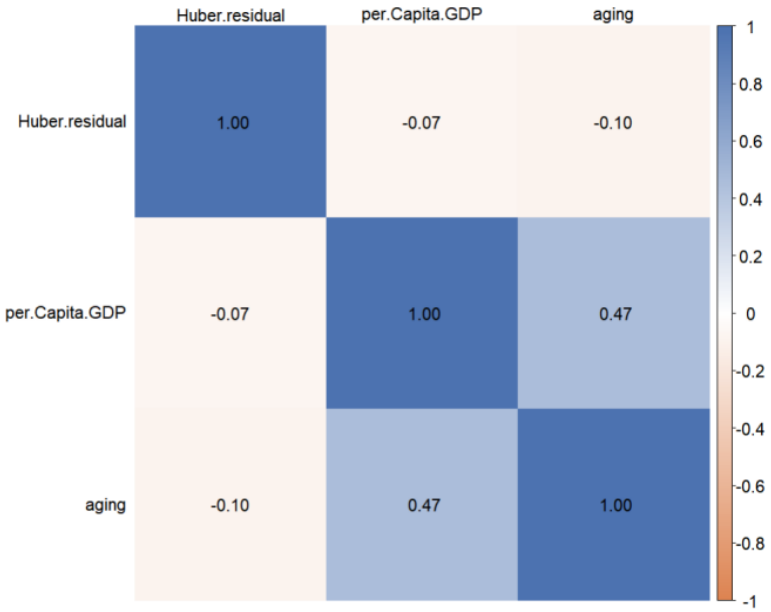


**Figure 12.** Huber residuals and covariate correlation analysis.

Let $Y_{it}$ be the per capita healthcare expenditure of province $i$ in year $t$, and let $x_{1,it}$ and $x_{2,it}$ denote per-capita GDP and the aging index. Based on log-transformation and standardization of the

variables, we construct a linear mixed-effects model $Y_{it} = \beta_0 + \beta_1 x_{1,it} + \beta_2 x_{2,it} + \alpha_i + \varepsilon_{it}$ with a province-specific random intercept only of the province, where $\alpha_i$ is the random intercept at the provincial level. The empirical analysis introduces only random effects in the intercept term.

The above model is estimated using BEL, empirical likelihood, restricted maximum likelihood (REML), maximum likelihood (ML), ordinary Bayes (Bayes), and generalized moment estimation (GMM). The BEL MCMC algorithm is diagnosed using the MH algorithm for sampling, running 10,000 iterations, with the first 5,000 iterations used as a burn-in period and samples retained every 10 iterations. We assess convergence using Effective Sample Size (ESS), Geweke's Z-statistic, Heidelberger-Welch stationarity test, and half-width test. The relevant diagnostic results are as follows:

**Table 14.** BEL's MCMC algorithm diagnostic results.

| Coefficients | ESS | Z-statistic | Stationarity Test P-value | half-width test |
|---|---|---|---|---|
| $\beta_0$ | 308.0775 | -1.189 | 0.0593 | 0.0034 |
| $\beta_1$ | 203.3085 | 1.625 | 0.0695 | 0.0030 |
| $\beta_2$ | 290.1714 | 1.048 | 0.6704 | 0.0020 |
| Mean Acceptance Rate | | | 0.212 | |

The diagnostic results show that the effective sample size for the three-dimensional parameters is approximately 200-300, the Geweke diagnostic values are all less than 2, and the Heidelberger-Welch stationarity and half-width tests pass, indicating that the MCMC chain has converged, and the Monte Carlo error is negligible. Based on the convergence of the algorithms, parameter estimation is performed for each method, and the estimation results are shown in the Table 15.

**Table 15.** Coefficient estimates under different estimation methods.

| Method | Gross Product per Capita | 95% confidence interval | Aging | 95% confidence interval |
|---|---|---|---|---|
| BEL | 0.2651 | （0.2164,0.3030） | 0.1999 | （0.1700,0.2317） |
| EL | 0.2702 | （0.2017,0.3386） | 0.1979 | （0.1551,0.2406） |
| REML | 0.3834 | （0.3388,0.4280） | 0.1371 | （0.1014,0.1728） |
| ML | 0.3825 | （0.3381,0.4268） | 0.1376 | （0.1021,0.1732） |
| Bayes | 0.3834 | （0.3349,0.4258） | 0.1382 | （0.0997,0.1743） |
| GMM | 0.3330 | （0.2561,0.4099） | 0.1808 | （0.1343,0.2273） |

As shown in Table 9, the six methods all estimate the coefficients of the two variables with the same sign, which are all positive. In practice, the increase in economic development level and population aging will significantly increase per capita medical and health care expenditure, and the estimation methods are consistent with the actual situation.

Numerically, BEL, EL, and GMM estimate GDP per capita lower than the estimates made by the classic Gaussian methods REML, ML, and Bayes; conversely, BEL, EL, and GMM estimate the degree of aging more than REML, ML, and Bayes. Compared to the classic methods, BEL, EL, and GMM, based on robust moment conditions, significantly reduce the marginal effect of economic variables and enhance the explanatory power of aging variables.

Observations that may have a strong leverage effect in classic LMM are calculated based on the standardized residuals $u_{it}$ based on the REML fitting results, and Huber weights $w_{it} = \begin{cases} 1, |u_{it}| \le c. \\ \dfrac{c}{|u_{it}|}, |u_{it}| > c \end{cases}$ are constructed accordingly. The average weights and truncation ratios are summarized by province, and the relevant results for the five provinces with the lowest average weights are shown in Table 16.

**Table 16.** Joint diagnosis of the random intercept and Huber weights of the five identified high-leverage provinces.

| Province | Average Weight | Truncated proportion | Average GDP per Capita | Coefficient of Variation of Medical Expenditure |
|---|---|---|---|---|
| Tibet | 0.553 | 1 | 10.7 | 0.498 |
| Jiangsu | 0.710 | 1 | 11.6 | 0.278 |
| Heilongjiang | 0.764 | 0.9 | 10.7 | 0.244 |
| Fujian | 0.784 | 1 | 11.4 | 0.292 |
| Gansu | 0.825 | 0.7 | 10.4 | 0.267 |

Huber weights, constructed from REML residuals, indicate that provinces with significantly lower average weights (below 1) and higher truncation rates primarily include a few regions such as Tibet, Jiangsu, Heilongjiang, and Fujian. Coastal developed provinces like Jiangsu and Fujian exhibit systematically higher healthcare expenditures at the same GDP and aging levels, while central and western provinces and northeastern provinces like Tibet, Gansu, Qinghai, and Heilongjiang show more pronounced fluctuations and structural heterogeneity in healthcare expenditures. This suggests that Huber weights do not simply apply a general deweighting to high-GDP provinces, but rather selectively weaken regions whose healthcare expenditures deviate significantly from the national linear pattern in relation to economic development and population structure.

To verify whether the difference between robust estimation and REML is primarily driven by these high-leverage provinces, we further conduct a sensitivity analysis: Tibet, Jiangsu, Heilongjiang, and five other provinces identified in the Huber diagnosis are excluded from the sample, and REML estimation is performed again on the remaining 26 provinces. Table 17 compares the REML results for the original sample and the sample with the removed provinces.

**Table 17.** Comparison of REML estimation under original and pruned samples.

| Coefficients | REML for all provinces | | | REML for samples with provinces removed | | |
|---|---|---|---|---|---|---|
| | Estimated value | Estimated standard deviation | t-value | Estimated value | Estimated standard deviation | t-value |
| $\beta_0$ | 7.3762 | 0.0512 | 144.159 | 7.4080 | 0.0360 | 205.776 |
| $\beta_1$ | 0.3834 | 0.0228 | 16.40 | 0.3470 | 0.0227 | 15.261 |
| $\beta_2$ | 0.1371 | 0.0182 | 7.524 | 0.1510 | 0.0182 | 8.285 |

The results show that after excluding high-leverage provinces, the coefficient for per capita GDP decreases from 0.3834 to approximately 0.3470, while the coefficient for aging population increases from 0.1371 to approximately 0.1510. The direction and magnitude of these changes are consistent with that of the BEL robust method, and the changes in the intercept and variance components are modest, indicating that the overall fit is not materially degraded. This suggests that the large marginal effect of GDP and the relatively small aging effect in the classic REML model are driven largely by a small number of high-leverage provinces, rather than reflecting a pervasive pattern of the entire sample.

## 5. Conclusions

In this paper, we investigate robust inference of linear mixed-effects models under complex data scenarios. We develop a BEL estimation method based on the Huber robust moment condition and annihilation of random effects, and design a structured Metropolis-Hastings sampling algorithm for posterior simulation. Through systematic numerical simulations and comparisons with other estimation methods, the results indicate that regardless of whether the perturbation originates from heavy tails and point anomalies at the observation level, or from misspecifications and outliers in the random effects distribution at the group-level, BEL exhibits stable properties in terms of parameter estimation bias and mean squared error. Empirical analysis based on panel data of healthcare expenditures from 31 provinces in China further demonstrates the practical applicability of this method, attenuating the impact of outlier observations.

## Authors' contributions

Hanfang Li conceived the project and Youxi Luo drafted the manuscript. Youxi Luo, Shiqi Zhou and Chaozhu Hu conducted the main analysis and contributed to the writing. All authors read and approved the final manuscript.

## Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article

## Acknowledgments

This work was supported by the National Social Science Fund of China [grant numbers 24BTJ068]; National Natural Science Fund of China [grant numbers 11701161]; Key Humanities and Social Science Fund of Hubei Provincial Department of Education [Grant Number 25D096].

## Conflict of interest

The authors declare no conflicts of interest.

## References


1. N. M. Larid, J. H. Ware, Random-Effects models for longitudinal data, *Biometrics*, **38** (1982), 963–974. https://doi.org/10.2307/2529876
2. C. J. Pinheiro, M. D. Bates, Mixed-Effects models in S and S-PLUS, *Springer, New York, NY*, (2000). https://doi.org/10.1198/jasa.2001.s411
3. M. Xin, D. A. Klinger, Hierarchical linear modelling of student and school effects on academic achievement, *Can. J. Educ.*, **25** (2000), 41–55. https://doi.org/10.2307/1585867
4. K. Huang, Z. Z. Ni, W. B. Cheng, Application of mixed linear models in repeated measures data in clinical trials, *Modern Preventive Medicine*, **32** (2005), 1584–1584. https://doi.org/10.3969/j.issn.1003-8507.2005.11.014
5. R. H. Baayen, D. J. Davidson, D. M. Bates, Mixed-effects Mmodeling with crossed random effects for subjects and items, *J. Mem. Lang.*, **59** (2007), 390–412. https://doi.org/10.1016/j.jml.2007.12.005
6. P. X. K. Song, P. Zhang, A. Qu, Maximum likelihood inference in robust linear mixed-effects models using multivariate t distributions, *Stat. Sinica*, **17** (2007), 929–943. https://www.jstor.org/stable/24307706
7. W. L. Wang, Mixture of multivariate t linear mixed models for multi-outcome longitudinal data with heterogeneity, *Stat. Sinica*, **27** (2017), 733–760. https://www.jstor.org/stable/26383298
8. P. J. Huber, Robust estimation of a location parameter, *Annals Math. Stat.*, **35** (1964), 73–101. https://doi.org/10.1007/978-1-4612-4380-9_35
9. R. Mohammadi, I. Kazemi, A robust linear mixed-effects model for longitudinal data using an innovative multivariate skew-huber distribution, *J. Multivariate Anal.*, **187** (2022), 104856. https://doi.org/10.1016/J.JMVA.2021.104856
10. S.Y. Wu, Huber Loss: A mixed effects model for longitudinal data and its application, *Guangxi Normal University*, (2025). https://doi.org/10.27036/d.cnki.ggxsu.2025.000626
11. Q. H. Wang, A review of the development of empirical likelihood statistical inference methods, *Adv. Math.*, **33** (2004), 141–151. https://doi.org/10.3969/j.issn.1000-0917.2004.02.002
12. J. Qin, J. Lawless, Empirical likelihood and general estimating equations, *Ann. Stat.*, **22** (2007), 300–325. https://doi.org/10.1214/AOS/1176325370
13. N. Tang, P. Zhao, H. Zhu, Empirical likelihood for estimating equations with Nonignorably missing data, *Stat. Sinica*, **24** (2014), 723–747. http://dx.doi.org/10.5705/ss.2012.254
14. C. Leng, Penalized empirical likelihood and growing dimensional general estimating equations, *Biometrika*, **99** (2012), 703–716. https://doi.org/10.1093/biomet/ass014
15. B. W. Chen, P. X. Zhao, X. R. Tang, Y. P. Yang, Efficient and robust empirical likelihood inference for linear mixed-effects models, *Acta Math. Appl. Sinica*, **33** (2020), 88–893. https://doi.org/10.13642/j.cnki.42-1184/o1.2020.04.008
16. H. Liu, Statistical inference and applications of mixed-effects Bayesian models, *Stat. Decision*, **36** (2020), 38–42. https://doi.org/10.13546/j.cnki.tjyjc.2020.13.008
17. J. Ouyang, H. Bondell, Bayesian analysis of longitudinal data via empirical likelihood, *Comput. Stat. Data An.*, **187** (2023), 107785. https://doi.org/10.1016/J.CSDA.2023.107785
18. A. N. Lazar, Bayesian empirical likelihood, *Biometrika*, **90** (2003), 319–326. https://doi.org/10.1093/biomet/90.2.319

19. T. Lancaster, S. Jae Jun, Bayesian quantile regression methods, *J. Appl. Economet.*, **25** (2010), 287–307. https://doi.org/10.1002/jae.1069
20. A. T. Porter, S. H. Holan, C. K. Wikle, Bayesian semiparametric hierarchical empirical likelihood spatial models, *J. Stat. Plan. Infer.*, **165** (2015), 78–90. https://doi.org/10.1016/j.jspi.2015.04.002
21. A. Vexler, J. Yu, N. Lazar, Bayesian empirical likelihood methods for quantile comparisons, *J. Korean Stat. Soc.*, **46** (2017), 518–538. https://doi.org/10.1016/j.jkss.2017.03.002
22. J. W. Wang, C. Z. Hu, H. F. Li, Y. X. Luo, Bayesian empirical likelihood inference of composite quantile regression, *J. Guangxi Normal University*, **42** (2024), 130–140. https://doi.org/10.16088/j.issn.1001-6600.2023110304
23. X. G. Dong, X. R. Liu, C. J. Wang, X. H. Yuan, Bayesian empirical likelihood of accelerating failure model under right-censored data, *J. Appl. Stat. Manag.*, **39** (2020), 838–844. https://doi.org/10.13860/j.cnki.sltj.20200818-001
24. Y. S. Qin, Q. Z. Lei, Progress in empirical likelihood research of spatial econometric models, *J. Guangxi Normal University*, **40** (2022), 138–149. https://doi.org/10.16088/j.issn.1001-6600.2022012002
25. S. Y. Zhou, X. Y. Qian, Application of empirical likelihood Bayesian calculation method in stochastic fluctuation models, *Math. Practice Theory*, **50** (2020), 8–15. https://doi.org/10.20266/j.math.2020.06.002