JIRSS (2027)
Vol. 26, No. 1, pp 26-53
DOI: 10.22034/jirss.2026.2096683.1234

Asymptotic Score Test Statistic in Multilevel Models with Application to the Income Tax Payments Data

Sedigheh Zamani Mehreyan[1], **Ramin Kazemi**[1]

[1]Department of Statistics, Imam Khomeini International University, Qazvin, Iran.

Received: 2026-08-05, Accepted: 2026-11-01, Published online: 2026-11-07

Abstract. Multilevel structural equation modeling is a statistical framework of major relevance for research concerned with people's interpersonal dynamics. The purpose of this study is to determine the asymptotic distribution of score test statistics for hypothesis tests in a multilevel regression. We estimate the model parameters using the EM algorithm. Then, we derive the asymptotic distribution of score test statistics for regression tests. Through simulation studies, the performance of the EM algorithm is evaluated. Additionally, the relative frequency of the regression test is examined. A real data analysis carried out to illustrate the theoretically obtained procedures. The proposed method is then applied to predict the logarithm of the total positive tax liability from individual characteristics and economic characteristics, to demonstrate the practical application of the theoretical findings.

Keywords. EM algorithm, multilevel model, random intercept, random slopes, score test.

MSC: 62H12, 62F03, 62E20.

1 Introduction

The simple linear regression model is defined as

$$y = B_0 + B_1 x + R$$

Sedigheh Zamani Mehreyan (s.zamani@sci.ikiu.ac.ir).
Corresponding Author: Ramin Kazemi (r.kazemi@sci.ikiu.ac.ir).

where, B_0 represents intercept, B_1 is the slope of the line and the random term R is assumed to follow a normal distribution with mean 0 and constant variance σ^2. Suppose we observe n data pairs and call them $\{(x_i, y_i), i = 1, 2, \cdots, n\}$. We can describe the underlying relationship between y_i and x_i involving this error term R_i by

$$y_i = B_0 + B_1 x_i + R_i,$$

$$(\hat{B}_0, \hat{B}_1) = argmin(Q(B_0, B_1)),$$

where $Q(B_0, B_1)$ is an objective function. If the same variables y and x are measured for m groups and for each group regression line is fitted then we have m regression lines. The simple linear regression model describing a simple linear relation between the response variable y and the explanatory variable x is

$$y_{ij} = B_{0j} + B_{1j} x_{ij} + R_{ij}, i = 1, 2 \cdots, n \quad j = 1, 2, \cdots, m \tag{1.1}$$

by fitting such m regression models we estimate $2m + 1$ parameters, $(B_{0j}, B_{ij}), j = 1, 2, \cdots, m$ and σ^2.

There exist a natural hierarchy in all multilevel problems and for the proper exploration, specialized analytical tools are required. Multilevel analytical tools provide the proper estimation of such type of problems. The variance components models (Aitkin and Longford, 1986), the hierarchical linear models (Raudenbush and Bryk, 1986) and the random coefficient models (Longford, 1987) are known as multilevel regression. A crucial issue in the specification of these models is the choice of a random-effects structure (i.e., random intercept and slopes), which can have important consequences both for the precision of parameter estimates (Heisig et al., 2017) and for statistical inference (Berkhof and Kampen, 2004; Barr, 2013; Bryan and Jenkins, 2016; Schmidt-Catran and Fairbrother, 2016).

Let y_{ij} be the ij^{th} observation of a response variable y measured at level 1 and the corresponding explanatory variable x observation x_{ij} (j refers to level 2 unit and i represents level 1 unit), then the simple linear regression model is model (1.1). Let B_{0j} and B_{1j} be the random variables as their magnitude varies in m linear regression lines and assume that

$$\begin{aligned} B_{0j} &= \gamma_{00} + u_{0j}, \\ B_{1j} &= \gamma_{10} + u_{1j}, \end{aligned} \tag{1.2}$$

where, u_{0j} and u_{1j} are the unexplained parts of (B_{0j}, B_{1j}) with parameters $E(u_{0j}) = 0$, $E(u_{1j}) = 0$, $Var(u_{0j}) = \tau_0^2$, $Var(u_{1j}) = \tau_1^2$ and $Cov(u_{0j}, u_{1j}) = \tau_{01}$ and follow normal distribution, $u_{0j} \sim N(0, \tau_0^2)$ and $u_{1j} \sim N(0, \tau_1^2)$.

The model (1.1) can be written as

$$\begin{aligned} y_{ij} &= \gamma_{00} + u_{0j} + (\gamma_{10} + u_{1j}) x_{ij} + R_{ij} \\ &= \gamma_{00} + \gamma_{10} x_{ij} + u_{0j} + u_{1j} x_{ij} + R_{ij}. \end{aligned}$$

It can also be expressed in matrix form as follows:

$$\mathbf{Y} = \mathbf{X\Gamma} + \boldsymbol{\epsilon},$$

where

$$\mathbf{Y} = \begin{bmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{nm} \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} \\ 1 & x_{21} \\ \vdots & \\ 1 & x_{nm} \end{bmatrix}, \mathbf{\Gamma} = \begin{bmatrix} \gamma_{00} \\ \gamma_{10} \end{bmatrix}, \boldsymbol{\epsilon} = \mathbf{R} + \mathbf{X}\boldsymbol{U}, \mathbf{R} = \begin{bmatrix} R_{11} \\ R_{21} \\ \vdots \\ R_{nm} \end{bmatrix}, \boldsymbol{U} = \begin{bmatrix} u_0 \\ u_1 \end{bmatrix}.$$

Note that $\boldsymbol{\epsilon}$ follows normal distribution with

$$E(\boldsymbol{\epsilon}) = \mathbf{0}$$

$$\mathbf{V} = Var(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{I}_n + \mathbf{X}\mathbf{\Sigma}_u \mathbf{X}'$$

where $\mathbf{I}_n$ is an $(n \times n)$ identity matrix and $\Sigma_u = \begin{bmatrix} \tau_0^2 & \tau_{01} \\ \tau_{10} & \tau_1^2 \end{bmatrix}$.
The Generalized Least Square (GLS) estimates of $\mathbf{\Gamma}$ can be obtained by using the relation:

$$\hat{\mathbf{\Gamma}} = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$$

with variance-covariance matrix $\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}$.
Frees (2004) extended the Mundlak approach to handle random slopes. Wooldridge (2005) showed that under seemingly benign conditions that traditional fixed-effects estimation of random-intercept models is robust against correlations between unit-level covariates and random slopes. Kim and Frees (2007) used generalized method of moments estimation to extend the Hausman-Taylor approach to multilevel models with random slopes. However, their method is difficult to implement, making the fixed-effects approach more feasible in practice. Bates et al. (2014) investigated estimation of the coefficients of unit-level and cluster-level covariates in multilevel models in the presence of two sources of endogeneity; (nonzero) correlations between unit-level covariates and both the random intercept and random slopes. Grund et al. (2016) discussed applications of multiple imputation in multilevel random-coefficient models, theoretical challenges posed by slope variation, and the current limitations of standard multiple imputation software.
Zhao and Edland (2021) derived power calculation formulas for cohort studies and clinical trials using the longitudinal mixed effects model with random slopes and intercepts to compare rate of change across groups and demonstrated how differences in variance parameters across groups, typically overlooked in power calculations, can have a dramatic effect on statistical power. For frequentist mixed models, excluding random slopes (differences between individuals in the direction and size of an effect) from a model when they are in the data can lead to a substantial increase in false-positive conclusions in null-hypothesis tests. Oberpriller et al. (2022) demonstrated

through five simulations that the same is true for Bayesian hypothesis testing with mixed models, which often yield Bayes factors reflecting very strong evidence for a mean effect on the population level even if there was no such effect. Barendse and Rosseel (2023) investigated the pairwise maximum likelihood estimation method with computationally intensive multilevel random intercept and random slope structural equation models in discrete data. Humberg et al. (2024) evaluated the nonlinear multilevel structural equation modeling approach for the three classes of nonlinear effects (U-shaped, moderation, congruence) and compared it with three simpler approaches: a simple two-step approach, a single-indicator approach, and a plausible values approach.
The rest of the paper is structured as follows: in Section 2, we consider random intercepts and slopes model and propose EM algorithm. In Section 3 the statistical behavior of the score test statistic for the regression test is derived. In Section 4, we study the obtained theoretical results by simulation. Also, we illustrate our theoretical results with the analysis of a real dataset in Section 5.

2 Estimation and model selection

The expectation-maximization (EM) algorithm is a well-established likelihood-based approach for parameter estimation in models involving latent or unobserved variables, originally formalized by Dempster et al. (1977). It has also been widely used for estimation in linear mixed-effects models, including models with random effects for longitudinal data (Laird and Ware, 1982). In this study, we employ the EM framework for the proposed linear mixed-effects regression model with random intercepts and random slopes. For the specific model structure considered here, we derive the corresponding likelihood components, estimation equations, and EM steps. These derivations provide the basis for developing and theoretically establishing the proposed inferential procedure for testing $H_0 : \mathbf{\Gamma} = \mathbf{0}$.
We consider the problem of estimation in a random-intercept and random-slope. Let y_{ij} for $i = 1, \ldots, n$ and $j = 1, \ldots, m$ be the observed data and follow random intercepts and slopes model as:
Level 1:

$$y_{ij} = B_{0j} + B_{1j}x_{ij} + R_{ij}, \tag{2.1}$$

where R_{ij}'s are normally distributed with zero mean and variance σ^2, i.e. iid $N(0, \sigma^2)$.
Level 2:

$$B_{0j} = \gamma_{00} + u_{0j},$$
$$B_{1j} = \gamma_{10} + u_{1j},$$

where

$$\mathbf{U}_j = \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_0^2 & \tau_{01} \\ \tau_{10} & \tau_1^2 \end{bmatrix}\right),$$

and is independent of R_{ij}. Model (2.1) can be rewritten as:

$$y_{ij} = \underbrace{\gamma_{00} + \gamma_{10}x_{ij}}_{\text{fixed}} + \underbrace{u_{0j} + u_{1j}x_{ij} + R_{ij}}_{\text{Random}}. \tag{2.2}$$

Also, the model (2.2) can be rewritten in matrix terms as:

$$\mathbf{Y}_i = \mathbf{X}_i\mathbf{\Gamma} + \mathbf{Z}_i\mathbf{U} + \mathbf{R}_i, \qquad i = 1, \ldots, n, \tag{2.3}$$

where for $i = 1, \ldots, n$,

$$\mathbf{Y}_i = \begin{bmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{im} \end{bmatrix}, \quad \mathbf{\Gamma} = \begin{bmatrix} \gamma_{00} \\ \gamma_{10} \end{bmatrix}, \mathbf{U} = \begin{bmatrix} \mathbf{U}_1 \\ \mathbf{U}_2 \\ \vdots \\ \mathbf{U}_m \end{bmatrix}, \quad \mathbf{U}_j = \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix}, \mathbf{R}_i = \begin{bmatrix} R_{i1} \\ R_{i2} \\ \vdots \\ R_{im} \end{bmatrix}, \quad \mathbf{X}_i = \begin{bmatrix} 1 & x_{i1} \\ 1 & x_{i2} \\ \vdots & \vdots \\ 1 & x_{im} \end{bmatrix}.$$

Here, $\mathbf{\Gamma}$ is a common 2×1 fixed-effects vector for all levels $j = 1, \ldots, m$. Thus, γ_{00} and γ_{10} are common fixed effects, whereas u_{0j} and u_{1j} represent the level-specific random effects. The matrix $\mathbf{Z}_i$ is an $m \times 2m$ design matrix for the random effects. Its jth row contains 1 and x_{ij} in columns $2j - 1$ and $2j$, respectively, and zeros elsewhere.

Note that the elements of $\mathbf{Z}_j$ are not the z_j's in the level 2 models. $\mathbf{Z}_j$ is a sub-matrix of $\mathbf{X}_j$.

Note that model (2.2) follows a normal distribution with

$$E(Y_{ij}) = \gamma_{00} + \gamma_{10}x_{ij}$$

and variance

$$Var(Y_{ij}) = \tau_0^2 + 2\tau_{01}x_{ij} + \tau_1^2x_{ij}^2 + \sigma^2.$$

In terms of matrix notation, for each $i = 1, \ldots, n$,

$$\mathbf{Y}_i \sim N_m(\boldsymbol{\mu}_y, \mathbf{\Sigma}_y),$$

where

$$\boldsymbol{\mu}_y = E(\mathbf{Y}_i) = \mathbf{X}_i\mathbf{\Gamma},$$

and

$$\mathbf{\Sigma}_y = Var(\mathbf{Y}_i) = \mathbf{Z}_i\mathbf{\Sigma}_U\mathbf{Z}_i^T + \mathbf{\Sigma}_R,$$

such that

$$\mathbf{\Sigma}_U = \mathbf{I}_m \otimes \mathbf{\Sigma}_u, \qquad \mathbf{\Sigma}_u = \begin{bmatrix} \tau_0^2 & \tau_{01} \\ \tau_{10} & \tau_1^2 \end{bmatrix},$$

and

$$\mathbf{\Sigma}_R = \sigma^2\mathbf{I}_m$$

and $\mathbf{I}_m$ is the identity matrix of order m. Here, $\mathbf{\Sigma}_u$ is the 2×2 covariance matrix of the random intercept and random slope for each level j, while $\mathbf{\Sigma}_U$ is the corresponding

$2m \times 2m$ covariance matrix of the complete random-effects vector $\mathbf{U}$. For the complete sample, define

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_n \end{bmatrix}, \qquad \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_n \end{bmatrix}, \qquad \mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \\ \vdots \\ \mathbf{Z}_n \end{bmatrix}.$$

Thus, $\mathbf{Y}$ is an $nm \times 1$ vector, $\mathbf{X}$ is an $nm \times 2$ matrix, and $\mathbf{Z}$ is an $nm \times 2m$ matrix. The expectation-maximization, EM, algorithm is an efficient iterative procedure to compute the maximum likelihood, ML, estimate in the presence of missing or hidden data. The EM algorithm consists of two processes: In the expectation-step, E-step, the missing data are estimated given the observed data and current estimate of the model parameters. In the maximization-step, M-step, the likelihood function is maximized under the assumption that the missing data are known.
Let $\mathbf{y}_c = (\mathbf{y}, \mathbf{u})^T$ be complete-data vector, where $\mathbf{y}$ is the observed data and $\mathbf{u}$ be the vector of unobserved or missing data. From model (2.3), we have

$$\mathbf{Y}_c = \begin{pmatrix} \mathbf{Y} \\ \mathbf{U} \end{pmatrix} \sim N_{3m}\left(\begin{bmatrix} \boldsymbol{\mu}_y \\ 0 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_y & \boldsymbol{\Sigma}_{yu} \\ \boldsymbol{\Sigma}_{uy} & \boldsymbol{\Sigma}_U \end{bmatrix} \right),$$

where

$$\boldsymbol{\mu}_y = E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\Gamma},$$

$$\boldsymbol{\Sigma}_y = Var(\mathbf{Y}) = \mathbf{Z}\boldsymbol{\Sigma}_U\mathbf{Z}^T + \sigma^2\mathbf{I}_{nm},$$

and

$$\boldsymbol{\Sigma}_{yu} = Cov(\mathbf{Y}, \mathbf{U}) = \mathbf{Z}\boldsymbol{\Sigma}_U, \qquad \boldsymbol{\Sigma}_{uy} = \boldsymbol{\Sigma}_U\mathbf{Z}^T.$$

Then, the distribution of $\mathbf{U}$ conditional on $\mathbf{Y}$ is

$$\mathbf{U}|\mathbf{Y} = \mathbf{y} \sim N_{2m}\left(\boldsymbol{\Sigma}_{uy}\boldsymbol{\Sigma}_y^{-1}(\mathbf{Y} - \boldsymbol{\mu}_y), \boldsymbol{\Sigma}_U - \boldsymbol{\Sigma}_{uy}\boldsymbol{\Sigma}_y^{-1}\boldsymbol{\Sigma}_{yu}\right).$$

We employ the EM framework for the proposed linear mixed-effects regression model with random intercepts and random slopes. For the specific model structure considered here, we derive the corresponding likelihood components, estimation equations, and EM steps. These derivations subsequently provide the basis for developing and theoretically establishing the proposed inferential procedure for testing $H_0 : \boldsymbol{\Gamma} = \mathbf{0}$.
The complete log-likelihood function can be expressed as

$$\begin{aligned} l(\boldsymbol{\theta}; y_c) &= l(\boldsymbol{\theta}; y|u) + l(\boldsymbol{\theta}; u) \\ &= C - \frac{n}{2}\log|\boldsymbol{\Sigma}_R| - \frac{1}{2}\sum_{i=1}^{n}(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\Gamma} - \mathbf{z}_i\mathbf{u})^T\boldsymbol{\Sigma}_R^{-1}(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\Gamma} - \mathbf{z}_i\mathbf{u}) \\ &\quad - \frac{m}{2}\log|\boldsymbol{\Sigma}_u| - \frac{1}{2}\sum_{j=1}^{m}\mathbf{u}_j^T\boldsymbol{\Sigma}_u^{-1}\mathbf{u}_j. \end{aligned}$$

where C is a constant that is independent of $\mathbf{\Sigma}_R$ and $\mathbf{\Sigma}_u$. Since the random effects are defined for $j = 1, \ldots, m$, their contribution to the complete log-likelihood is expressed as a sum over j. The covariance matrix $\mathbf{\Sigma}_u$ is the covariance matrix for each $\mathbf{U}_j$. Let $\boldsymbol{\theta}^{(0)} = \left(\mathbf{\Gamma}^{(0)}, \mathbf{\Sigma}_R^{(0)}, \mathbf{\Sigma}_u^{(0)}\right)$ be starting values of $\boldsymbol{\theta} = (\mathbf{\Gamma}, \mathbf{\Sigma}_R, \mathbf{\Sigma}_u)$, so

$$\begin{aligned} Q &= E\left(l(\boldsymbol{\theta}; y_c \mid y, \boldsymbol{\theta}^{(0)})\right) \\ &= C - \frac{n}{2}\log|\mathbf{\Sigma}_R| - \frac{1}{2}\sum_{i=1}^{n} E\left\{(\mathbf{y}_i - \mathbf{x}_i\mathbf{\Gamma} - \mathbf{z}_i\mathbf{U})^T \mathbf{\Sigma}_R^{-1}(\mathbf{y}_i - \mathbf{x}_i\mathbf{\Gamma} - \mathbf{z}_i\mathbf{U}) \mid \mathbf{y}, \boldsymbol{\theta}^{(0)}\right\} \\ &- \frac{m}{2}\log|\mathbf{\Sigma}_u| - \frac{1}{2}\sum_{j=1}^{m} E\left\{\mathbf{U}_j^T \mathbf{\Sigma}_u^{-1}\mathbf{U}_j \mid \mathbf{y}, \boldsymbol{\theta}^{(0)}\right\}. \end{aligned} \tag{2.4}$$

The third term of Equation (2.4) is equal to

$$\begin{aligned} E\left[\mathbf{R}_i^T \mathbf{\Sigma}_R^{-1}\mathbf{R}_i\right] &= E\left[(\mathbf{y}_i - \mathbf{x}_i\mathbf{\Gamma} - \mathbf{z}_i\mathbf{U})^T \mathbf{\Sigma}_R^{-1}(\mathbf{y}_i - \mathbf{x}_i\mathbf{\Gamma} - \mathbf{z}_i\mathbf{U}) \mid \mathbf{y}, \boldsymbol{\theta}^{(0)}\right] \\ &= E\left[tr\left\{\mathbf{\Sigma}_R^{-1}(\mathbf{y}_i - \mathbf{x}_i\mathbf{\Gamma} - \mathbf{z}_i\mathbf{U})(\mathbf{y}_i - \mathbf{x}_i\mathbf{\Gamma} - \mathbf{z}_i\mathbf{U})^T \mid \mathbf{y}, \boldsymbol{\theta}^{(0)}\right\}\right] \\ &= tr\left\{\mathbf{\Sigma}_R^{-1} E\left[(\mathbf{y}_i - \mathbf{x}_i\mathbf{\Gamma} - \mathbf{z}_i\mathbf{U})(\mathbf{y}_i - \mathbf{x}_i\mathbf{\Gamma} - \mathbf{z}_i\mathbf{U})^T \mid \mathbf{y}, \boldsymbol{\theta}^{(0)}\right]\right\} \\ &= (\mathbf{y}_i - \mathbf{x}_i\mathbf{\Gamma} - \mathbf{z}_i\tilde{\mathbf{u}}_1)^T \mathbf{\Sigma}_R^{-1}(\mathbf{y}_i - \mathbf{x}_i\mathbf{\Gamma} - \mathbf{z}_i\tilde{\mathbf{u}}_1) \\ &+ tr\left\{\mathbf{\Sigma}_R^{-1}\mathbf{z}_i\tilde{\mathbf{b}}_1\mathbf{z}_i^T\right\}, \end{aligned} \tag{2.5}$$

where tr denotes trace,

$$\tilde{\mathbf{u}}_1 = E\left(\mathbf{U} \mid \mathbf{y}, \boldsymbol{\theta}^{(0)}\right) = \mathbf{\Sigma}_{uy}^{(0)}\left(\mathbf{\Sigma}_y^{(0)}\right)^{-1}\left(\mathbf{y} - \boldsymbol{\mu}_y^{(0)}\right)$$

and

$$\tilde{\mathbf{b}}_1 = Var\left(\mathbf{U} \mid \mathbf{y}, \boldsymbol{\theta}^{(0)}\right) = \mathbf{\Sigma}_U^{(0)} - \mathbf{\Sigma}_{uy}^{(0)}\left(\mathbf{\Sigma}_y^{(0)}\right)^{-1}\mathbf{\Sigma}_{yu}^{(0)}.$$

The last term of Equation (2.4) is equal to

$$\begin{aligned} E\left(\mathbf{U}_j^T \mathbf{\Sigma}_u^{-1}\mathbf{U}_j \mid \mathbf{y}, \boldsymbol{\theta}^{(0)}\right) &= tr\left\{\mathbf{\Sigma}_u^{-1} E\left(\mathbf{U}_j \mid \mathbf{y}, \boldsymbol{\theta}^{(0)}\right) E\left(\mathbf{U}_j^T \mid \mathbf{y}, \boldsymbol{\theta}^{(0)}\right)\right\} \\ &+ tr\left\{\mathbf{\Sigma}_u^{-1} Var\left(\mathbf{U}_j \mid \mathbf{y}, \boldsymbol{\theta}^{(0)}\right)\right\} \\ &= \tilde{\mathbf{u}}_{1j}^T \mathbf{\Sigma}_u^{-1}\tilde{\mathbf{u}}_{1j} + tr\left\{\mathbf{\Sigma}_u^{-1}\widetilde{\mathbf{B}}_j\right\}, \end{aligned} \tag{2.6}$$

where

$$\tilde{\mathbf{u}}_{1j} = E\left(\mathbf{U}_j \mid \mathbf{y}, \boldsymbol{\theta}^{(0)}\right)$$

and

$$\widetilde{\mathbf{B}}_j = Var\left(\mathbf{U}_j \mid \mathbf{y}, \boldsymbol{\theta}^{(0)}\right).$$

By substituting Equations (2.5) and (2.6) in Equation (2.4), we have

$$\begin{aligned} Q = C &- \frac{n}{2}\log|\boldsymbol{\Sigma}_R| \\ &- \frac{1}{2}\sum_{i=1}^{n}\left[\left(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\Gamma} - \mathbf{z}_i\tilde{\mathbf{u}}_1\right)^T\boldsymbol{\Sigma}_R^{-1}\left(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\Gamma} - \mathbf{z}_i\tilde{\mathbf{u}}_1\right) + tr\left\{\boldsymbol{\Sigma}_R^{-1}\mathbf{z}_i\tilde{\mathbf{b}}_1\mathbf{z}_i^T\right\}\right] \\ &- \frac{m}{2}\log|\boldsymbol{\Sigma}_u| - \frac{1}{2}\sum_{j=1}^{m}\left[\tilde{\mathbf{u}}_{1j}^T\boldsymbol{\Sigma}_u^{-1}\tilde{\mathbf{u}}_{1j} + tr\left\{\boldsymbol{\Sigma}_u^{-1}\widetilde{\mathbf{B}}_j\right\}\right]. \end{aligned} \tag{2.7}$$

We obtain estimators by solving the estimating equations as:

$$\begin{aligned} 0 = \frac{\partial Q}{\partial\boldsymbol{\Gamma}} &= -\frac{1}{2}\frac{\partial}{\partial\boldsymbol{\Gamma}}\left[\sum_{i=1}^{n}\left(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\Gamma} - \mathbf{z}_i\tilde{\mathbf{u}}_1\right)^T\boldsymbol{\Sigma}_R^{-1}\left(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\Gamma} - \mathbf{z}_i\tilde{\mathbf{u}}_1\right)\right. \\ &\left. + tr\left\{\boldsymbol{\Sigma}_R^{-1}\mathbf{z}_i\tilde{\mathbf{b}}_1\mathbf{z}_i^T\right\}\right] \\ &= \sum_{i=1}^{n}\mathbf{x}_i^T\boldsymbol{\Sigma}_R^{-1}\left(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\Gamma} - \mathbf{z}_i\tilde{\mathbf{u}}_1\right), \end{aligned}$$

so

$$\hat{\boldsymbol{\Gamma}} = \left(\sum_{i=1}^{n}\mathbf{x}_i^T\boldsymbol{\Sigma}_R^{-1}\mathbf{x}_i\right)^{-1}\left(\sum_{i=1}^{n}\mathbf{x}_i^T\boldsymbol{\Sigma}_R^{-1}\left(\mathbf{y}_i - \mathbf{z}_i\tilde{\mathbf{u}}_1\right)\right). \tag{2.8}$$

Also,

$$\begin{aligned} \frac{\partial Q}{\partial\boldsymbol{\Sigma}_R} &= \frac{\partial Q}{\partial\boldsymbol{\Sigma}_R^{-1}}\left(\frac{\partial\boldsymbol{\Sigma}_R^{-1}}{\partial\boldsymbol{\Sigma}_R}\right) \\ &= \frac{1}{2}\sum_{i=1}^{n}Vec^T\left(\boldsymbol{\Sigma}_R^T\right)\left(\boldsymbol{\Sigma}_R^{-T}\otimes\boldsymbol{\Sigma}_R^{-1}\right) \\ &\quad - \frac{1}{2}\sum_{i=1}^{n}\left(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\Gamma} - \mathbf{z}_i\tilde{\mathbf{u}}_1\right)^T\otimes\left(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\Gamma} - \mathbf{z}_i\tilde{\mathbf{u}}_1\right) \\ &\quad + tr\left\{\mathbf{z}_i\tilde{\mathbf{b}}_1\mathbf{z}_i^T\right\}\left(\boldsymbol{\Sigma}_R^{-T}\otimes\boldsymbol{\Sigma}_R^{-1}\right) = 0, \end{aligned}$$

where $\otimes$ denotes Kronecker product and $Vec(A)$ denotes vectorization of matrix A. So we have

$$\hat{\boldsymbol{\Sigma}}_R = \frac{1}{n}\sum_{i=1}^{n}\left[\left(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\Gamma} - \mathbf{z}_i\tilde{\mathbf{u}}_1\right)\left(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\Gamma} - \mathbf{z}_i\tilde{\mathbf{u}}_1\right)^T + \mathbf{z}_i\tilde{\mathbf{b}}_1\mathbf{z}_i^T\right]. \tag{2.9}$$

Similarly,

$$\hat{\boldsymbol{\Sigma}}_u = \frac{1}{m}\sum_{j=1}^{m}E\left(\mathbf{U}_j\mathbf{U}_j^T \mid \mathbf{y}, \boldsymbol{\theta}^{(0)}\right). \tag{2.10}$$

The corresponding covariance matrix of the complete random-effects vector is

$$\hat{\boldsymbol{\Sigma}}_U = \mathbf{I}_m \otimes \hat{\boldsymbol{\Sigma}}_u.$$

Thus the EM algorithm is described as follows:
Step 0: Set $r = 0$ and choose starting value $\boldsymbol{\theta}^{(0)}$.
Step 1: For $r \geq 0$, calculate

$$\tilde{\mathbf{u}}_1^{(r+1)} = E\left(\mathbf{U} \mid \mathbf{y}, \boldsymbol{\theta}^{(r)}\right) = \boldsymbol{\Sigma}_{uy}^{(r)} \left(\boldsymbol{\Sigma}_y^{(r)}\right)^{-1} \left(\mathbf{y} - \boldsymbol{\mu}_y^{(r)}\right)$$

and for each $j = 1, \ldots, m$ calculate

$$\tilde{\mathbf{u}}_{1j}^{(r+1)} = E\left(\mathbf{U}_j \mid \mathbf{y}, \boldsymbol{\theta}^{(r)}\right)$$

and

$$\widetilde{\mathbf{B}}_j^{(r+1)} = Var\left(\mathbf{U}_j \mid \mathbf{y}, \boldsymbol{\theta}^{(r)}\right).$$

Moreover,

$$E\left(\mathbf{U}_j \mathbf{U}_j^T \mid \mathbf{y}, \boldsymbol{\theta}^{(r)}\right) = \widetilde{\mathbf{B}}_j^{(r+1)} + \tilde{\mathbf{u}}_{1j}^{(r+1)} \left(\tilde{\mathbf{u}}_{1j}^{(r+1)}\right)^T.$$

Step 2: For $r \geq 0$, compute $\hat{\boldsymbol{\Gamma}}^{(r+1)}$, $\hat{\boldsymbol{\Sigma}}_R^{(r+1)}$ and $\hat{\boldsymbol{\Sigma}}_u^{(r+1)}$ based on Equations (2.8), (2.9) and (2.10). The covariance matrix of the complete random-effects vector is then updated as

$$\hat{\boldsymbol{\Sigma}}_U^{(r+1)} = \mathbf{I}_m \otimes \hat{\boldsymbol{\Sigma}}_u^{(r+1)}.$$

Step 3: Iterate Steps 1 and 2 from $r = 1$ until reaching convergence.

3 Limiting behavior of the score test

Several established inferential procedures are available for testing hypotheses on fixed effects in linear mixed-effects models, including Wald-type tests (Wald, 1943) and likelihood ratio tests (Wilks, 1938). For finite-sample inference, degrees-of-freedom approximations such as the Satterthwaite method (Satterthwaite, 1946) and the Kenward–Roger method (Kenward and Roger, 1997) have also been widely used. In contrast to these existing approaches, the present study develops a score-based test specifically for the proposed linear mixed-effects regression model with random intercepts and random slopes. The score statistic and its asymptotic distribution are derived from the likelihood structure of the proposed model.
In this section, following Cox and Hinkley (1974) and Zamani Mehreyan et al. (2024), the limiting behavior of the score test statistic for regression test is derived. We assume that model (2.3) holds and define the score vector as

$$\boldsymbol{S}_n(\boldsymbol{\theta}) = \frac{\partial Q}{\partial \boldsymbol{\theta}} = \left(\frac{\partial Q}{\partial \boldsymbol{\Gamma}}, \frac{\partial Q}{\partial \boldsymbol{\Sigma}_R}, \frac{\partial Q}{\partial \boldsymbol{\Sigma}_u}\right)^T,$$

and the Hessian matrix as

$$\boldsymbol{F}_n(\boldsymbol{\theta}) = -\frac{\partial^2 Q}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T},$$

where $\boldsymbol{\theta} = (\boldsymbol{\Gamma}, \boldsymbol{\Sigma}_R, \boldsymbol{\Sigma}_u)^T$. According to the model specification in Section 2, $\boldsymbol{\Gamma} = \begin{bmatrix} \gamma_{00} \\ \gamma_{10} \end{bmatrix}$ is a common 2×1 fixed-effects vector for all levels $j = 1, \ldots, m$. The level-specific effects are represented by the random-effects vectors $\mathbf{U}_j = (u_{0j}, u_{1j})^T$. Therefore, although the complete random-effects vector $\mathbf{U}$ has dimension $2m$, the fixed-effects vector $\boldsymbol{\Gamma}$ has dimension two. The components of $\boldsymbol{F}_n(\boldsymbol{\theta})$ corresponding to the fixed-effects vector are

$$\boldsymbol{F}_{\boldsymbol{\Gamma}\boldsymbol{\Gamma}}(\boldsymbol{\theta}) = -\frac{\partial^2 Q}{\partial \boldsymbol{\Gamma} \partial \boldsymbol{\Gamma}^T} = \sum_{i=1}^{n} \mathbf{X}_i^T \boldsymbol{\Sigma}_R^{-1} \mathbf{X}_i.$$

Here, $\mathbf{X}_i$ is the $m \times 2$ fixed-effects design matrix defined in Section 2. The remaining blocks of the Hessian matrix involve the covariance parameters $\boldsymbol{\Sigma}_R$ and $\boldsymbol{\Sigma}_u$. For the derivation below, we partition the Hessian matrix as

$$\boldsymbol{F}_n(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{F}_{\boldsymbol{\Gamma}\boldsymbol{\Gamma}}(\boldsymbol{\theta}) & \boldsymbol{F}_{\boldsymbol{\Gamma}\boldsymbol{\Sigma}}(\boldsymbol{\theta}) \\ \boldsymbol{F}_{\boldsymbol{\Sigma}\boldsymbol{\Gamma}}(\boldsymbol{\theta}) & \boldsymbol{F}_{\boldsymbol{\Sigma}\boldsymbol{\Sigma}}(\boldsymbol{\theta}) \end{bmatrix}, \qquad \boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_R, \boldsymbol{\Sigma}_u)^T.$$

This block representation avoids assigning an incorrect dimension to the fixed-effects block. In particular, $\boldsymbol{F}_{\boldsymbol{\Gamma}\boldsymbol{\Gamma}}$ is a 2×2 matrix. We derive the limiting distribution of the score test statistic for the regression test.

Theorem 3.1. *(Regression test). Suppose* $H_0 : \boldsymbol{\Gamma} = \mathbf{0}$ *holds. Then*

$$SC = \left\{ \boldsymbol{S}_{\boldsymbol{\Gamma}}^T(\boldsymbol{\theta}) \boldsymbol{F}^{\boldsymbol{\Gamma}\boldsymbol{\Gamma}}(\boldsymbol{\theta}) \boldsymbol{S}_{\boldsymbol{\Gamma}}(\boldsymbol{\theta}) \right\}_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0} \xrightarrow{d} \chi^2(2), \qquad \text{as } n \to \infty,$$

where $\boldsymbol{F}^{\boldsymbol{\Gamma}\boldsymbol{\Gamma}}(\boldsymbol{\theta})$ *denotes the* 2×2 *block of* $\boldsymbol{F}_n^{-1}(\boldsymbol{\theta})$ *corresponding to the two-dimensional fixed-effects vector* $\boldsymbol{\Gamma}$.

Proof. Suppose that $\hat{\boldsymbol{\theta}} = \left(\hat{\boldsymbol{\Gamma}}, \hat{\boldsymbol{\Sigma}}_R, \hat{\boldsymbol{\Sigma}}_u\right)^T$ is the EM algorithm estimator of $\boldsymbol{\theta}$, and $\hat{\boldsymbol{\theta}}_0 = \left(\hat{\boldsymbol{\Gamma}}_0, \hat{\boldsymbol{\Sigma}}_{R,0}, \hat{\boldsymbol{\Sigma}}_{u,0}\right)^T$ is the EM algorithm estimator of $\boldsymbol{\theta}$ under H_0. We define

$$\Delta\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_0,$$

and

$$\boldsymbol{S}_n(\boldsymbol{\theta}) = \frac{\partial Q}{\partial \boldsymbol{\theta}} = \left(\boldsymbol{S}_{\boldsymbol{\Gamma}}(\boldsymbol{\theta}), \boldsymbol{S}_{\boldsymbol{\Sigma}_R}(\boldsymbol{\theta}), \boldsymbol{S}_{\boldsymbol{\Sigma}_u}(\boldsymbol{\theta})\right)^T = \left(\boldsymbol{S}_{\boldsymbol{\Gamma}}(\boldsymbol{\theta}), \boldsymbol{S}_{\boldsymbol{\Sigma}}(\boldsymbol{\theta})\right)^T,$$

where

$$\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_R, \boldsymbol{\Sigma}_u)^T.$$

Using the Taylor expansion of $\boldsymbol{S}_n(\hat{\boldsymbol{\theta}})$ around $\hat{\boldsymbol{\theta}}_0$, and the definition

$$\boldsymbol{F}_n(\boldsymbol{\theta}) = -\frac{\partial^2 Q}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T},$$

we have

$$\boldsymbol{S}_n(\hat{\boldsymbol{\theta}}) = \boldsymbol{S}_n(\hat{\boldsymbol{\theta}}_0) - \boldsymbol{F}_n(\hat{\boldsymbol{\theta}}_0)\Delta\hat{\boldsymbol{\theta}} + O_p(1).$$

Since $\boldsymbol{S}_n(\hat{\boldsymbol{\theta}}) = \mathbf{0}$ and $\boldsymbol{S}_{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\theta}}_0) = \mathbf{0}$, we obtain

$$\begin{bmatrix} \boldsymbol{S}_{\boldsymbol{\Gamma}}(\hat{\boldsymbol{\theta}}_0) \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{F}_{\boldsymbol{\Gamma\Gamma}}(\hat{\boldsymbol{\theta}}_0) & \boldsymbol{F}_{\boldsymbol{\Gamma\Sigma}}(\hat{\boldsymbol{\theta}}_0) \\ \boldsymbol{F}_{\boldsymbol{\Sigma\Gamma}}(\hat{\boldsymbol{\theta}}_0) & \boldsymbol{F}_{\boldsymbol{\Sigma\Sigma}}(\hat{\boldsymbol{\theta}}_0) \end{bmatrix} \begin{bmatrix} \Delta\hat{\boldsymbol{\Gamma}} \\ \Delta\hat{\boldsymbol{\Sigma}} \end{bmatrix} + O_p(1).$$

Thus, we have

$$\boldsymbol{S}_{\boldsymbol{\Gamma}}\left(\hat{\boldsymbol{\theta}}_0\right) = \boldsymbol{F}_{\boldsymbol{\Gamma\Gamma}}\left(\hat{\boldsymbol{\theta}}_0\right)\Delta\hat{\boldsymbol{\Gamma}} + \boldsymbol{F}_{\boldsymbol{\Gamma\Sigma}}\left(\hat{\boldsymbol{\theta}}_0\right)\Delta\hat{\boldsymbol{\Sigma}} + O_p(1),$$

and

$$\mathbf{0} = \boldsymbol{F}_{\boldsymbol{\Sigma\Gamma}}\left(\hat{\boldsymbol{\theta}}_0\right)\Delta\hat{\boldsymbol{\Gamma}} + \boldsymbol{F}_{\boldsymbol{\Sigma\Sigma}}\left(\hat{\boldsymbol{\theta}}_0\right)\Delta\hat{\boldsymbol{\Sigma}} + O_p(1).$$

From the second relation, we obtain

$$\Delta\hat{\boldsymbol{\Sigma}} = -\boldsymbol{F}_{\boldsymbol{\Sigma\Sigma}}^{-1}\left(\hat{\boldsymbol{\theta}}_0\right)\boldsymbol{F}_{\boldsymbol{\Sigma\Gamma}}\left(\hat{\boldsymbol{\theta}}_0\right)\Delta\hat{\boldsymbol{\Gamma}} + O_p(n^{-1/2}).$$

Substituting this expression into the first relation gives

$$\boldsymbol{S}_{\boldsymbol{\Gamma}}\left(\hat{\boldsymbol{\theta}}_0\right) = \left[\boldsymbol{F}_{\boldsymbol{\Gamma\Gamma}}\left(\hat{\boldsymbol{\theta}}_0\right) - \boldsymbol{F}_{\boldsymbol{\Gamma\Sigma}}\left(\hat{\boldsymbol{\theta}}_0\right)\boldsymbol{F}_{\boldsymbol{\Sigma\Sigma}}^{-1}\left(\hat{\boldsymbol{\theta}}_0\right)\boldsymbol{F}_{\boldsymbol{\Sigma\Gamma}}\left(\hat{\boldsymbol{\theta}}_0\right)\right]\Delta\hat{\boldsymbol{\Gamma}} + O_p(1).$$

Define

$$\boldsymbol{F}_{\boldsymbol{\Gamma}\cdot\boldsymbol{\Sigma}}(\boldsymbol{\theta}) = \boldsymbol{F}_{\boldsymbol{\Gamma\Gamma}}(\boldsymbol{\theta}) - \boldsymbol{F}_{\boldsymbol{\Gamma\Sigma}}(\boldsymbol{\theta})\boldsymbol{F}_{\boldsymbol{\Sigma\Sigma}}^{-1}(\boldsymbol{\theta})\boldsymbol{F}_{\boldsymbol{\Sigma\Gamma}}(\boldsymbol{\theta}),$$

which is the efficient information matrix for the fixed-effects vector $\boldsymbol{\Gamma}$ after accounting for the covariance parameters. By the block inverse formula,

$$\boldsymbol{F}^{\boldsymbol{\Gamma\Gamma}}(\boldsymbol{\theta}) = \left[\boldsymbol{F}_{\boldsymbol{\Gamma}\cdot\boldsymbol{\Sigma}}(\boldsymbol{\theta})\right]^{-1}.$$

Therefore,

$$\boldsymbol{S}_{\boldsymbol{\Gamma}}\left(\hat{\boldsymbol{\theta}}_0\right) = \left[\boldsymbol{F}^{\boldsymbol{\Gamma\Gamma}}\left(\hat{\boldsymbol{\theta}}_0\right)\right]^{-1}\Delta\hat{\boldsymbol{\Gamma}} + O_p(1).$$

Because $\boldsymbol{\Gamma}$ is a two-dimensional parameter vector, the difference $\Delta\hat{\boldsymbol{\Gamma}}$ is also two-dimensional. Under H_0, the asymptotic distribution of the estimator is therefore

$$\sqrt{n}\,\Delta\hat{\boldsymbol{\Gamma}} \xrightarrow{d} N_2\left(\mathbf{0}, \boldsymbol{I}_{\boldsymbol{\Gamma}\cdot\boldsymbol{\Sigma}}^{-1}\right),$$

where

$$\boldsymbol{I}_{\boldsymbol{\Gamma}\cdot\boldsymbol{\Sigma}} = \lim_{n\to\infty}\frac{1}{n}\boldsymbol{F}_{\boldsymbol{\Gamma}\cdot\boldsymbol{\Sigma}}(\hat{\boldsymbol{\theta}}_0)$$

denotes the limiting efficient information matrix for the two fixed effects. Therefore, the score test statistic is

$$\begin{aligned} SC &= \left\{\boldsymbol{S}_{\boldsymbol{\Gamma}}^T(\boldsymbol{\theta})\boldsymbol{F}^{\boldsymbol{\Gamma\Gamma}}(\boldsymbol{\theta})\boldsymbol{S}_{\boldsymbol{\Gamma}}(\boldsymbol{\theta})\right\}_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0} \\ &= \left\{\Delta\hat{\boldsymbol{\Gamma}}^T\left[\boldsymbol{F}^{\boldsymbol{\Gamma\Gamma}}(\boldsymbol{\theta})\right]^{-1}\Delta\hat{\boldsymbol{\Gamma}}\right\}_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0} + o_p(1). \end{aligned} \tag{3.1}$$

Using

$$\left[\boldsymbol{F}^{\boldsymbol{\Gamma\Gamma}}(\hat{\boldsymbol{\theta}}_0)\right]^{-1} = \boldsymbol{F}_{\boldsymbol{\Gamma\cdot\Sigma}}(\hat{\boldsymbol{\theta}}_0),$$

Equation (3.1) can be written as

$$\begin{aligned} SC &= \Delta\hat{\boldsymbol{\Gamma}}^T \boldsymbol{F}_{\boldsymbol{\Gamma\cdot\Sigma}}(\hat{\boldsymbol{\theta}}_0)\Delta\hat{\boldsymbol{\Gamma}} + o_p(1) \\ &= \sqrt{n}\Delta\hat{\boldsymbol{\Gamma}}^T \left[\frac{1}{n}\boldsymbol{F}_{\boldsymbol{\Gamma\cdot\Sigma}}(\hat{\boldsymbol{\theta}}_0)\right] \sqrt{n}\Delta\hat{\boldsymbol{\Gamma}} + o_p(1) \\ &\xrightarrow{d} \chi^2(2). \end{aligned} \tag{3.2}$$

Equation (3.2) is valid under H_0 and follows from the asymptotic normality of the two-dimensional estimator $\hat{\boldsymbol{\Gamma}}$ and the convergence of the normalized efficient information matrix. Thus, the score test for

$$H_0 : \boldsymbol{\Gamma} = \boldsymbol{0}$$

has an asymptotic chi-square distribution with two degrees of freedom. □

4 Simulation study

In this section, we examine by simulation the relative performance of the material presented so far in the paper. In particular, we examine more closely the performance of the proposed algorithm for estimation and also the performance of the proposed score test.

4.1 Part one

We consider

$$y_{ij} = B_{0j} + B_{1j}x_{ij} + R_{ij}, i = 1, \cdots, 200, \quad j = 1, \cdots, 4 \tag{4.1}$$

where R_{ij}'s are independent and identically distributed and are generated from the standard normal distribution $N(0, 1)$. Also,

$$\begin{aligned} B_{0j} &= \gamma_{00} + u_{0j}, \\ B_{1j} &= \gamma_{10} + u_{1j} \end{aligned}$$

where $\boldsymbol{\Gamma} = \begin{pmatrix} \gamma_{00} \\ \gamma_{10} \end{pmatrix} = \begin{pmatrix} 2 \\ 0.9 \end{pmatrix}$, $\mathbf{U}_j = \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma}_u\right)$ and $\boldsymbol{\Sigma}_u = \begin{pmatrix} 1 & 0.6 \\ 0.6 & 1 \end{pmatrix}$. The observations are generated from model (4.1), where x_{ij}'s are generated from $U(0, 1)$. We generate $R = 10^4$ independent samples $\left\{y_{ij}^{(r)}; i = 1, \cdots, 200, \quad j = 1, \cdots, 4, \quad r = 1, \cdots, R\right\}$. For each iteration, we first consider (y_i, x_i) as a dataset with sample size of 800 and fit a simple linear regression model

$$y_i = B_0 + B_1 x_i + R_i, i = 1, \cdots, 800, \tag{4.2}$$

where R_i's follow $N(\mu, \sigma^2)$. The unknown parameters of the model have been estimated using the maximum likelihood method. The mean of the estimated values, coefficient of determination (R^2), adjusted coefficient of determination (R^2_{adj}), p-value of Shapiro-Wilk test ($P-value$), Akaike information criterion (AIC) and mean squared error of prediction (MSE) obtained from all iterations are presented in Table 1. It shows that, the value of the adjusted coefficient of determination of the estimated model is too low, and the value of MSE of the model is also large. In other words, this model is not suitable for fitting this type of data. The scatter plot of the data and the fitted regression line are given in Figure 1, which confirms the obtained results.

Table 1: The maximum likelihood estimators of simple linear regression model (4.2).

$\hat{B}_0$	$\hat{B}_1$	μ	σ^1	R^2	R^2_{adj}	AIC	MSE	$P-value$
1.9717	0.8988	-1.62e-16	3.5039	0.1022	0.1011	3170.6710	3.4995	0.0823

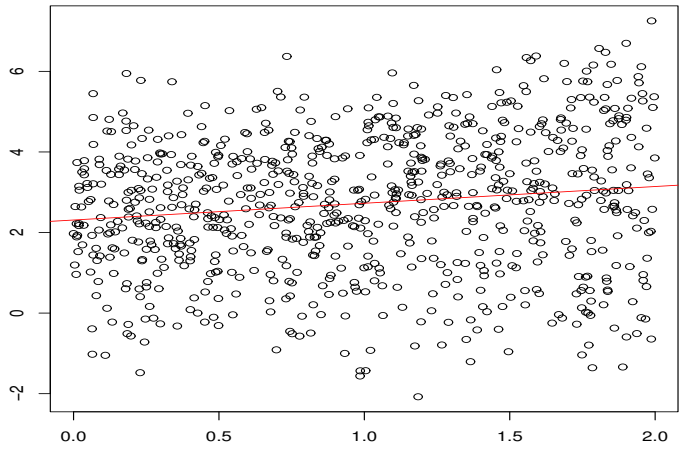


Figure 1: The scatter plot of the data and the fitted regression line.

We consider model

$$y_{ij} = B_{0j} + B_{1j}x_{ij} + R_{ij}, i = 1, \cdots, 200, \quad j = 1, \cdots, 4 \tag{4.3}$$

where R_{ij}'s follow $N(\mu, \sigma^2)$. The previous statistics are also calculated and the results are presented in Table 2. The results of Table 2 show that the values of MSE and AIC of the estimated model (4.3) are lower than the estimated model (4.2). Also, the p-value of Shapiro-Wilk test for model (4.3) is larger than that for model (4.2). Although the model (4.3) is more reliable when models (4.2) and (4.3) are proposed as competing models, the values of the R^2 and R^2_{adj} of the model are also small. Also, we consider the

Table 2: The maximum likelihood estimators of model (4.3).

	$\hat{B}_0$	$\hat{B}_1$	μ	σ^2	R^2	R^2_{adj}	*AIC*	*MSE*	$P-value$
1	2.0110	0.8829	3.27e-18	0.9979	0.2753	0.2717	571.1497	0.9929	0.5117
2	2.0194	0.9114	4.25e-19	1.0040	0.2828	0.2792	572.3661	0.9990	0.5084
3	1.9883	0.8944	-9.12e-19	0.9982	0.2724	0.2687	571.2709	0.9932	0.4980
4	1.8836	0.8882	3.02e-18	0.9911	0.2611	0.2574	569.7010	0.9862	0.5052

random intercept and slope model

$$y_{ij} = B_{0j} + B_{1j}x_{ij} + R_{ij}, i = 1, \cdots, 40, \quad j = 1, \cdots, 4 \tag{4.4}$$

as competing model, where R_{ij}'s are independent and identically distributed as normal distribution $N(\mu, \sigma^2)$ and

$$B_{0j} = \gamma_{00} + u_{0j},$$
$$B_{1j} = \gamma_{10} + u_{1j}$$

where $\mathbf{U}_j = \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{\Sigma}_u\right)$. The model is estimated based on the proposed EM algorithm. For the EM algorithm, the random initialization was employed. The maximum number of iterations was set to 500 (for 10^4 replications), with an absolute convergence tolerance of 10^{-4}. For each simulation setting, 10^4 replications were performed. The number of iterations varied across replications, while the EM procedure successfully produced parameter estimates in all replications, resulting in an empirical convergence failure rate of 0%. As an illustration, one representative run converged at iteration 161, substantially below the maximum allowed number of iterations. The same initialization and convergence settings were maintained across all simulation scenarios.

The results of parameter estimation are given in Table 3. The results of Table 3 show that, the estimated random intercept and slope model based on the proposed algorithm is selected as optimal model based on the R^2, R^2_{adj}, *MSE* and *AIC*. The scatter plot of the data and the estimated model (4.4) are given in Figure 2.

Table 3: The results of parameter estimation of model (4.4).

	$\hat{\gamma}_{00}$	$\hat{\gamma}_{01}$	μ	σ^2	μ_u	Σ_u	R^2	R^2_{adj}	*AIC*	*MSE*	$P-value$
1	1.9940	0.9076	3.71e-17	0.9981	$\begin{pmatrix} -0.0187 \\ -0.0089 \end{pmatrix}$	$\begin{pmatrix} 0.9642 & 0.5705 \\ 0.5705 & 0.9271 \end{pmatrix}$	0.9193	0.9189	571.1422	0.9904	0.5118
2	2.0020	0.8912	2.21e-16	1.0041	$\begin{pmatrix} -0.0187 \\ -0.0089 \end{pmatrix}$	$\begin{pmatrix} 0.9642 & 0.5705 \\ 0.5705 & 0.9271 \end{pmatrix}$	0.9190	0.9186	572.3486	0.9988	0.5080
3	1.9864	0.9104	1.36e-17	0.9985	$\begin{pmatrix} -0.0187 \\ -0.0089 \end{pmatrix}$	$\begin{pmatrix} 0.9642 & 0.5705 \\ 0.5705 & 0.9271 \end{pmatrix}$	0.9194	0.9190	571.2634	0.9926	0.4981
4	1.9948	0.9034	5.03e-18	0.9912	$\begin{pmatrix} -0.0187 \\ -0.0089 \end{pmatrix}$	$\begin{pmatrix} 0.9642 & 0.5705 \\ 0.5705 & 0.9271 \end{pmatrix}$	0.9199	0.9195	569.7005	0.9860	0.5053

For theoretical results of score test, using obtained data and proposed algorithm, the

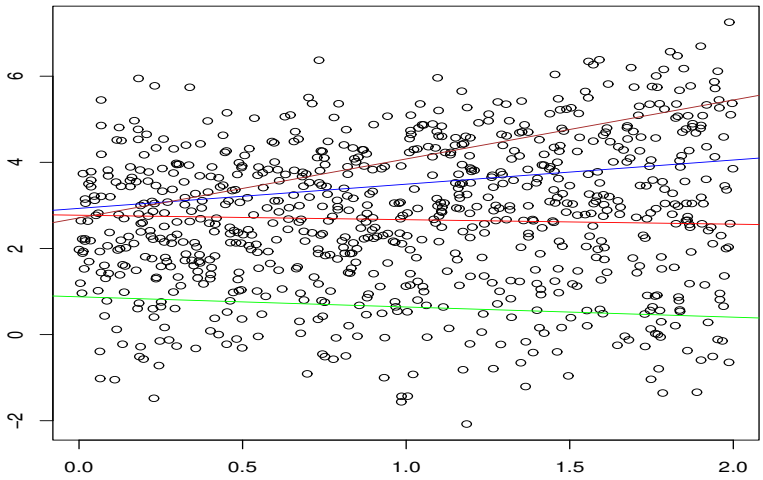


Figure 2: The scatter plot of the data and the estimated model (4.4).

values of score test statistic are computed and compared with $\chi^2_{0.975}(2)$. The relative frequency of proposed test for each of rejection-acceptance regions is computed and the results are summarized in Table 4. It shows that, based on the proposed test, hypothesis $H_0 : \mathbf{\Gamma} = \mathbf{0}$ is rejected, in other words, the estimated model (4.4) is significant. Also, the result shows that the value of the relative frequency is 1.0000 which confirms that the proposed test correctly detects the significance of the model.

Table 4: The relative frequency of score test.

Accept H_0	Reject H_0
0.0000	1.0000

4.2 Part two

Similar to the simulation of part 1, the data are generated from model (4.1) where

$$\mathbf{U}_j = \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{\Sigma}_u\right)$$

and $\mathbf{\Sigma}_u = \begin{pmatrix} 9 & 2 \\ 2 & 4 \end{pmatrix}$. We consider models (4.3) and (4.4) as competing models and estimate competing models based on the maximum likelihood method and proposed EM algorithm, respectively. The statistics presented in Part One are also calculated and the results are presented in Table 5. The results of Table 5 show that, the estimated

Table 5: The results of parameter estimation of models (4.3) and (4.4).

	$\hat{B}_0$	$\hat{B}_1$	μ	σ^2	μ_u	Σ_u	R^2	R^2_{adj}	AIC	MSE	$P-value$
						model (4.3)					
1	0.5653	1.8889	3.12e-17	0.9971	-	-	0.5434	0.5411	571.7503	0.9921	0.3731
2	2.1194	2.0721	1.96e-17	0.9306	-	-	0.4419	0.4391	556.9993	0.9259	0.4523
3	1.0887	0.6233	1.57e-17	1.0215	-	-	0.2897	0.2861	576.4454	1.0164	0.6048
4	0.7968	-0.3404	1.78e-19	0.9725	-	-	0.4029	0.3999	566.1941	0.9677	0.4465
	$\hat{\gamma}_{00}$	$\hat{\gamma}_{01}$	μ	σ^2	μ_u	Σ_u	R^2	R^2_{adj}	AIC	MSE	$P-value$
						model (4.4)					
1	1.9620	0.9264	1.20e-16	0.9971	$\begin{pmatrix} -0.8372 \\ 0.1330 \end{pmatrix}$	$\begin{pmatrix} 8.9583 & 2.2343 \\ 2.2343 & 4.4379 \end{pmatrix}$	0.9729	0.9727	570.4928	0.4124	0.6801
2	1.9697	0.9402	-7.83e-16	0.9306	$\begin{pmatrix} -0.8372 \\ 0.1330 \end{pmatrix}$	$\begin{pmatrix} 8.9583 & 2.2343 \\ 2.2343 & 4.4379 \end{pmatrix}$	0.9743	0.9742	556.3019	0.4765	0.6539
3	2.0663	0.8906	-1.79e-16	1.0215	$\begin{pmatrix} -0.8372 \\ 0.1330 \end{pmatrix}$	$\begin{pmatrix} 8.9583 & 2.2343 \\ 2.2343 & 4.4379 \end{pmatrix}$	0.9726	0.9725	574.8249	0.4185	0.7240
4	1.9212	0.9545	8.85e-16	0.9725	$\begin{pmatrix} -0.8372 \\ 0.1330 \end{pmatrix}$	$\begin{pmatrix} 8.9583 & 2.2343 \\ 2.2343 & 4.4379 \end{pmatrix}$	0.9726	0.9724	565.0766	0.3801	0.6458

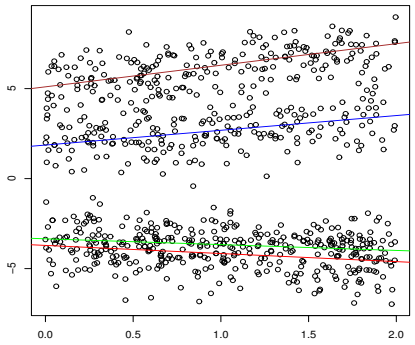


Figure 3: The scatter plot of the data and the estimated competing models.

random intercept and slope model base on the proposed algorithm is selected as the optimal model based on the R^2, R^2_{adj}, MSE and AIC. It can also be seen that for large values of Σ_u, model (4.3) does not perform well in model estimation. Figure 3 confirms these results. The relative frequency of proposed test for each of rejection-acceptance regions is computed. Similar to the simulation of part 1, the value of the relative frequency is 1.0000 which confirms that the proposed test correctly detects the significance of the model.

4.3 Part three

The observations are generated from model (4.1), where R_{ij}'s follow the standard normal distribution $N(0,1)$, $\mathbf{\Gamma} = \begin{pmatrix} \gamma_{00} \\ \gamma_{10} \end{pmatrix} = \begin{pmatrix} 0.01 \\ 0.006 \end{pmatrix}$, $\mathbf{U}_j = \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N_3\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{\Sigma}_u\right)$ and $\mathbf{\Sigma}_u = \begin{pmatrix} 9 & 2 \\ 2 & 4 \end{pmatrix}$. We consider model (4.3) and (4.4) as competing models and estimate competing models based on the maximum likelihood method and proposed EM algorithm, respectively. The previous statistics are also calculated and the results are

Table 6: The results of parameter estimation of competing models.

	$\hat{B}_0$	$\hat{B}_1$	μ	σ^2	μ_u	Σ_u	R^2	R^2_{adj}	AIC	MSE	$P-value$
						model (4.3)					
1	-0.0905	0.1786	2.17e-18	1.0035	-	-	0.4196	0.4166	572.5800	0.9985	0.5281
2	-0.1130	0.1648	-4.47e-18	0.9875	-	-	0.3807	0.3776	569.1881	0.9825	0.5478
3	-0.1542	-0.2003	2.73e-18	0.9865	-	-	0.3791	0.3760	568.9382	0.9816	0.4975
4	0.0284	-0.0110	5.87e-19	1.0017	-	-	0.4195	0.4166	571.9121	0.9967	0.4965
	$\hat{\gamma}_{00}$	$\hat{\gamma}_{01}$	μ	σ^2	μ_u	Σ_u	R^2	R^2_{adj}	AIC	MSE	$P-value$
						model (4.4)					
1	0.0147	0.0612	4.80e-16	1.0035	$\begin{pmatrix} -0.0888 \\ -0.0325 \end{pmatrix}$	$\begin{pmatrix} 8.2771 & 2.1220 \\ 2.1220 & 4.2768 \end{pmatrix}$	0.9726	0.9724	572.5825	0.9985	0.5281
2	0.0066	0.0689	-2.00e-16	0.9875	$\begin{pmatrix} -0.0888 \\ -0.0325 \end{pmatrix}$	$\begin{pmatrix} 8.2771 & 2.1220 \\ 2.1220 & 4.2768 \end{pmatrix}$	0.9730	0.9728	569.1906	0.9825	0.5478
3	-0.0154	0.0816	-4.01e-17	0.9865	$\begin{pmatrix} -0.0888 \\ -0.0325 \end{pmatrix}$	$\begin{pmatrix} 8.2771 & 2.1220 \\ 2.1220 & 4.2768 \end{pmatrix}$	0.9734	0.9732	568.9407	0.9816	0.4975
4	0.0200	0.0501	1.09e-16	1.0017	$\begin{pmatrix} -0.0888 \\ -0.0325 \end{pmatrix}$	$\begin{pmatrix} 8.2771 & 2.1220 \\ 2.1220 & 4.2768 \end{pmatrix}$	0.9727	0.9726	571.9146	0.9967	0.4965

Table 7: The relative frequency of score test.

Accept H_0	Reject H_0
0.9917	0.0083

presented in Table 6. The results of Table 6 show that, the estimated random intercept and slope model based on the proposed algorithm is selected as optimal model based on the R^2, R^2_{adj}, MSE and AIC. The relative frequency of proposed test for each of rejection-acceptance regions is computed and the results are summarized in Table 7. It shows that, based on the proposed test, Hypothesis $H_0 : \mathbf{\Gamma} = \mathbf{0}$ is accepted. The value of the relative frequency is 0.9917 which confirms the proposed test correctly identified that the model is not significant.

4.4 Part four: type I error analysis

To further investigate the finite-sample performance of the proposed score test, an additional simulation study is conducted to evaluate its empirical type I error rate under the true null hypothesis. Similar to the simulation settings considered in the previous parts, the data are generated from model (4.1). Under the null hypothesis, the common fixed-effects vector is specified as

$$H_0 : \mathbf{\Gamma} = \begin{bmatrix} \gamma_{00} \\ \gamma_{10} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

The random-effects vectors are generated according to

$$\mathbf{U}_j = \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{\Sigma}_u \right),$$

where

$$\mathbf{\Sigma}_u = \begin{pmatrix} 9 & 2 \\ 2 & 4 \end{pmatrix}.$$

Table 8: The relative frequency of score test.

Accept H_0	Reject H_0	Empirical Type I Error
0.9845	0.0155	0.0155

The remaining model parameters and simulation settings are kept the same as those used in the previous simulation study. For each simulated data set, the proposed score test is performed for testing

$$H_0 : \mathbf{\Gamma} = \mathbf{0} \quad \text{versus} \quad H_1 : \mathbf{\Gamma} \neq \mathbf{0}.$$

The test is conducted at the nominal significance level of $\alpha = 0.05$.
The empirical type I error rate is calculated as the relative frequency with which the null hypothesis is rejected over the simulation replications. Specifically, if R denotes the total number of simulation replications, the empirical type I error rate is calculated as

$$\widehat{\alpha} = \frac{1}{R} \sum_{s=1}^{R} I\left(SC_s > \chi^2_{2,0.95}\right),$$

where SC_s denotes the score test statistic obtained from the sth simulated data set.
The results are reported in Table 8. These results show that the null hypothesis is accepted in 98.45% of the simulation replications, whereas it is rejected in 1.55% of the replications. Since the data are generated under the true null hypothesis, the empirical type I error rate of the proposed score test is 0.0155. Thus, for this simulation configuration, the empirical type I error rate is below the nominal significance level of 0.05, indicating that the proposed test does not exhibit an inflated type I error rate under the null hypothesis.

5 Real data analysis

We consider the income tax payments data consisting of information on 258 taxpayers in the United States, collected during 1982 and 1986–1987. These data have been previously analyzed to investigate the effects of individual and economic characteristics on tax payments (Frees, 2004; Rabe-Hesketh and Skrondal, 2008). Each taxpayer contributes five repeated observations, resulting in a total of 1290 observations. Thus, the data have a natural two-level structure, with the five observations nested within each of the 258 taxpayers. The dataset includes one response variable, four explanatory variables of individual characteristics and four explanatory variables of economic characteristics. The response variable is LNTAX, the logarithm of the total positive tax liability as stated on the tax return in 1983 dollars. The eight candidate explanatory variables considered in the variable-selection stage are MS, HH, DEPEND, AGE, MR, EMP, PREP, and LNTPI. The variables MS, HH, AGE, EMP, and PREP are binary indicator variables, while DEPEND represents the number of dependents. The economic variables LNTPI and MR represent the logarithm of total positive income and the

marginal tax rate measure, respectively. The data were analyzed in their original form and were organized according to taxpayer and time. No missing observations were present in the dataset, and no observations were removed or modified because of outlying values. No additional standardization or rescaling of the explanatory variables was performed. The logarithmic variables LNTAX and LNTPI were used as provided in the original dataset. The analysis was conducted in two stages. In the first stage, all eight candidate explanatory variables were considered in the variable-selection procedures. In the second stage, the variables selected by the variable-selection procedures were used to specify the multilevel model. In particular, LNTPI and MR were retained as explanatory variables in the final random-intercept and random-slope model. Frees (2004) and Rabe-Hesketh and Skrondal (2008) considered model

$$\begin{aligned} y_{ij} = \beta_0 + \beta_1 MS_{ij} + \beta_2 HH_{ij} + \beta_3 DEPEND_{ij} + \beta_4 AGE_{ij} + \beta_5 MR_{ij} \\ + \beta_6 EMP_{ij} + \beta_7 PREP_{ij} + \beta_8 LNTPI_{ij} + \epsilon_{ij}, \end{aligned} \tag{5.1}$$

where $j = 1, 2, \cdots, 258$, $i = 1, 2, \cdots, 5$ and ϵ_{ij}'s follow normal distribution. In this section, the data were analyzed in two steps. First, suitable variables are selected using different methods and we compare the performance of these methods in variable selection. In the second step, we fit the random intercept and slope model using the proposed algorithm and compare the performance of this model with the models of the first step.

5.1 Variable selection

We consider multiple linear regression model

$$y_l = \beta_0 + \beta_1 x_{1,l} + \cdots + \beta_k x_{k,l} + \epsilon_l, l = 1, 2, \cdots, N,$$

where $N = n \times m = 258 \times 5 = 1290$, ϵ_l's are independent and identically distributed as normal distribution, $N(\mu, \sigma^2)$. The correlation between Y and X and the p-value of their correlation test,

$$H_0 : \rho = 0$$

are presented in Table 9, where $X = (MS, HH, DEPEND, AGE, MR, EMP, PREP, LNTPI)^T$. These results show that all p-values are less than 0.05, indicating that the response variable Y is significantly associated with all explanatory variables in X. Also, variable Y has the highest correlation with variables *MR* and *LNTPI*.

Table 9: The correlation between Y and X and the p-value of correlation test.

	MS	HH	DEPEND	AGE	MR	EMP	PREP	LNTPI
ρ	0.2621	-0.1602	0.0851	-0.0604	0.7466	-0.0908	0.0991	0.7176
p-value	$< 2.2e-16$	7.131e-09	0.0021	0.0298	$< 2.2e-16$	0.0010	0.0003	$< 2.2e-16$

The results of the estimated full model using the likelihood method and presented statistics in Section 4 are given in Table 10. These results show that the estimated

Table 10: The results of the estimated full model.

	Estimate	Std. Error	t-value	$Pr(>\|t\|)$
β_0	-3.4303	0.5647	-6.074	1.64e-09
β_1	0.0417	0.1546	0.270	0.7873
β_2	-0.6969	0.1924	-3.621	0.0003
β_3	-0.1395	0.0496	-2.809	0.0050
β_4	0.0121	0.1562	0.078	0.9379
β_5	0.1077	0.0069	15.502	$< 2e-16$
β_6	-0.6216	0.1351	-4.600	4.65e-06
β_7	-0.0004	0.0976	-0.004	0.9966
β_8	0.8332	0.0715	11.651	$< 2e-16$
μ	-1.15e-16			
σ^2	2.7926			
$R^2 = 0.6155$				
$R^2_{adj} = 0.6131$				
AIC=5004.681				
MSE= 2.7904				

Table 11: The results of the estimated reduced model.

	Estimate	Std. Error	t-value	$Pr(>\|t\|)$
β_0	-3.4675	0.5491	-6.315	$3.72e-10$
β_2	-0.7239	0.1675	-4.320	$1.68e-05$
β_3	-0.1319	0.0370	-3.557	0.0003
β_5	0.1076	0.0068	15.735	$< 2e-16$
β_6	-0.6165	0.1325	-4.653	3.61e-06
β_8	0.8384	0.0693	12.081	$< 2e-16$
μ	1.59e-16			
σ^2	2.7928			
$R^2 = 0.6155$				
$R^2_{adj} = 0.614$				
AIC=4998.774				
MSE= 2.7906				

parameters β_1, β_4 and β_7 are not significant, so we removed variables *MS*, *AGE* and *PREP* from the model and estimated the reduced model using the maximum likelihood method, Table 11. The results of Tables 10 and 11 show that the reduced model has lower Akaike information criterion and higher adjusted coefficient of determination than the full model. The results of variable selection based on the stepwise method are given in Table 12. Variables MR, LNTPI, EMP, HH and DEPEND were selected using a step-by-step method. The results obtained from the step-by-step method confirm the results of Table 11. The variable importance scores obtained from the boosting algorithm are reported in Table 13. The results indicate that MS and HH have the largest importance scores, followed by DEPEND and AGE. These scores quantify the relative importance assigned to the candidate predictors by the boosting algorithm. The fitted model using the boosting algorithm has the lowest *AIC* and *MSE* and the highest R^2 and R^2_{adj} compared to the full model and the reduced model.

Table 12: The results of variable selection based on the step-by-step method.

	Estimate	Std. Error	Lower 2.5%	*Upper*97.5%
β_0	-3.4676	0.5825	-4.6103	-2.3248
β_5	0.1076	0.0073	0.0934	0.1218
β_8	0.8384	0.0736	0.6940	0.9828
β_6	-0.6165	0.1406	-0.8923	-0.3408
β_2	-0.7239	0.1778	-1.0727	-0.3752
β_3	-0.1319	0.0393	-0.2091	-0.0547
μ	3.72e-17			
σ^2	2.7928			
$R^2 = 0.6155$				
$R^2_{adj} = 0.614$				
AIC=4998.774				
MSE= 2.7906				

Table 13: The results of variable importance scores obtained from the boosting algorithm.

β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	μ	σ^2
46.1550	40.0279	4.8867	2.9273	2.6902	1.9939	1.0526	0.2660	-0.0004	0.4116
$R^2 = 0.9433$									
$R^2_{adj} = 0.9429$									
AIC=2530.907									
MSE= 0.4113									

5.2 Random intercept and slope model analysis

We first considered the multiple linear regression model (5.1) for variable selection. Although variables MS, HH, DEPEND, AGE, MR, EMP, PREP and LNTPI were selected based on the boosting algorithm and variables MR, LNTPI, EMP, HH and DEPEND were selected based on the step-by-step method for the multiple linear regression model, but only variables LNTPI and MR were selected for all $j = 1, 2, \cdots, 258$ except $j = 12, 28, 92, 148, 185, 190$ based on the boosting algorithm and the step-by-step method for the model (5.1). So, six taxpayers (IDs 12, 28, 92, 148, 185, and 190) were excluded from the final analysis. Consequently, the final multilevel analysis included 252 taxpayers, each with five repeated observations, resulting in a total of 1260 observations. Following the variable-selection stage, a multilevel linear regression model with a random intercept and random slopes for LNTPI and MR was considered.
Let y_{ij} denote the logarithm of the tax liability for taxpayer j at observation time i, where $j = 1, \ldots, 252$ and $i = 1, \ldots, 5$. The explanatory variables are $LNTPI_{ij}$, representing the logarithm of total positive income, and MR_{ij}, representing the marginal tax rate measure. We considered the following multilevel linear regression model with a random intercept and random slopes:

$$y_{ij} = B_{0j} + B_{1j}LNTPI_{ij} + B_{2j}MR_{ij} + R_{ij}, \tag{5.2}$$

where B_{0j} denotes the taxpayer-specific random intercept, and B_{1j} and B_{2j} denote the taxpayer-specific random slopes associated with *LNTPI* and *MR*, respectively. These subject-specific coefficients are decomposed into fixed and random components as

follows:

$$\begin{aligned} B_{0j} &= \gamma_{00} + u_{0j}, \\ B_{1j} &= \gamma_{10} + u_{1j}, \\ B_{2j} &= \gamma_{20} + u_{2j}. \end{aligned} \tag{5.3}$$

The parameters γ_{00}, γ_{10}, and γ_{20} represent the population-level fixed effects, while u_{0j}, u_{1j}, and u_{2j} represent the taxpayer-specific random effects. Define the random-effects vector as

$$\mathbf{U}_j = \begin{pmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \end{pmatrix} \sim N_3(\mathbf{0}, \mathbf{\Sigma}_u),$$

where $\mathbf{\Sigma}_u$ is the 3×3 covariance matrix of the random effects. The within-taxpayer errors are assumed to satisfy

$$R_{ij} \overset{\text{iid}}{\sim} N(0, \sigma^2),$$

and are assumed to be independent of the random effects. The random-effects vectors are also assumed to be independent across taxpayers. This specification allows both the baseline level of tax liability and the associations of *LNTPI* and *MR* with tax liability to vary across taxpayers.

The parameters of model (5.2) were estimated using both the proposed expectation-maximization (EM) algorithm and the maximum likelihood (ML) method. The individual-level estimates and model-evaluation results for the 252 taxpayers are displayed graphically in Figures 4 and 5, respectively. Displaying all individual-level results in a single table would be impractical; therefore, the figures are used to present these results concisely. Table 14 summarizes the parameter estimates and other model results obtained from fitting the multilevel model using the two estimation procedures. The EM algorithm was implemented according to the theoretical formulation presented in Section 2, with initial values obtained from ML estimates. The ML estimates were obtained through numerical optimization using the `optim` function in R. Thus, the figures present the individual-level results, whereas the table reports the summary results for the fitted multilevel model.

The results reported in Table 14 and Figure 5 show differences between the estimates obtained from the EM-based and numerical ML procedures. In particular, the EM-based procedure yields lower reported AIC and MSE values and higher R^2 and adjusted R^2 values.

The estimated fixed effects and their associated standard errors, confidence intervals, and p-values are reported in Table 14. Under the adopted inferential procedure, all three fixed effects were statistically different from zero at the 0.001 significance level. The estimated coefficient of *LNTPI* was positive, indicating a positive conditional association between log total positive income and log tax liability, after accounting for *MR* and the taxpayer-specific random effects. The estimated coefficient of *MR* was also positive, indicating a positive conditional association between the marginal tax rate measure and log tax liability. Table 14 also reports the estimated residual-scale parameter and

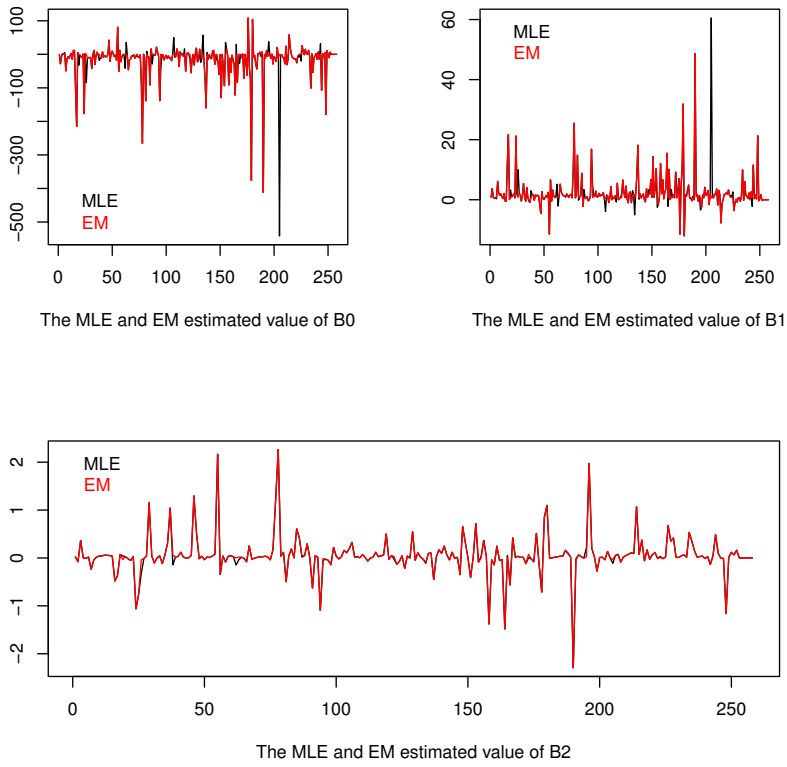


Figure 4: Individual-level parameter estimates obtained for model (5.2) using the considered estimation procedures.

the estimated random-effects mean vector and covariance matrix. To assess the normality assumption of the residuals in the multilevel model, the Shapiro–Wilk normality test was performed, and the residual distribution was further examined using a histogram and a normal quantile–quantile (Q–Q) plot, Figure 6. The Shapiro–Wilk test yielded a test statistic of (W=0.99923) and a p-value of (0.9006). Since the p-value is substantially greater than the conventional significance level of (0.05), the null hypothesis of normality is not rejected. Furthermore, the histogram and the normal Q–Q plot provide complementary graphical assessments of the residual distribution. The histogram exhibits an approximately bell-shaped pattern and the points in the Q–Q plot lie reasonably close to the reference line, these graphical results support the normality assumption. Overall, the results provide no statistically significant evidence against the normality of the residuals.

To assess the multivariate normality assumption for the random effects, the squared Mahalanobis distances of the estimated random-effect vectors were calculated and compared with the theoretical quantiles of a chi-squared distribution with three degrees

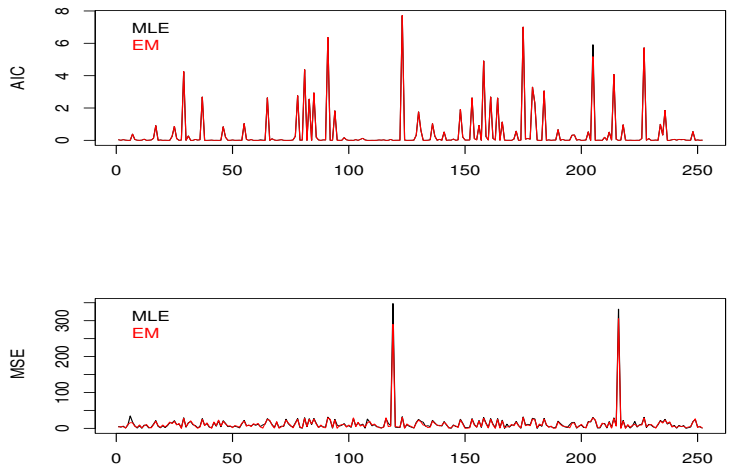


Figure 5: Individual-level AIC and MSE values obtained for model (5.2) using the considered estimation procedures.

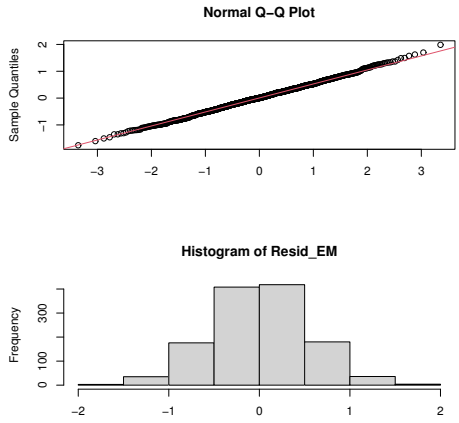


Figure 6: Histogram and normal Q–Q Plot of the model residuals.

of freedom, see Johnson and Wichern (2007). The Q–Q plot (Figure 7) showed that the points were approximately aligned with the reference line, providing graphical support for the multivariate normality assumption of the random effects. The proposed score test statistic for model (5.2) is 60.8657. We compared the value of proposed score test statistic with $\chi^2(0.975, 3) = 9.3484$. Since $60.8657 > 9.3484$, the null hypothesis $H_0 : \mathbf{\Gamma} = \mathbf{0}$ was rejected. This provides evidence that the three population-level fixed effects are not jointly equal to zero and the estimated model is significant.

Table 14: Parameter estimates, standard errors, confidence intervals, and model summary statistics for model (5.2) obtained using the proposed EM algorithm and maximum likelihood (ML) estimation.

Method	Parameter			Estimate	SE	95% CI	p-value
EM	γ_{00}			-16.9426	0.1535	$[-17.2435, -16.6417]$	< 0.001
	γ_{10}			2.3399	0.0018	$[2.3364, 2.3434]$	< 0.001
	γ_{20}			0.0564	0.0014	$[0.0537, 0.0591]$	< 0.001
	$\hat{\mu}$			-6.14×10^{-6}	–	–	–
	$\hat{\sigma}$			0.5367	–	–	–
	$\hat{\boldsymbol{\mu}}_u$			$\begin{pmatrix} 2.46 \times 10^{-14} \\ 2.81 \times 10^{-15} \\ 4.44 \times 10^{-17} \end{pmatrix}$	–	–	–
	$\widehat{\boldsymbol{\Sigma}}_u$	$\begin{pmatrix} 23.1865 \\ 9.72 \times 10^{-12} \\ 2.28 \times 10^{-13} \end{pmatrix}$	$\begin{matrix} 9.72 \times 10^{-12} \\ 2.5712 \\ 2.15 \times 10^{-15} \end{matrix}$	$\begin{pmatrix} 2.28 \times 10^{-13} \\ 2.15 \times 10^{-15} \\ 0.1702 \end{pmatrix}$	–	–	–
ML	γ_{00}			-19.4204	0.4464	$[-20.2953, -18.5455]$	< 0.001
	γ_{10}			2.6135	0.0057	$[2.6023, 2.6247]$	< 0.001
	γ_{20}			0.0544	0.0048	$[0.0450, 0.0638]$	< 0.001
	$\hat{\mu}$			-2.30×10^{-19}	–	–	–
	$\hat{\sigma}$			0.4294	–	–	–
	$\hat{\boldsymbol{\mu}}_u$			$\begin{pmatrix} 4.35 \times 10^{-21} \\ 3.48 \times 10^{-23} \\ 1.56 \times 10^{-26} \end{pmatrix}$	–	–	–
	$\widehat{\boldsymbol{\Sigma}}_u$	$\begin{pmatrix} 37.7548 \\ 1.72 \times 10^{-17} \\ 3.05 \times 10^{-15} \end{pmatrix}$	$\begin{matrix} 1.72 \times 10^{-17} \\ 3.9457 \\ 7.821 \times 10^{-26} \end{matrix}$	$\begin{pmatrix} 3.05 \times 10^{-15} \\ 7.821 \times 10^{-26} \\ 0.5201 \end{pmatrix}$	–	–	–
MSE (EM / ML)				0.4044/0.4156			
AIC (EM / ML)				12.4344/13.2683			
R^2 (EM / ML)				0.9989/0.9045			
R^2_{adj} (EM / ML)				0.9978/0.8110			

Note: SE denotes standard error and CI denotes confidence interval. The confidence intervals and p-values for the fixed effects are based on the inferential procedure used for the fitted model.

6 Conclusion

In this paper, we make a number of contributions to the literature that relates to an extension of the basic multilevel regression model involving random intercept and slope model. First, we estimated the parameters of model based on the EM algorithm. This is important, because the standard methods of variance component estimation used in the multilevel regression model produce a negative or zero estimate. Second, we proposed score test for the regression test of the multilevel regression model. Also, we examined the properties of estimators and score test statistics using simulations which validate our theoretical results. In summary, the theoretical derivations and the simulations support the use of EM algorithm and the use of proposed score test as a useful tool in a multilevel analytical toolbox for empirical analysis of random intercept and slope models.

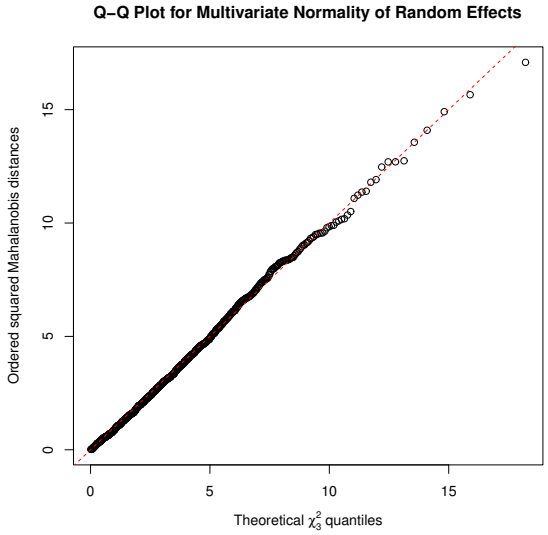


Figure 7: Chi-squared Q–Q plot of the squared Mahalanobis distances for assessing the multivariate normality of the estimated random effects.

References

Aitkin M, Longford N. Statistical modelling in school effectiveness studies (with discussion). Journal of the Royal Statistical Society. 1986;149:1–43.

Barendse MT, Rosseel Y. Multilevel SEM with random slopes in discrete data using the pairwise maximum likelihood. British Journal of Mathematical and Statistical Psychology. 2023;76(2):327–352.

Barr DJ. Random effects structure for confirmatory hypothesis testing: keep it maximal. Journal of Memory and Language. 2013;68:255–278.

Bates MD, Castellano KE, Rabe-Hesketh S, Skrondal A. Handling correlations between covariates and random slopes in multilevel models. Journal of Educational and Behavioral Statistics. 2014;39(6):524–549.

Berkhof J, Kampen JK. Asymptotic effect of misspecification in the random part of the multilevel model. Journal of Educational and Behavioral Statistics. 2004;29:201–218.

Bryan ML, Jenkins SP. Multilevel modelling of country effects: a cautionary tale. European Sociological Review. 2016;32:3–22.

Cox DR, Hinkley DV. Theoretical statistics. London: Chapman and Hall; 1974.

Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society: Series B (Methodological). 1977;39(1):1–38.

Frees EW. Longitudinal and panel data: Analysis and applications in the social sciences. New York: Cambridge University Press; 2004.

Grund S, Lüdtke O, Robitzsch A. Multiple imputation of missing covariate values in multilevel models with random slopes: a cautionary note. Behavior Research Methods. 2016;48:640–649.

Heisig JP, Schaeffer M, Giesecke J. The costs of simplicity: why multilevel models may benefit from accounting for cross-cluster differences in the effects of controls. American Sociological Review. 2017;82:796–827.

Humberg S, Grund S, Nestler S. Estimating nonlinear effects of random slopes: A comparison of multilevel structural equation modeling with a two-step, a single-indicator, and a plausible values approach. Behavior Research Methods. 2024;.

Johnson RA, Wichern DW. Applied multivariate statistical analysis. 6 ed. Pearson Prentice Hall; 2007.

Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. Biometrics. 1997;53(3):983–997.

Kim JS, Frees EW. Multilevel modeling with correlated effects. Psychometrika. 2007;72:505–533.

Laird NM, Ware JH. Random-effects models for longitudinal data. Biometrics. 1982;38(4):963–974.

Longford NT. A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. Biometrika. 1987;74:817–827.

Oberpriller J, Leite M, Pichler M. Fixed or random? On the reliability of mixed-effects models for a small number of levels in grouping variables. Ecology and Evolution. 2022;12(7):e9062.

Rabe-Hesketh S, Skrondal A. Multilevel and longitudinal modeling using Stata. 2 ed. StataCorp LP; 2008.

Raudenbush SW, Bryk AS. A hierarchical model for studying school effects. Sociology of Education. 1986;12:241–269.

Satterthwaite FE. An approximate distribution of estimates of variance components. Biometrics Bulletin. 1946;2(6):110–114.

Schmidt-Catran AW, Fairbrother M. The random effects in multilevel models: getting them wrong and getting them right. European Sociological Review. 2016;32:23–38.

Wald A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. Transactions of the American Mathematical Society. 1943;54(3):426–482.

Wilks SS. The large-sample distribution of the likelihood ratio for testing composite hypotheses. The Annals of Mathematical Statistics. 1938;9(1):60–62.

Wooldridge JM. Fixed-effects and related estimators for correlated random coefficient and treatment-effect panel data models. Review of Economics and Statistics. 2005;51:385–390.

Zamani Mehreyan S, Sharafi M, Nematollahi A. Large sample properties of the score test statistic in nonlinear regression involving stationary and non-stationary errors: an application in age estimation of European wild rabbits from eye lens weight. Journal of Agricultural, Biological and Environmental Statistics. 2024;.

Zhao Y, Edland SD. Power formulas for mixed effects models with random slope and intercept comparing rate of change across groups. Biostat. 2021;18(1):173–182.