

CONSTRAINED INFERENCE IN SCALED MIXED EFFECTS MODELS WITH APPLICATIONS TO LONDON JUNIOR SCHOOL PROJECT DATA

MD MOSTAK AHAMMED

Department of Statistics, Grand Valley State University, Allendale, MI 49401, USA
Email: ahammedm@gvsu.edu

BHASKAR BHATTACHARYA*

School of Mathematical and Statistical Sciences, Southern Illinois University, Carbondale, IL 62901, USA
Email: bhaskar@siu.edu

SUMMARY

Multiple outcomes arise frequently in many different settings. Mixed effect models are a useful tool for the analyses of such data. However, when outcomes are measured on different scales, analyses based on any one scale are misleading. Often parameters of the model are subject to known order restrictions. To incorporate heterogeneity across different outcomes, we propose a *scaled* linear mixed effect model. To estimate parameters, we propose a maximum likelihood estimation procedure based on a restricted version of the expectation-conditional maximization either algorithm. Constrained hypotheses testing procedures are developed using likelihood ratio tests. The empirical significance levels and powers are studied using simulations. This article shows that incorporating the restrictions improves the mean squared errors of the estimates and the power of the tests. The methodology is applied on the London Junior School Project data incorporating known restrictions of patterns of scores.

Keywords and phrases: Chi-bar square; Constrained estimation and testing; ECME algorithms; Mixed effect models.

1 Introduction

The primary goal of educators is to equip students with the knowledge and skills necessary to think critically, solve complex problems, and succeed in today's rapidly changing world. Measurement of such knowledge and skills is essential to tracking students' development and assessing the effectiveness of educational policies and practices. Numerous authors have written on the merits of measuring students' abilities using their verbal/English and mathematics scores on standardized tests (Jacob et al., 2016; Koretz 2008; Rubin et al., 2004). A higher verbal score indicates better fluency, control of both receptive and productive languages, being able to follow analogies, comprehend difficult written material and produce creative writing. A higher mathematics score indicates conceptual understanding, procedural fluency, strategic competence, adaptive reasoning, and productive

* Corresponding author

© Institute of Statistical Research and Training (ISRT), University of Dhaka, Dhaka 1000, Bangladesh.

disposition. It is also well-documented that gender differences exist on verbal and mathematics test scores (Fryer et al., 2010; Lindberg et al., 2010; Cornwell et al., 2013; Quinn et al., 2015). Girls are known to score higher on verbal tests while boys tend to score higher on mathematics tests. Despite the fact that gender gaps in mathematics test scores have been found to narrow or even vanish over recent decades (Hyde et al., 1990; Joensen et al., 2009) they remain present in large-scale assessments such as the Programme for International Student Assessment (PISA) (Guiso et al., 2010). More recently, Balart and Oosterveen (2019) examined the hypothesis that females show more sustained overall performance during test-taking than males and investigated its potential implications for the gender gaps in test scores.

Many studies have found a strong association between the economic outcomes of nations and students' performance on international cognitive tests such as the PISA, Trends in International Mathematics and Science Study (TIMSS) or Progress in International Reading Literacy Study (PIRLS) (see, for example, Hanushek and Kimko, 2000; Hanushek and Woessmann, 2008; Hanushek and Woessmann, 2012). This association may be interpreted as evidence for cognitive skills being an important determinant of social economic growth among perhaps many other noncognitive skills (Wechsler, 1940). Duckworth, et al. (2011) found that under low-stakes testing conditions, such as in the international cognitive tests, some individuals try harder than others. Moreover, scores of low performers can be substantially improved by offering a reward (e.g. Borghans et al., 2008; Gneezy and Rustichini, 2000; Segal 2012). The website of Educational Opportunity Project at edopportunity.org lists many publications connecting educational opportunities and performances of students. All these indicate that students' scores improve with better economic conditions. The better a student is prepared for learning, the more successful he or she is likely to be during the school career. So it is reasonable (Kitsantas, et al., 2008) to assume that students with higher scores on school admission tests are better prepared and later on will generally perform superior on any test that they study for, including, English and mathematics tests. A similar argument is made by Allensworth and Clark (2020) to postulate that better prepared students perform well in future studies.

Therefore, while analyzing the English and mathematics test scores, it is natural to assume that, generally, females would score higher than males. Scores would improve with better household socioeconomic conditions and would also improve with higher school admission test results. In practice, these assumptions lead to imposing constraints on the parameters of the model. Constrained environments arise naturally in many fields. Statistical inference under such constraints are more efficient than their counterparts wherein such constraints are ignored (Silvapulle & Sen, 2005, and other references therein). However, any additional restriction complicates the associated inferential procedures. Here, we develop the methodology for constrained estimation and testing in longitudinal mixed models.

We analyze the Junior School Project data (available as `jsp` in the R package `faraway`) that were collected from primary schools in inner London on scores of English and mathematics tests along with observations on variables including school level, gender, social class, etc. While different outcomes of English and mathematics test scores are affected by the same cognitive and non-cognitive conditions of a student, their effect may be different. The English scores range from 0 to 100 while mathematics scores range from 0 to 40 (see Figure 1). The same data were analyzed

by Mortimer et al. (1989) who studied the influences associated with the home backgrounds of students and examined the effects of schools over and above the influence of home backgrounds. More recently, de Luma (2021) studied effects of 3rd year versus 5th year after combining English and mathematics scores. None of these authors considered heterogeneity of the English and mathematics scores or restrictions on the regression parameters in their analyses.

When outcomes are measured on different scales, then analysis using one selected scale of measurement is misleading. Lin et al. (2004) proposed a scaled linear mixed model (SLMM) for analyzing multiple continuous outcomes which considered the heterogeneity of variances over different outcomes. Correlations among different outcomes within the same subject are accommodated using random effects. Roy et al. (2003) proposed scaled marginal models for multiple outcomes to test for a common exposure effect. Kennedy et al. (2017) proposed scaled effect measures (via potential outcomes) that translate effects on multiple outcomes to a common scale.

There is an extensive literature on statistical inferences under inequality constraints (Silvapulle and Sen, 2005, and references therein) verifying real gains in terms of improved power and/or a better model fit over unrestricted. Singh and Wright (1990) considered order-restricted inference on fixed effects in a two-factor mixed model. They presented an analogue to the usual F-test for homogeneity and obtained several closed-form results. In the absence of any covariates, especially continuous covariates, Mukerjee (1988) noted that the usual tests for order restrictions on the means of independent normal populations can be extended to the case when normal populations are correlated as in a repeated measurements design. Later Silvapulle (1997) generalized this methodology to some unbalanced designs with incomplete data and showed that mixed linear models can be reduced to fixed-effect models where the usual one-sided tests for ordered hypothesis can be applied. When this approach is applicable, the asymptotic null distribution of the test statistic is chi-squared with weights that do not depend on the unknown variance components. Davidov and Rosen (2011) worked on constrained inference of the fixed effect parameter for linear mixed effects model under homoscedastic errors. Farnan et al. (2014) considered constrained inference of the fixed effect parameter under heterogeneity of errors with independent random effects in model. They used an empirical best linear unbiased predictor type residual based bootstrap methodology for deriving critical values of the proposed test.

Constrained estimation using expectation maximization (EM)-type algorithms (Mclachlan and Krishnan, 1997) has been used earlier. Fang and others (2006) proposed a modified EM algorithm for deriving the maximum likelihood (ML) estimator in a multivariate random-effects model which imposed constraints on the intercepts. Kim and Taylor (1995) considered a restricted EM algorithm for ML estimation under equality constraints. Nettleton (1999) and Shi et al. (2005) considered inequality constraints by performing a constrained maximization within the M-step, as was done by Davidov and Rosen (2011).

For ML estimation of fixed effect parameters in SLMM, Lin et al (2004) proposed an algorithm using existing software but it might produce negative estimates of the variances. Motivated by the need to develop a methodology for incorporating the heterogeneity in data and known restrictions on fixed parameters, we have proposed a variation of the SLMM model. We used the expectation-conditional maximization either (ECME) algorithm (Liu and Rubin, 1994; Laird et al., 1987) to

estimate its regression parameters with and without constraints. Tests of hypotheses regarding the regression parameters of the SLMM has not been considered before. We proposed the likelihood ratio (LR) tests for the constrained hypotheses. Their empirical significance levels and power are studied using simulations. The rest of this paper is organized as follows. Section 2 introduces the SLMM with linear inequality constraint. In Section 3, we propose the ECME algorithm for the ML estimation of the model parameters. In Section 4, the LR tests are considered. In Section 5, simulation results are presented. Section 6 applies the proposed methods to the London Junior School Project data set. We conclude, in Section 7, with a short discussion.

2 The Scaled Linear Mixed Model

Let $\mathbf{y}_i = (y_{i1}, \dots, y_{im})^T$ denote the m measurements taken on i th individual, $i = 1, \dots, n$. Since m outcomes are often measured on different scales, we propose a linear mixed model for outcomes standardized by error standard deviations as follows

$$\frac{y_{ij}}{\sigma_j} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i + \epsilon_{ij}, \quad (2.1)$$

where $j = 1, \dots, m, i = 1, \dots, n$, σ_j is the standard deviation of y_{ij} ; $\mathbf{x}_{ij}, \mathbf{z}_{ij}$ are $(p \times 1$ and $q \times 1$ respectively) known design vectors; $\epsilon_{ij} \sim N(0, 1)$ are independent random errors. The $p \times 1$ regression coefficients $\boldsymbol{\beta}$ are unknown fixed effects, and the \mathbf{b}_i is a $q \times 1$ vector of unknown subject-specific random effects. This version of the SLMM model generalizes the standard linear mixed model by introducing different variances across different outcomes.

The \mathbf{b}_i are distributed as $N_q(\mathbf{0}, \mathbf{D}(\boldsymbol{\theta}))$ and can be used to model correlations among different outcomes of the same subject where $\boldsymbol{\theta}$ is a $g \times 1$ vector of variance components to be estimated. Let $\boldsymbol{\Psi} = \text{diag}(\sigma_1^2, \dots, \sigma_m^2)$. Let $\mathbf{X}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{im})^T$ (of order $m \times p$), $\mathbf{Z}_i = (\mathbf{z}_{i1}, \dots, \mathbf{z}_{im})^T$ (of order $m \times q$) and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{im})^T$. Then model (2.1) can be expressed in matrix notation as

$$\boldsymbol{\Psi}^{-1/2} \mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad (2.2)$$

where $\boldsymbol{\Psi}^{-1/2} = \text{diag}(\sigma_1^{-1}, \dots, \sigma_m^{-1})$.

The marginal distribution of \mathbf{y}_i is $N_m(\boldsymbol{\Psi}^{1/2} \mathbf{X}_i \boldsymbol{\beta}, \boldsymbol{\Psi}^{1/2} \mathbf{V}_i(\boldsymbol{\theta}) \boldsymbol{\Psi}^{1/2})$, where $\mathbf{V}_i(\boldsymbol{\theta}) = \mathbf{Z}_i \mathbf{D}(\boldsymbol{\theta}) \mathbf{Z}_i^T + \mathbf{I}$. Let $\boldsymbol{\delta} = (\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\sigma}^2)$, where $\boldsymbol{\sigma}^2 = (\sigma_1^2, \dots, \sigma_m^2)^T$. The log-likelihood for the observed data is given by

$$\begin{aligned} \ell_0(\boldsymbol{\delta}) = & \sum_{i=1}^n \left\{ -\frac{m}{2} \ln 2\pi - \frac{1}{2} \ln |\boldsymbol{\Psi}| - \frac{1}{2} \ln |\mathbf{V}_i(\boldsymbol{\theta})| \right. \\ & \left. - \frac{1}{2} \left(\boldsymbol{\Psi}^{-1/2} \mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} \right)^T \mathbf{V}_i(\boldsymbol{\theta})^{-1} \left(\boldsymbol{\Psi}^{-1/2} \mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} \right) \right\}. \end{aligned} \quad (2.3)$$

Note that due to the presence of the scale matrix $\boldsymbol{\Psi}$ in the marginal mean and marginal covariance of \mathbf{y}_i , the standard linear mixed model methodology cannot be used to fit the scaled linear mixed model. So a new estimation procedure is warranted.

Often the fixed effects β may be subject to constraints which typically reflect prior information about the value of the parameters. These are represented by $R\beta \geq \mathbf{0}$ where R is a $k \times p$ matrix of full row rank, $\mathbf{0}$ is a $k \times 1$ vector of zeroes and the inequalities are coordinate wise.

3 Estimation

For estimation of parameters of (??), we propose an ECME algorithm which is a variant of the EM algorithm in which the M-step is decomposed into a series of conditional maximization (CM) steps.

Let $\mathbf{y} = (\mathbf{y}_1^T, \dots, \mathbf{y}_n^T)^T$ be the observed data. Considering the random effects as the missing data, we take $\mathbf{b} = (\mathbf{b}_1^T, \dots, \mathbf{b}_n^T)^T$ to be the unobserved data. Let $\mathbf{c} = (\mathbf{c}_1^T, \dots, \mathbf{c}_n^T)^T$, where $\mathbf{c}_i = (\mathbf{y}_i^T, \mathbf{b}_i^T)^T$ denote the complete data. The log-likelihood for the complete data is given by (except for trivial constants)

$$\ell_c(\delta) = -\frac{1}{2} \sum_{i=1}^n \{ \ln |\Sigma_i| + (\mathbf{c}_i - \boldsymbol{\mu}_i)^T \Sigma_i^{-1} (\mathbf{c}_i - \boldsymbol{\mu}_i) \} \quad (3.1)$$

where

$$\boldsymbol{\mu}_i = \begin{bmatrix} \Psi^{1/2} X_i \beta \\ \mathbf{0} \end{bmatrix} \text{ and } \Sigma_i = \begin{bmatrix} \Psi^{1/2} V_i(\theta) \Psi^{1/2} & \Psi^{1/2} Z_i D \\ D Z_i^T \Psi^{1/2} & D \end{bmatrix}.$$

For estimation of δ , we present an algorithm below. In Step 4 of the algorithm, to avoid negative estimates of σ^2 , we use a Fisher approximation which is developed as follows. Let $S(\sigma^2)$ denote the term obtained by differentiating $\ell_0(\delta)$ in (??) with respect to σ^2 . For a given σ_0^2 , use Taylor expansion of $S(\sigma^2)$ around σ_0^2 , which produces the approximation $S(\sigma^2) \approx S(\sigma_0^2) + I(\sigma_0^2)(\sigma^2 - \sigma_0^2)$ where $I(\sigma_0^2)$ is the Fisher information evaluated at σ_0^2 . Assuming that σ^2 is the MLE, $S(\sigma^2) = 0$, and after simplifying the Taylor expansion, $\sigma^2 \approx \sigma_0^2 + [I(\sigma_0^2)]^{-1} S(\sigma_0^2)$. This equation is used in (??) below to update σ^2 starting from a known σ_0^2 . Expressions for $I(\sigma^2)$, $S(\sigma^2)$ are given in Lin et al. (2000).

To estimate δ , the scheme is given below.

- (1) Initialize $\theta = \theta_0$, for some θ_0 , and, $\sigma^2 = \sigma_0^2$; for example, for the j th component σ_{j0}^2 use the sample variance of y_{ij} , for any $j = 1, \dots, m$. Then, maximize (??) with respect to β to obtain β_0 for given θ_0, σ_0^2 .
- (2) Calculate $y_{ij}^* = y_{ij}/\sigma_j$. Let $\mathbf{y}^* = (\mathbf{y}_1^{*T}, \dots, \mathbf{y}_n^{*T})^T$ where $\mathbf{y}_i^* = (y_{i1}^*, \dots, y_{im}^*)^T$.
- (3) Estimate β and θ : The ECME algorithm iterates between its E- and M-steps. In the $(k+1)$ th iteration of the E-step, the expectation $Q(\delta|\delta_k) = E[\ell_c(\delta)|\mathbf{y}^*; \delta_k]$ is computed where $\delta_k = (\beta_k, \theta_k, \sigma_k^2)$ is the estimated parameter after the k th iteration. This reduces to computing $E(\mathbf{b}_i|\mathbf{y}_i^*, \delta_k)$ and $E(\mathbf{b}_i \mathbf{b}_i^T|\mathbf{y}_i^*, \delta_k)$, the expected values of the sufficient statistics of the missing data (Laird et al., 1987; Liu et al., 1994).

The maximization of the $Q(\delta|\delta_k)$ in $(k+1)$ th iteration is carried out using 3 CM steps. In the first CM-step, θ_{k+1} is found by maximizing $Q(\beta_k, \theta, \sigma_k^2|\delta_k)$ with respect to θ . In the second

CM-step, first the unrestricted estimate β_{k+1} is found by maximizing the observed likelihood, $\ell_0(\beta, \theta_{k+1}, \sigma_k^2)$ with respect to β .

Next, to find the restricted MLE under $R\beta \geq 0$, solve for $\tilde{\beta}_{k+1}$, where

$$\tilde{\beta}_{k+1} = \arg \min \left\{ \sum_{i=1}^n (\mathbf{y}_i^* - \mathbf{X}_i \beta)^T [\mathbf{V}_i(\theta_{k+1})]^{-1} (\mathbf{y}_i^* - \mathbf{X}_i \beta) : R\beta \geq 0 \right\}. \quad (3.2)$$

In the third CM step, updates for σ^2 are computed using a one-step Fisher approximation (as described above) for iteration $k + 1$ of the form

$$\sigma_{k+1}^2 = \sigma_k^2 + I(\sigma_k^2)^{-1} S(\sigma_k^2). \quad (3.3)$$

- (4) Check for convergence. At $(k + 1)$ th iteration, comparing the j th components of δ (say, $\delta_{j,k}$), if $|\delta_{j,k+1} - \delta_{j,k}|, \forall j$, is less than a pre-specified convergence criterion, then stop. Otherwise, go back to step 2.

It can be shown that the estimators produced by the algorithm are consistent (Nettleton, 1999; Liu and Rubin, 1994).

4 Likelihood Ratio Tests

We consider testing hypotheses of the form

$$H_0 : R\beta = 0 \text{ versus } H_1 : R_1\beta \geq 0 \quad (4.1)$$

(with at least one strict inequality under H_1) where $R = (R_1^T, R_2^T)^T$ is a $k \times p$ matrix of full row rank, R_1 is $q \times p$ with $q \leq k$ and R_2 is $(k - q) \times p$. For example, if one is interested in testing whether the fixed effects have positive effect, then one would first set $k = p = q$, and then form $H_0 : \beta = 0$ and $H_1 : \beta_i > 0, 1 \leq i \leq q$, where $R = R_1$ are identity matrices.

The regularity conditions needed for the validity of the theorem are stated as Condition Q in Silvapulle and Sen (2005, page 146); see also Self and Liang (1987). These conditions can be verified for the scaled linear model similar to the proofs of Davis et al. (2012) and Sinha (2004).

Let $\bar{\delta} = (\bar{\beta}, \bar{\theta}, \bar{\sigma}^2)$ be the ML estimate of δ under H_0 (obtained by setting $R\beta = 0$ in (??)) and $\tilde{\delta} = (\tilde{\beta}, \tilde{\theta}, \tilde{\sigma}^2)$ be the ML estimate of δ under H_1 . From (??), the likelihood ratio test statistic is given by

$$LRT = 2[\ell_0(\tilde{\delta}) - \ell_0(\bar{\delta})], \quad (4.2)$$

whose asymptotic distribution is given below.

Theorem 3.1 Under (??) and its assumptions, for $t > 0$ and under H_0 ,

$$\lim_{n \rightarrow \infty} P(LRT \geq t) = \sum_{i=0}^q w_i P(\chi_i^2 \geq t), \quad (4.3)$$

where the weights w_i depend on the constraints and the related covariance matrix.

Proof. The LRT in (??) equals

$$\begin{aligned} & \sum_{i=1}^n \left\{ \ln |\tilde{\Psi}| + \ln |V_i(\tilde{\theta})| + \left(\tilde{\Psi}^{-1/2} \mathbf{y}_i - \mathbf{X}_i \tilde{\beta} \right)^T V_i(\tilde{\theta})^{-1} \left(\tilde{\Psi}^{-1/2} \mathbf{y}_i - \mathbf{X}_i \tilde{\beta} \right) \right\} \\ & - \sum_{i=1}^n \left\{ \ln |\bar{\Psi}| + \ln |V_i(\bar{\theta})| + \left(\bar{\Psi}^{-1/2} \mathbf{y}_i - \mathbf{X}_i \bar{\beta} \right)^T V_i(\bar{\theta})^{-1} \left(\bar{\Psi}^{-1/2} \mathbf{y}_i - \mathbf{X}_i \bar{\beta} \right) \right\}, \end{aligned} \quad (4.4)$$

where $\tilde{\Psi}, \bar{\Psi}$ are obtained using $\tilde{\sigma}, \bar{\sigma}$, respectively. For $\delta = (\beta, \theta, \sigma^2)$ being the true parameter value, under H_0 , both $\tilde{\delta}, \bar{\delta} \rightarrow \delta$ as $n \rightarrow \infty$. Using Proposition 4.3.1(1) of Silvapulle and Sen (2005), the asymptotic distribution of (??) as $n \rightarrow \infty$ is same as that of

$$\min_{\mathbf{R}, \beta=0} \{ (\mathbf{Z}_n - \delta)^T [\mathbf{I}(\delta_0)] (\mathbf{Z}_n - \delta) \} - \min_{\mathbf{R}_1, \beta \geq 0} \{ (\mathbf{Z}_n - \delta)^T [\mathbf{I}(\delta_0)] (\mathbf{Z}_n - \delta) \} + o_p(1), \quad (4.5)$$

where $\mathbf{Z}_n = n^{-1/2} [\mathbf{I}(\delta_0)]^{-1} \mathbf{S}(\delta_0)$, $\mathbf{S}(\delta) = (\partial/\partial \delta) \ell(\delta)$, $\mathbf{I}(\delta) = E_{\delta} \{ (\partial/\partial \delta) \log(f(\mathbf{y}, \delta)) (\partial/\partial \delta)^T \log(f(\mathbf{y}, \delta)) \}$. Let \mathbf{R}^* be the $k \times (p + g + m)$ matrix obtained by augmenting the matrix \mathbf{R} by $g + m$ columns of zeroes. The parameters regions under H_0 and H_1 may be redefined as $H_0 : \{ \delta : \mathbf{R}^* \delta = 0 \}$ and $H_1 : \{ \delta : \mathbf{R}_1^* \delta \geq 0 \}$ (with at least one strict inequality), respectively, where $\mathbf{R}^* = (\mathbf{R}_1^{*T}, \mathbf{R}_2^{*T})^T$ is a $k \times (p + g + m)$ matrix of full row rank, \mathbf{R}_1^* is $q \times (p + g + m)$ with $q \leq k$ and \mathbf{R}_2^* is $(k - q) \times (p + g + m)$. Let the elements of the 3×3 block matrix $[\mathbf{I}(\delta)]^{-1}$ be denoted by $I^{ij}(\delta)$. Then, it follows that $\mathbf{R}_1^* [\mathbf{I}(\delta_0)]^{-1} \mathbf{R}_1^{*T} = \mathbf{R}_1 \mathbf{I}^{11}(\delta_0) \mathbf{R}_1^T$.

Using Proposition 4.3.1(2) of Silvapulle and Sen (2005), the asymptotic null distribution of (??) is given by

$$\lim_{n \rightarrow \infty} P_{\delta_0}(LRT \geq t | H_0) = \sum_{i=0}^q w_i(q, \mathbf{R}_1 \mathbf{I}^{11}(\delta_0) \mathbf{R}_1^T) P(\chi_i^2 \geq t), \quad (4.6)$$

from which (??) follows. \square

As mentioned earlier, the distribution on the right side of (??) is known as *chi-bar square*. The weights $w_i = w_i(q, \Sigma)$, where $\Sigma = \mathbf{R}_1 \mathbf{I}^{11}(\delta_0) \mathbf{R}_1^T$ is the true asymptotic variance of the unconstrained MLE, can be computed exactly for $q \leq 3$. For larger values of q , weights are estimated or simulated (see Section 3.5 in Silvapulle and Sen, 2005).

As the chi-bar squared distribution may be difficult to calculate, the following bounds (Silvapulle and Sen, 2005) on its tail probabilities may be useful. As it is known that

$$\frac{1}{2} [P(\chi_{k-q}^2 \geq c) + P(\chi_{k-q+1}^2 \geq c)] \leq \sum_{i=0}^q w_i P(\chi_{k-q+i}^2 \geq c) \leq \frac{1}{2} [P(\chi_{k-1}^2 \geq c) + P(\chi_k^2 \geq c)]. \quad (4.7)$$

So one may reject H_0 if the upper bound in (??) is less than α , or do not reject H_0 if the lower bound in (??) is greater than α where α is the significance level. But the test remains inconclusive if $\frac{1}{2} [P(\chi_{k-q}^2 \geq c) + P(\chi_{k-q+1}^2 \geq c)] \leq \alpha \leq \frac{1}{2} [P(\chi_{k-1}^2 \geq c) + P(\chi_k^2 \geq c)]$.

5 Simulation Studies

In our coding, we begin by importing the necessary packages `MASS`, `psych`, and `quadprog` from R library. An additional package `nleqslv` has been used for the estimation related with chi-bar square distribution. Our goal is to compare the performances of the unrestricted and restricted estimates of δ in model (??) produced by the ECME algorithm. We also consider the empirical significance levels of the LR test in (??) and its power.

The setup of our simulation is similar to that in Davidov and Rosen (2011) with modifications for suitability of the heterogeneous model (??). Using $n = 20, m = 5, p = 3, q = 2, g = 4$, the j th row in the \mathbf{X}_i and \mathbf{Z}_i matrices are $X_{ij} = (1, x_{1ij}, x_{2ij})$ and $Z_{ij} = (1, x_{2ij})$, respectively. The variable x_1 is chosen independently of index j and given by $\mathbf{x}_{1i}^T = (i, i, i, i, i)$. However, the variable x_2 depends on j , and is given by $\mathbf{x}_{2i}^T = (0, 1, 2, 3, 4), 1 \leq i \leq n$. The errors are $\epsilon_i \stackrel{iid}{\sim} N(\mathbf{0}, \mathbf{I}_5)$. The terms $\mathbf{b}_i \stackrel{iid}{\sim} N(\mathbf{0}, \mathbf{D})$, with \mathbf{D} being 2×2 with $d_{ii} = 1, d_{ij} = .5, i \neq j$. The scale matrix is set as $\Psi = \text{diag}(0.64, 0.81, 1.0, 1.21, 1.44)$. We ran simulations using other values of \mathbf{D} and σ^2 as well as other covariate configurations with similar results.

The regression coefficients $\beta = (\beta_0, \beta_1, \beta_2)$ were restricted to satisfy $\beta_1 \geq 0, \beta_2 \geq 0$.

Estimation: We used the algorithm developed in Section 3 to estimate the parameters of the model. Let MSE_0 denote the mean square errors (MSE) of the unconstrained estimators, and, let MSE_1 denote the MSE of the constrained estimators. The relative MSE of the constrained to the unconstrained estimators is given by MSE_1/MSE_0 .

Table 1: Relative mean square errors MSE_1/MSE_0 for the estimates

β	β_0	β_1	β_2	D_{11}	D_{12}	D_{22}	σ_1^2	σ_2^2	σ_3^2	σ_4^2	σ_5^2
(0, 0, 0)	0.56	0.40	0.55	1.04	1.02	1.03	1.0	1.0	1.03	1.01	0.99
(0, 0.05, 0.05)	0.70	0.65	0.55	0.93	1.04	0.99	0.97	0.99	1.01	0.98	0.95
(0, 0.10, 0.10)	0.85	0.84	0.58	0.85	1.02	0.97	0.97	0.97	0.99	0.97	0.93

Table 1 shows the relative mean square errors for the estimates of β, \mathbf{D} and Ψ using 3,000 replications. As can be seen, the proposed method improves on the unconstrained estimates, especially for the regression coefficients. For smaller values of the regression parameters, the improvement can be more than 50%. The relative MSE of the variance components \mathbf{D} and Ψ are close to 1 indicating the constraints have little effect on estimation of these parameters. We experimented using larger sample sizes and different initial values for the parameters which showed similar results.

Testing: We consider testing $H_0 : \mathbf{R}\beta = \mathbf{0}$ versus $H_1 - H_0, H_1 : \mathbf{R}\beta \geq \mathbf{0}$, where

$$\mathbf{R} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \beta = (\beta_0, \beta_1, \beta_2)^T.$$

We present the empirical significance levels in Table 2 using 3,000 replications for different sample sizes. As is evident from Table 2, the empirical significance levels approach the nominal level

$\alpha = 0.05$ as the sample size grows, as expected. Also, the restricted test seems to perform slightly better than the unrestricted test as their empirical levels are closer to the nominal level.

Table 2: Empirical significance levels of unrestricted and restricted tests

n	unrestricted	restricted
20	0.076	0.047
50	0.058	0.045
100	0.055	0.051

Table 3: Simulated powers of the unrestricted and restricted tests

Sample sizes	Unrestricted		Restricted	
	$\beta_1 = .01$	$\beta_2 = .03$	$\beta_1 = .01$	$\beta_2 = .03$
20	0.075	0.101	0.066	0.111
30	0.102	0.174	0.122	0.209
40	0.114	0.314	0.158	0.368
50	0.138	0.539	0.352	0.599
60	0.429	0.759	0.488	0.807
100	0.529	0.999	0.591	1.000

We also compared the powers of unrestricted and restricted tests. Table 3 shows that incorporating the constraints improves the power of the likelihood ratio tests, especially in small and medium sample sizes. Improvement is substantial in situations where the power is low (near H_0). We conclude from Table 3 that the restricted tests perform slightly better than the unrestricted counterpart as their empirical levels are closer to the nominal level (Table 2).

6 Analysis of the Junior School Project Data

The Junior School Project data is a longitudinal study of 1,618 pupils from 50 primary schools chosen at random among the 636 schools under the Inner London Education Authority in 1980. The 'jisp' in R programming software is a data frame with 3236 observations on 9 variables as follows. There are six categorical independent variables with levels as follows: school (1 to 50), class (1 to 4), student ID (1 to 1402), gender (f or m), year (0, 1 or 2), social (father's socioeconomic status 1 to 9 from highest to lowest) and one quantitative independent variable: Raven (School admission test score 0 – 40). The response variables are mathematics test score (range 0 – 40) and English test score (range 0 – 100).

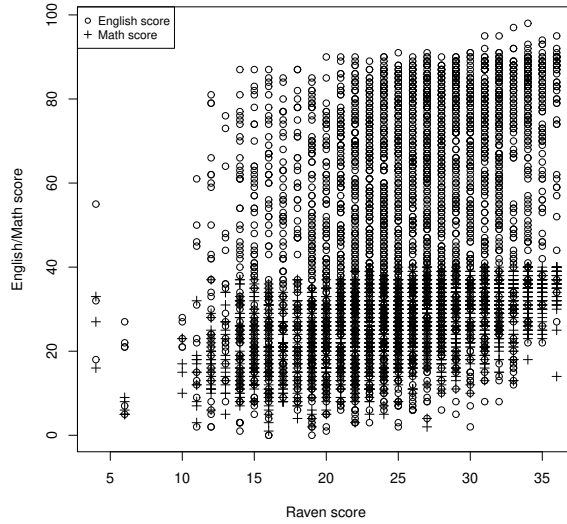


Figure 1: Plots of English versus Raven and mathematics versus Raven scores to show the differences in dispersion of output variables

The variables (gender, social, Raven) are taken as fixed effects (with an intercept term), and the corresponding fixed effect vector is β (hence, $p = 4$). However, a student could be admitted to any one school from a pool of different schools, to any one of different classes available in that school, and be assigned a random ID, so the variables school, class and student ID are taken as random effects, which are associated (in that order with an intercept term, hence, $q = 4$) with the random vector \mathbf{b} as those variables could have different factor effects on a student’s overall test score.

Outcomes are $y_{ij}, j = 1, 2, i = 1, 2, \dots, 1618$, where $y_{i1} =$ English test score and $y_{i2} =$ mathematics test score. The model (??) is fit to data with $n = 1618, m = 2, p = 4, q = 4, g = 10$ (as the 4×4 covariance matrix of \mathbf{b} has 10 non-duplicated terms). The standard deviation of the English test scores is 24.7123, and the standard deviation of Mathematics test scores is 7.63738 (see Figure 1). Since the standard deviations of these test groups are quite different, appropriate scaling is needed, and model (??) would be appropriate.

Based on the discussion in Section 1 on prior knowledge regarding relation between the test scores and the fixed effect variables, we set constraints as $H_1 : \beta_1 > 0, \beta_2 < 0, \beta_3 > 0$. The maximum likelihood estimates of various parameters are obtained using the ECME algorithm as described in Section 3 along with the Fisher approximation method. The unrestricted estimates are $\hat{\beta} = (0.836, 0.281, -0.041, 0.085), \hat{\sigma}_{11}^2 = 389.052, \hat{\sigma}_{21}^2 = 69.805$, and with $\mathbf{D} = (d_{ij})$, for $i \geq j$, $\hat{d}_{ij} = 0.741, 0.232, 0.241, -0.009, 0.723, 0.232, -0.026, 0.741, -0.009, 0.0009$. The restricted estimate under H_1 of β is identical to its unrestricted estimate since the unrestricted estimate satisfies

the constraints specified in H_1 spontaneously. Under H_0 , the restricted estimates of variances are $\hat{\sigma}_{10}^2 = 252.796$, $\hat{\sigma}_{20}^2 = 55.496$, and $\hat{\beta}_0 = 3.440$. The restricted estimates, under H_0 , of D remained the same as those under H_1 .

When testing $H_0 : \beta = \mathbf{0}$ versus $H_1 - H_0$, where $H_1 : \beta_1 \geq 0, \beta_2 \leq 0, \beta_3 \geq 0$, the likelihood ratio test statistic is 4649371 which is larger than the chi-bar square critical value of 2.786 (using estimates under H_0), so we reject H_0 with a p -value of 0.0000 (with 5% significance level). Using an unconstrained test, one would still reject H_0 with a p -value of 0.0000 (with 5% significance level) and conclude that the variables gender, social and Raven affect the English and Mathematics test scores but would not be able to detect the direction of influence with respect to the levels of each variable.

7 Discussion

Scaled mixed models are improved versions of mixed models which considers heterogeneity of multiple response variables. We considered an SLMM model where its parameters may be subject to some restrictions which arise naturally in many researches due to the nature of the problem or the data. For example, in Junior School Project data, one would expect a pattern of female test scores to be higher than those of male, students from higher social class to have higher scores, etc. Such constraints improve the analysis of data and are easily incorporated at the expense of additional computations. We develop the necessary methodology for the constrained inference by proposing an ECME-based algorithm for estimating the regression parameters. In simulations, the proposed methods with constrained estimators demonstrate an improvement in terms of MSE over the unconstrained estimators. In some settings, the improvement are substantial. For example, for values of the parameters close to the boundary of the parameter space up to 50% improvement was observed in simulations.

Hypotheses testing procedures that incorporate the constraints are also developed. It is shown with simulations that incorporating constraints results in higher power overall, and in some cases substantial gains are noted. The methodology is used to analyze English and mathematics test scores from Junior School Project study where effects of specific directions of gender, social and admission test are shown to be significant only if one accounts for the constraints.

any real-world problems have inherent restrictions that must be considered in order to obtain a feasible solution. Analyzing multiple outcomes can be challenging, but it also offers opportunities to gain insight into complex systems. We scale the multiple outcomes using the error standard deviation σ_j . We use the ECME algorithm along with the Fisher approximation method to obtain the MLEs of various parameters of the model. In this method, the unrestricted and restricted estimation methods take approximately similar processing times so the restricted method would be recommended when constraints are present. We did not compare our method with that of Lin et al. (2000) because of the inaccuracy in variance estimation of the latter as mentioned earlier.

For future directions, a natural extension would be to consider a constrained multivariate version of the mixed linear model with heteroscedastic scales. It would also be of interest to consider marginal models with heteroscedasticity where parameters satisfy some restrictions. We used R for

all the programming in this paper and the codes are available from authors to implement the method. Finally, we note that the methodology we propose is general and can be used in any longitudinal setting and not only in the context of test score data.

References

- Allensworth, E. M. and Clark, K. (2020), "High School GPAs and ACT Scores as Predictors of College Completion: Examining Assumptions About Consistency Across High Schools," *Educational Researcher*, 49, 198-211.
- Balart, P., Oosterveen, M. (2019), "Females Show More Sustained Performance During Test-Taking Than Males," *Nat Commun.*, 10, 37-98.
- Borghans, L., Meijers, H., & Ter Weel, B. (2008), "The Role of Noncognitive Skills in Explaining Cognitive Test Scores," *Economic Inquiry*, 46, 2-12.
- Cornwell, C., Mustard, D. B. & Van Parys, J. (2013), "Noncognitive Skills and the Gender Disparities in Test Scores and Teacher Assessments: Evidence From Primary School," *J. Hum. Resour.*, 48, 236-264.
- Davidov, O., Rosen, S. (2011), "Constrained Inference in Mixed-Effects Models for Longitudinal Data with Application to Hearing Loss," *Biostatistics*, 12, 327-340.
- Davis, K.A., Park, C.G. and Sinha, S.K. (2012), "Testing for Generalized Linear Mixed Models with Cluster Correlated Data Under Linear Inequality Constraints," *Canadian Journal of Statistics*, 40, 243-258.
- de Luma, O. (2021), The Junior School Project, www.odelama.com.
- Duckworth, A. L., Quinn, P. D., Lynam, D. R., Loeber, R., & Stouthamer-Loeber, M. (2011), "Role of Test Motivation in Intelligence Testing," *Proceedings of the National Academy of Sciences*, 108, 7716-7720.
- Fang, H. B., Tian, G. L., Xiong, X. and Tan, M. (2006), "A Multivariate Random-Effects Model with Restricted Parameters: Application to Assessing Radiation Therapy for Brain Tumors," *Statistics in Medicine*, 25, 1948-1959.
- Farnan, L., Ivanova, A. and Peddada, S.D. (2014), "Linear Mixed Effects Models Under Inequality Constraints with Applications," *PLoS ONE*, 9, e847-878.
- Fryer, R. G. and Levitt, S. D. (2010), "An Empirical Analysis of the Gender Gap in mathematics," *Am. Econ. J.-Appl. Econ.*, 2, 210-240.
- Gneezy, U., & Rustichini, A. (2000), "Pay Enough or Don't Pay at All," *Quarterly Journal of Economics*, 791-810.

- Guiso, L., Sapienza, P., Monte, F. & Zingales, L. (2010), "Culture, Gender and Math.," *Science*, 320, 1164–1165.
- Hanushek, E. A., & Kimko, D. D. (2000), "Schooling, Labor-Force Quality, and the Growth of Nations," *The American Economic Review*, 1184–1208.
- Hanushek, E. A., & Woessmann, L. (2008), "The Role of Cognitive Skills in Economic Development," *Journal of Economic Literature*, 607–668.
- Hanushek, E. A., & Woessmann, L. (2012), "Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation," *Journal of Economic Growth*, 17, 267–321.
- Hyde, J. S., Fennema, E. & Lamon, S. J. (1990), "Gender Differences in Mathematics Performance: a Meta-Analysis," *Psychol. Bull.*, 107, 139–155.
- Jacob, B. and Rothstein, J. (2016), "The Measurement of Student Ability in Modern Assessment Systems," *Journal of Economic Perspectives*, 30, 85-108.
- Joensen, J. S. & Nielsen, H. S. (2009), "Is There a Causal Effect of High School Math on Labor Market Outcomes?," *J. Hum. Resour.*, 44, 171–198.
- Koretz, D. (2008), "Measuring up. What Educational Testing Really Tells Us," *Harvard University Press, Cambridge, Mass.*
- Kitsantas, A., Winsler, A., Huie, F., Duckworth, A. L., Quinn, P. D., Lynam, D. R., Loeber, R., & Stouthamer-Loeber, M. (2008), "Self-Regulation and Ability Predictors of Academic Success During College," *Journal of Advanced Academics*, 20, 42-68.
- Kennedy, E. H., Kangovi, S., Mitra, N. (2019), "Estimating Scaled Treatment Effects with Multiple Outcomes," *Stat Methods Med Res*, 28, 1094-1104.
- Kim, D. K. and Taylor, J. M. G. (1995), "The Restricted EM Algorithm for Maximum Likelihood Estimation Under Linear Restrictions on the Parameters," *Journal of the American Statistical Association*, 90, 708–716.
- Laird, N., Lange, N. and Stram, D. (1987), "Maximum Likelihood Computations with Repeated Measures: Application of the EM Algorithm," *Journal of the American Statistical Association*, 82, 97-105.
- Lin, X., Ryan, L., Sammel, M., Zhang, D., Padungtod, C. and XU, X. (2000), "A Scaled Linear Mixed Model for Multiple Outcomes," *Biometrics*, 56, 593–601.
- Lindberg, S. M., Hyde, J. S., Petersen, J. L. & Linn, M. C. (2010), "New Trends in Gender and Mathematics Performance: a Meta-Analysis," *Psychol. Bull.*, 136, 1123–1135.
- Liu, C. and Rubin, D. B. (1994), "The ECME Algorithm: a Simple Extension of EM and ECM with Fast Monotone Convergence," *Biometrika*, 81, 633–648.

- Mortimer, P., Ecob, and Sammons P., Louise, S., Lewis, D. and Ecob, R. (1989), "A Study of Effective Junior Schools," *International Journal of Educational Research*, 13, 75–768.
- Mclachlan, G. J. and Krishnan, T. (1997), "The EM Algorithm and Extensions," *New York: Wiley*.
- Mukerjee H (1988), "Order Restricted Inference in a Repeated Measures Model," *Biometrika*, 75, 616–617.
- Nettleton, D. (1999), "Convergence Properties of the EM Algorithm in Constrained Parameter Spaces," *The Canadian Journal of Statistics*, 27, 639–648.
- Quinn, D. M. & Cooc, N. (2015), "Science Achievement Gaps by Gender and Race/Ethnicity in Elementary and Middle School Trends and Predictors," *Educ. Res.*, 44, 336–346.
- Roy, J., Lin, X., Louise, M.R. (2003), "Scaled Marginal Models for Multiple Continuous Outcomes," *Biostatistics*, 4, 371–383.
- Rubin, D. B., Stuart, E. A. and Zanutto, E. L. (2004), "A Potential Outcomes View of Value-Added Assessment in Education," *Journal of Educational and Behavioral Statistics*, 29, 103–116.
- Segal, C. (2012), "Working When No One is Watching: Motivation, Test Scores, and Economic Success," *Management Science*, 58, 1438–1457.
- Self, S. G. and Liang, K-Y. (1987), "Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions," *Journal of the American Statistical Association*, 82, 60–610.
- Shi, N. Z., Zheng, S. R. and Guo, J. (2005), "The Restricted EM Algorithm Under Inequality Restrictions on the Parameters," *Journal of Multivariate Analysis*, 92, 53–76.
- Silvapulle MJ, Sen PK (2005), "Constrained Statistical Inference: Inequality, Order, and Shape Restrictions," *Hoboken, N.J.: Wiley-Interscience*
- Silvapulle MJ (1997), "On Order Restricted Inference in Some Mixed Linear Models," *Statistics & Probability Letters*, 36, 23–27.
- Singh B, Wright FT (1990), "Testing for and Against an Order Restriction in Mixed Effects Models," *Statistics & Probability Letters*, 9, 195–200.
- Sinha, S. K. (2004), "Robust Analysis of Generalized Linear Mixed models," *Journal of the American Statistical Association*, 99, 451–460.
- Wechsler, D. (1940), "Non-Intellective Factors in General Intelligence," *The Journal of Abnormal and Social Psychology*, 37, 444–445.

Received: October 18, 2023

Accepted: April 5, 2024