Fundamentals of Contemporary Mathematical Sciences
doi:10.54974/fcmathsci.1263030
(**2024**) **5**(1) 25 – 35
Research Article

Generalized Kibria-Lukman Prediction Approximation in Linear Mixed Models

Özge Kuran *
Dicle University, Faculty of Science, Department of Statistics
Diyarbakır, Türkiye

Received: 10 March 2023 **Accepted:** 25 December 2023

Abstract: One of the new suggested prediction methods is the Kibria-Lukman's prediction approach under multicollinearity in linear mixed models and in this article, the generalized Kibria-Lukman estimator and predictor are introduced to combat multicollinearity problem. The comparisons between the proposed generalized Kibria-Lukman estimator/predictor and several other estimators/predictors, namely the best linear unbiased estimator/predictor and Kibria-Lukman estimator/predictor are done by using the matrix mean square error criterion. Lastly, the selection of the biasing parameter is given and to demonstrate the performance of our new defined prediction method, the greenhouse gases data analysis is made.

Keywords: Linear mixed model, mean square error, generalized Kibria-Lukman predictor, multicollinearity.

1. Introduction

The linear mixed model (LMM) is described the following form for $i = 1, \ldots, m$,

$$y_i = X_i\beta + Z_i u_i + \varepsilon_i,$$

where y_i is an $n_i \times 1$ vector of response variables measured on subject i, β is a $p \times 1$ parameter vector of fixed effects, X_i and Z_i are $n_i \times p$ and $n_i \times q$ known design matrices of the fixed and random effects, respectively, u_i is a $q \times 1$ random vector, the components of which are called random effects and ε_i is an $n_i \times 1$ random vector of errors. LMM mostly has the assumptions given below

$$u_i \overset{iid}{\sim} N_q\left(0, \sigma^2 F\right) \text{ and } \varepsilon_i \overset{iid}{\sim} N_{n_i}\left(0, \sigma^2 W_i\right), \quad i = 1, \ldots, m,$$

where u_i and ε_i are independent, F and W_i are $q \times q$ and $n_i \times n_i$ known positive definite (pd) matrices.

*Correspondence: ozge.kuran@dicle.edu.tr
2020 *AMS Mathematics Subject Classification*: 62J05, 62J07.

$y = (y_1^T, \ldots, y_m^T)^T$, $X = (X_1^T, \ldots, X_m^T)^T$, $Z = \oplus_{i=1}^m Z_i$ ($\oplus$ is the direct sum), $u = (u_1^T, \ldots, u_m^T)^T$ and $\varepsilon = (\varepsilon_1^T, \ldots, \varepsilon_m^T)^T$ is taken. So, the more compact model can be written as

$$y = X\beta + Zu + \varepsilon, \tag{1}$$

this means $\begin{pmatrix} u \\ \varepsilon \end{pmatrix} \sim N_{qm+n}\left(\begin{pmatrix} 0_{qm} \\ 0_n \end{pmatrix}, \begin{pmatrix} \sigma^2 G & 0 \\ 0 & \sigma^2 W \end{pmatrix}\right)$, where $n = \sum_{i=1}^m n_i$, $G = I_m \otimes F$, $W = \oplus_{i=1}^m W_i$ ($\otimes$ is the Kronecker product) and I_m is the identity matrix of order m. $y \sim N\left(X\beta, \sigma^2 H\right)$ is written under model (1), where $H = ZGZ^T + W$. It is assumed that the G and W matrices are known for ease of theoretical calculations. But, if this assumption is not satisfied, we substitute their maximum likelihood (ML) or restricted maximum likelihood (REML) estimates for the G and W. $\hat{\beta}$ and $\hat{u}$ were obtained by [4, 5] as follows

$$\begin{aligned} \hat{\beta} &= (X^T H^{-1} X)^{-1} X^T H^{-1} y, \\ \hat{u} &= GZ^T H^{-1}(y - X\hat{\beta}), \end{aligned} \tag{2}$$

and they were, respectively, named as BLUE (the best linear unbiased estimator of β) and BLUP (the best linear unbiased predictor of u).

This article aims to reveal a new prediction method, which is an alternative to the existing estimators/predictors defined below in the LMM literature under multicollinearity and, for the sake of actualizing this aim, is to introduce a generalized form of Kibria-Lukman prediction method in LMMs by following [1] generalized Kibria-Lukman estimator in linear regression models. Thus, the rest of our study is configured as follows: We give our preliminaries in Section 2. We obtain the generalized Kibria-Lukman estimator and predictor in LMMs via [1] in linear regression models in Section 3. Matrix mean square error (MMSE) performances are evaluated in Section 4. We mention about biasing parameter selection in Section 5 and in Section 6, greenhouse gases data analysis is ensured to show our theoretical findings. Finally, in Section 7, we discuss some conclusions.

2. Preliminaries

Multicollinearity is defined as the linear dependence between the columns of X. The statistical consequences of this effect, such as the parameter estimates having large variances and being different from the true values, are well known in all linear regression models, including LMM. In order to eliminate the effects of this effect, there are many methods defined in both linear regression models and LMM, and ridge regression in the linear regression models recommended by [6] is the most well-known method among these methods. Under LMM, [11, 13] identified the ridge estimator and predictor with $k > 0$ ridge biasing parameter as follows

$$\begin{aligned} \hat{\beta}_k &= (X^T H^{-1} X + kI_p)^{-1} X^T H^{-1} y, \\ \hat{u}_k &= GZ^T H^{-1}(y - X\hat{\beta}_k). \end{aligned} \tag{3}$$

In addition to ridge regression, [7, 10] suggested Liu's approach in linear regression models. By following [14, 15, 20] proposed the Liu estimator predictor via $0 < d < 1$ Liu biasing parameter under LMM as follows

$$\begin{aligned} \hat{\beta}_d &= (X^T H^{-1} X + I_p)^{-1}(X^T H^{-1} y + d\hat{\beta}), \\ \hat{u}_d &= GZ^T H^{-1}(y - X\hat{\beta}_d), \end{aligned} \tag{4}$$

where $\hat{\beta}$ is the BLUE in Equation (2).

In linear regression models, [9] proposed a new one-parameter estimator in the class of ridge and Liu estimators and they called their new estimator as the Kibria-Lukman (KL) estimator. By following [9] in linear regression models, [12] suggested respectively the KL estimator and the KL predictor in LMMs as

$$\begin{aligned} \hat{\beta}_{KL} &= (X^T H^{-1} X + kI_p)^{-1}(X^T H^{-1} y - k\hat{\beta}) = (X^T H^{-1} X + kI_p)^{-1}(X^T H^{-1} X - kI_p)\hat{\beta} \\ &= (I_p + k(X^T H^{-1} X)^{-1})^{-1}(I_p - k(X^T H^{-1} X)^{-1})\hat{\beta}, \\ \hat{u}_{KL} &= GZ^T H^{-1}(y - X\hat{\beta}_{KL}). \end{aligned} \tag{5}$$

Now, we will introduce a new prediction approximation as an alternative to the estimators/predictors defined above under multicollinearity.

3. Introduced New Prediction Approximation

Via [1] in linear regression models, a new prediction approximation is handled in LMMs in this part. With model (1) assumptions, we have

$$\begin{pmatrix} u \\ y \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ X\beta \end{pmatrix}, \sigma^2 \begin{pmatrix} G & GZ^T \\ ZG & H \end{pmatrix} \right), y|u \sim N\left(X\beta + Zu, \sigma^2 W\right),$$

[5] maximize

$$\begin{aligned} f(y,u) &= f(y|u)\, f(u) \\ &= (2\pi\sigma^2)^{-(n+qm)/2} |W|^{-1/2} |G|^{-1/2} \\ &\quad \times \exp\{-\frac{1}{2\sigma^2}[(y - X\beta - Zu)^T W^{-1}(y - X\beta - Zu) + u^T G^{-1} u]\}, \end{aligned}$$

where $|.|$ is a matrix determinant and thus, $\log f(y,u)$ is obtained

$$\begin{aligned}\log f(y,u) &= \log f(y|u) + \log f(u)\\ &= -\frac{1}{2}\{(n+qm)\log(2\pi) + (n+qm)\log\sigma^2 + \log|W| + \log|G|\\ &\quad +[(y-X\beta-Zu)^T W^{-1}(y-X\beta-Zu) + u^T G^{-1}u]/\sigma^2\}.\end{aligned}$$

Our goal is to describe a new prediction method which is resistant to multicollinearity alternative to ridge, Liu and KL prediction approaches in LMMs. Via [1], $\log f(y,u)$ is minimized under $(\beta+\hat{\beta})^T(\beta+\hat{\beta}) = c$ with $\delta = \frac{1}{2\sigma^2} \geq 0$ regularization parameter

$$\log f(y,u) - \frac{1}{2\sigma^2}K[(\beta+\hat{\beta})^T(\beta+\hat{\beta}) - c], \tag{6}$$

where $K = diag(k_1,\dots,k_p)$ for $0 < k_i < 1$, $i = 1,\dots,p$, as the ridge biasing parameters and c is a constant. Substituting the log function into Equation (6) and removing the constant term from the model,

$$\begin{aligned}&-\frac{1}{2\sigma^2}\{(y-X\beta)^T W^{-1}(y-X\beta) + K[(\beta+\hat{\beta})^T(\beta+\hat{\beta}) - c]\}\\ &-\frac{1}{2\sigma^2}\{u^T(Z^TW^{-1}Z+G^{-1})u - 2(y-X\beta)^T W^{-1}Zu\},\end{aligned} \tag{7}$$

is written. Initially, we get partial derivatives of Equation (7) corresponding to β and u. Later, we equalize these derivatives to zero. Thus, we derive the following equations

$$X^TW^{-1}(y-X\hat{\beta}_{GKL}) - K\hat{\beta} - K\hat{\beta}_{GKL} - X^TW^{-1}Z\hat{u}_{GKL} = 0, \tag{8}$$

$$Z^TW^{-1}(y-X\hat{\beta}_{GKL}) - (Z^TW^{-1}Z+G^{-1})\hat{u}_{GKL} = 0 \tag{9}$$

and we name as $\hat{\beta}_{GKL}$ and $\hat{u}_{GKL}$, respectively, as the generalized KL (GKL) estimator and predictor, respectively.

We present Equations (8) and (9) as

$$\begin{pmatrix} X^TW^{-1}X+K & X^TW^{-1}Z\\ Z^TW^{-1}X & Z^TW^{-1}Z+G^{-1}\end{pmatrix}\begin{pmatrix}\hat{\beta}_{GKL}\\ \hat{u}_{GKL}\end{pmatrix} = \begin{pmatrix}X^TW^{-1}y - K\hat{\beta}\\ Z^TW^{-1}y\end{pmatrix}. \tag{10}$$

We write Equation (10) via [3] as follows:

$$C\hat{\Psi} = \omega^TW^{-1}y + \kappa, \tag{11}$$

where $\hat{\Psi} = (\hat{\beta}_{GKL}^T, \hat{u}_{GKL}^T)^T$, $\omega = (X, Z)$, $\kappa = (-K\hat{\beta}^T, 0^T)^T$ and $C = \omega^T W^{-1}\omega + \c{G}^+$ is full rank with the Moore-Penrose inverse '+'

$$\c{G} = \begin{pmatrix} \frac{I_p}{K} & 0 \\ 0 & G \end{pmatrix} \text{ and } \c{G}^+ = \begin{pmatrix} K & 0 \\ 0 & G^{-1} \end{pmatrix}.$$

After Equation (11) is found, we obtain

$$\hat{\Psi} = C^{-1}\omega^T W^{-1} y + C^{-1}\kappa, \tag{12}$$

where C^{-1} is calculated from the inverse partitioned matrix [18] as

$$C^{-1} = \begin{pmatrix} \acute{N} & -\acute{N}X^T H^{-1} ZG \\ -GZ^T H^{-1} X\acute{N} & \Upsilon + GZ^T H^{-1} X\acute{N}X^T H^{-1} ZG \end{pmatrix},$$

where $\acute{N} = (X^T H^{-1} X + K)^{-1}$ and $\Upsilon = (Z^T W^{-1} Z + G^{-1})^{-1}$. Then, after C^{-1} puts in Equation (12), the GKL estimator and the GKL predictor are derived, respectively, as

$$\begin{aligned} \hat{\beta}_{GKL} &= (X^T H^{-1} X + K)^{-1}(X^T H^{-1} y - K\hat{\beta}) = (X^T H^{-1} X + K)^{-1}(X^T H^{-1} X - K)\hat{\beta} \\ &= (I_p + K(X^T H^{-1} X)^{-1})^{-1}(I_p - K(X^T H^{-1} X)^{-1})\hat{\beta}, \end{aligned} \tag{13}$$

$$\hat{u}_{GKL} = GZ^T H^{-1}(y - X\hat{\beta}_{GKL}). \tag{14}$$

4. Mean Square Error Performances

Prediction of linear combinations of β and u is explained as $\mu = L^T\beta + M^T u$ for specific $L \in \mathbb{R}^{p\times 1}$ and $M \in \mathbb{R}^{q\times 1}$ matrices (see [16, 17, 21]). With the help of [19], the MMSEs for $\hat{\mu}$, $\hat{\mu}_{KL}$ and $\hat{\mu}_{GKL}$ are written as

$$MMSE(\hat{\mu}) = \mathbb{Q}MMSE(\hat{\beta})\mathbb{Q}^T + \sigma^2 M^T(G - GZ^T H^{-1} ZG)M, \tag{15}$$

$$MMSE(\hat{\mu}_{KL}) = \mathbb{Q}MMSE(\hat{\beta}_{KL})\mathbb{Q}^T + \sigma^2 M^T(G - GZ^T H^{-1} ZG)M, \tag{16}$$

$$MMSE(\hat{\mu}_{GKL}) = \mathbb{Q}MMSE(\hat{\beta}_{GKL})\mathbb{Q}^T + \sigma^2 M^T(G - GZ^T H^{-1} ZG)M, \tag{17}$$

where $\hat{\mu} = L^T\hat{\beta} + M^T\hat{u} = \mathbb{Q}\hat{\beta} + M^T GZ^T H^{-1}y$, $\hat{\mu}_{KL} = L^T\hat{\beta}_{KL} + M^T\hat{u}_{KL} = \mathbb{Q}\hat{\beta}_{KL} + M^T GZ^T H^{-1}y$, $\hat{\mu}_{GKL} = L^T\hat{\beta}_{GKL} + M^T\hat{u}_{GKL} = \mathbb{Q}\hat{\beta}_{GKL} + M^T GZ^T H^{-1}y$, $\mathbb{Q} = L^T - M^T GZ^T H^{-1}X$,

$$MMSE(\hat{\beta}) = \sigma^2(X^T H^{-1}X)^{-1}, \tag{18}$$

$$\begin{aligned} MMSE(\hat{\beta}_{KL}) = {} & \sigma^2(I_p + k(X^T H^{-1}X)^{-1})^{-1}(I_p - k(X^T H^{-1}X)^{-1})(X^T H^{-1}X)^{-1} \\ & \times(I_p - k(X^T H^{-1}X)^{-1})(I_p + k(X^T H^{-1}X)^{-1})^{-1} \\ & +[(I_p + k(X^T H^{-1}X)^{-1})^{-1}(I_p - k(X^T H^{-1}X)^{-1}) - I_p] \\ & \times\beta\beta^T[(I_p + k(X^T H^{-1}X)^{-1})^{-1}(I_p - k(X^T H^{-1}X)^{-1}) - I_p]^T, \end{aligned} \tag{19}$$

$$\begin{aligned} MMSE(\hat{\beta}_{GKL}) = {} & \sigma^2(I_p + K(X^T H^{-1}X)^{-1})^{-1}(I_p - K(X^T H^{-1}X)^{-1})(X^T H^{-1}X)^{-1} \\ & \times(I_p - K(X^T H^{-1}X)^{-1})(I_p + K(X^T H^{-1}X)^{-1})^{-1} \\ & +[(I_p + K(X^T H^{-1}X)^{-1})^{-1}(I_p - K(X^T H^{-1}X)^{-1}) - I_p] \\ & \times\beta\beta^T[(I_p + K(X^T H^{-1}X)^{-1})^{-1}(I_p - K(X^T H^{-1}X)^{-1}) - I_p]^T. \end{aligned} \tag{20}$$

When we examine Equations (15), (16) and (17), it can be said that the superiority of $MMSE(\hat{\mu}_{GKL})$ over $MMSE(\hat{\mu})$ and $MMSE(\hat{\mu}_{KL})$ is equivalent to the superiority of $MMSE(\hat{\beta}_{GKL})$ over $MMSE(\hat{\beta})$ and $MMSE(\hat{\beta}_{KL})$ derived by, respectively, Equations (18), (19) and (20). Then, via orthogonal transformation, our model (1) is transformed to a canonical form. Because H is pd, there exists a nonsingular symmetric matrix N such that $H = N^T N$. Our new model is

$$y^* = X^*\beta + Z^*u + \varepsilon^*, \tag{21}$$

with $y^* = N^{-1}y$, $X^* = N^{-1}X$, $Z^* = N^{-1}Z$, $\varepsilon^* = N^{-1}\varepsilon$ and $Var(y^*) = \sigma^2 I$ is derived.

The spectral decomposition of the matrix $X^T H^{-1}X$ is $P^T\Lambda P$ with $\Lambda = diag(\lambda_i)$ the $p\times p$ orthogonal matrix of the eigenvalues of $X^T H^{-1}X$ ($\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$) and $P = [P_1 \ldots P_p]$ the $p\times p$ orthogonal matrix of the standardized eigenvectors corresponding to the eigenvalues. Then, the model (21) can be written as $y^* = K^*\alpha + Z^*u + \varepsilon^*$, where $K^* = X^*P^T$ and $\alpha = P\beta$. In the transformed model, $MMSE(\tilde{\alpha}) = P[MMSE(\tilde{\beta})]P^T$ for any estimator $\tilde{\alpha}$ is derived. Hence, we have the following MMSE formulas via Equations (18), (19) and (20)

$$MMSE(\hat{\alpha}) = \sigma^2\Lambda^{-1}, \quad (22)$$

$$\begin{aligned} MMSE(\hat{\alpha}_{KL}) &= \sigma^2(I_p + k\Lambda^{-1})^{-1}(I_p - k\Lambda^{-1})\Lambda^{-1}(I_p - k\Lambda^{-1})(I_p + k\Lambda^{-1})^{-1} \\ &+[(I_p + k\Lambda^{-1})^{-1}(I_p - k\Lambda^{-1}) - I_p]\alpha\alpha^T[(I_p + k\Lambda^{-1})^{-1}(I_p - k\Lambda^{-1}) - I_p]^T, \end{aligned} \quad (23)$$

$$\begin{aligned} MMSE(\hat{\alpha}_{GKL}) &= \sigma^2(I_p + K\Lambda^{-1})^{-1}(I_p - K\Lambda^{-1})\Lambda^{-1}(I_p - K\Lambda^{-1})(I_p + K\Lambda^{-1})^{-1} \\ &+[(I_p + K\Lambda^{-1})^{-1}(I_p - K\Lambda^{-1}) - I_p]\alpha\alpha^T[(I_p + K\Lambda^{-1})^{-1}(I_p - K\Lambda^{-1}) - I_p]^T. \end{aligned} \quad (24)$$

We will define the two theorems given below, respectively, the GKL estimator vs the BLUE and the GKL estimator vs the KL estimator.

Theorem 4.1 $MMSE(\hat{\alpha}) - MMSE(\hat{\alpha}_{GKL}) > 0$ *iff*

$$\begin{aligned} &\alpha^T[(I_p + K\Lambda^{-1})^{-1}(I_p - K\Lambda^{-1}) - I_p]^T \\ &\times[\sigma^2(\Lambda^{-1} - (I_p + K\Lambda^{-1})^{-1}(I_p - K\Lambda^{-1})\Lambda^{-1}(I_p - K\Lambda^{-1})(I_p + K\Lambda^{-1})^{-1})] \\ &\times[(I_p + K\Lambda^{-1})^{-1}(I_p - K\Lambda^{-1}) - I_p]\alpha < 1. \end{aligned}$$

Theorem 4.2 $MMSE(\hat{\alpha}_{KL}) - MMSE(\hat{\alpha}_{GKL}) > 0$ *iff*

$$\begin{aligned} &\alpha^T[(I_p + K\Lambda^{-1})^{-1}(I_p - K\Lambda^{-1}) - I_p]^T[\Omega + [(I_p + k\Lambda^{-1})^{-1}(I_p - k\Lambda^{-1}) - I_p] \\ &\times\alpha\alpha^T[(I_p + k\Lambda^{-1})^{-1}(I_p - k\Lambda^{-1}) - I_p]^T][(I_p + K\Lambda^{-1})^{-1}(I_p - K\Lambda^{-1}) - I_p]\alpha < 1, \end{aligned}$$

where

$$\begin{aligned} \Omega = \; &\sigma^2((I_p + k\Lambda^{-1})^{-1}(I_p - k\Lambda^{-1})\Lambda^{-1}(I_p - k\Lambda^{-1})(I_p + k\Lambda^{-1})^{-1} \\ &-(I_p + K\Lambda^{-1})^{-1}(I_p - K\Lambda^{-1})\Lambda^{-1}(I_p - K\Lambda^{-1})(I_p + K\Lambda^{-1})^{-1}). \end{aligned}$$

[1] can be investigated for Theorems 4.1 and 4.2 proofs.

5. About Biasing Parameter Selection

Under our proposed new prediction approximation, an appropriate parameter k calculation is important. For this purpose, differentiating Equation (24) corresponding to k and then, equating to zero, we find

$$k_i = \frac{\sigma^2}{2\alpha_i^2 + (\sigma^2/\lambda_i)}, i = 1, \ldots, p, \quad (25)$$

Since the optimal value of k in Equation (25) depends on the unknown parameters σ^2 and α^2, we replace with their unbiased estimate and so, we have

$$\hat{k}_i = \frac{\hat{\sigma}^2}{2\hat{\alpha}_i^2 + (\hat{\sigma}^2/\lambda_i)}, i = 1, \dots, p, \tag{26}$$

and then, we introduce the minimum version of Equation (26) as

$$\hat{k}_{\min} = \min\left[\frac{\hat{\sigma}^2}{2\hat{\alpha}_i^2 + (\hat{\sigma}^2/\lambda_i)}\right]. \tag{27}$$

6. Gases of Greenhouse Data Example

Greenhouse gases have increased greatly in the last 150 years and the most important reason for this increase is human activities. The burning of fossil fuels for heat, transportation and electricity is the largest cause of gas emissions from these human activities [2]. The transportation sector receives the largest portion of greenhouse gas emissions from these three sectors in the United States. In this data example, we employ data on 297 fuel combustion in transport from randomly selected 27 areas for the years including 2006-2016 (see [2]). To identify fuel combustion in transport (y), repeated measurements are taken from the cars (x_1), the light duty trucks (x_2), the heavy duty trucks-buses (x_3), the motorcycles (x_4) and railways (x_5). The areas factor effect is random effect. Thus, our model is yielded

$$y_{ij} = \beta_1 x_{ij1} + \beta_2 x_{ij2} + \beta_3 x_{ij3} + \beta_4 x_{ij4} + \beta_5 x_{ij5} + u_1 + u_2 t_{ij} + \varepsilon_{ij},\ i = 1, \dots, 27,\ j = 1, \dots, 11,$$

where y_{ij} shows the ith observation of the jth area of the response, x_{ijs} shows the ith observation of the jth area of the explanatory variable x_s, $s = 1, \dots, 5$, t_{ij} denotes time corresponding to y_{ij}. In this example, we benefit from Matlab R2014a. Initially, we think covariance structures given below and then, for comparing these covariance models with ML and REML, we benefit from the Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC) (see Table 1).

Table 1: Covariance structures [1]

Cov. Struc.	Est. Met. for Cov. Par.	AIC	BIC
Unstructured (UN)	ML	**337.30**	374.24
	REML	**362.03**	398.76
Diagonal (UN(1))	ML	339.42	**372.67**
	REML	362.87	**395.93**
Variance Components (VC)	ML	391.56	421.11
	REML	416.72	446.11
Compound Symmetry (CS)	ML	393.42	426.67
	REML	418.60	451.66

The best models for modeling covariance matrix structure by response variable, which are the minimum values corresponding to AIC and BIC criteria, are the UN under AIC and UN(1) under BIC. By following [8] and [13]'s ideas, we choose UN(1) under ML and $\hat{G}_{ML} = \begin{bmatrix} 2.1913 & 0 \\ 0 & 0.0755 \end{bmatrix}$, $\hat{W}_{ML} = 0.25451 I_{297}$ are computed. Therefore, with $H = ZGZ^T + W$ formula, $\hat{H}_{ML}$ is derived. $X^T \hat{H}_{ML}^{-1} X$ matrix eigenvalues are computed as $(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5) = (1.4326 \times 10^{+7}, 1.5085 \times 10^{+4}, 4.7251 \times 10^{+3}, 247.7243, 41.5100)$. Since condition number $\lambda_{\max}/\lambda_{\min} = 345120 > 1000$ is obtained, one can say that severe multicollinearity is appeared.

To derive the GKL estimators/predictors, we get

$$K = diag(\hat{k}_i) = diag(1.03488, 5.56847, 6.80586, 9.04688, 0.10696), i = 1, \ldots, p,$$

by using Equation (26) and to get the KL estimators/predictors, we use $\hat{k} = \hat{k}_{\min} = 0.10696$ where $\hat{\sigma}^2$ is computed as 5.17298 given by Equation (27). In Table 2, fixed/random effects parameter estimates and scalar mean square error (SMSE) values are given. $\hat{\beta}_{GKL}$ outperforms $\hat{\beta}$ and $\hat{\beta}_{KL}$ in the sense of SMSE values under Table 2.

Table 2: Fixed/random effects parameter estimates and SMSE values

	β_1	β_2	β_3	β_4	β_5	$SMSE$		u_1	u_2
$\hat{\beta}$	1.02474	1.05007	0.93304	3.34361	3.67898	0.14693	$\hat{u}$	0.54883	−0.07806
$\hat{\beta}_{KL}$	1.02549	1.05044	0.93246	3.32847	3.65880	0.14599	$\hat{u}_{KL}$	0.54997	−0.07823
$\hat{\beta}_{GKL}$	1.03151	1.06769	0.89854	2.17688	3.65997	**0.05558**	$\hat{u}_{GKL}$	1.69062	−0.08354

Theorems 4.1 and 4.2 conditions are computed as, respectively, $0.01205 < 1$ and $0.01186 < 1$, hence $\hat{\beta}_{GKL}$ is also better than $\hat{\beta}$ and $\hat{\beta}_{KL}$ under the MMSE criterion.

Gases of greenhouse data example confirms that $\hat{\beta}_{GKL}$ is superior than $\hat{\beta}$ and $\hat{\beta}_{KL}$ when appropriate k values are employed.

[1]The abbreviations " Cov. Struc." and " Est. Met. for Cov. Par." refer to " Covariance Structures" and " Estimation Methods for Covariance Parameters".

7. Conclusion

The GKL prediction approach is extended to LMMs by using the method given in [1]. We also perform MMSE comparisons then, we give biasing parameter selection. Eventually, we support with our findings with gases of greenhouse data example.

This article presents that one can use the GKL estimator/predictor alternative to KL estimator/predictor in an LMM when multicollinearity problem exists and additionally, this article has affirmed that the GKL approach usage ensures a smaller MSE than the BLUE and KL estimator for appropriate selected ridge biasing parameter.

Declaration of Ethical Standards

The author declares that the materials and methods used in her study do not require ethical committee and/or legal special permission.

Conflicts of Interest

The author declares no conflict of interest.

References

[1] Dawoud I., Abonazel M.R., Awwad F.A., *Generalized Kibria-Lukman estimator: Method, simulation, and application*, Frontiers in Applied Mathematics and Statistics, 8, 1-6, 2022.

[2] Eurostat website, Greenhouse gas emissions by source sector (Source:EEA, env_air_gge), https://ec.europa.eu/eurostat/cache/metadata/en/env_air_gge_esms.htm, Accessed 2018.

[3] Gilmour A.R., Thompson R., Cullis B.R., *Average information REML: an efficient algorithm for variance parameter estimation in linear mixed models*, Biometrics, 51(4), 1440-1450, 1995.

[4] Henderson C.R., *Estimation of genetic parameters*, Annals of Mathematical Statistics, 21, 309-310, 1950.

[5] Henderson C.R., Kempthorne O., Searle S.R., Von Krosig C.N., *Estimation of environmental and genetic trends from records subject to culling*, Biometrics, 15(2), 192-218, 1959.

[6] Hoerl A.E., Kennard R.W., *Ridge regression: Biased estimation for nonorthogonal problems*, Technometrics, 12(1), 55-67, 1970.

[7] Kaçıranlar S., Sakallıoğlu S., Akdeniz F., Styan G.P.H., Werner H.J., *A new biased estimator in linear regression and a detailed analysis of the widely analyzed dataset on Portland cement*, Sankhya The Indian Journal of Statistics, 61(3), 443-459, 1999.

[8] Kass R.E., Raftery A.E., *Bayes factors*, Journal of the American Statistical Association, 90(430), 773-795, 1995.

[9] Kibria B.M.G., Lukman A.F., *A new ridge-type estimator for the linear regression model: Simulations and applications*, Scientifica, Article ID 9758378, 2020.

[10] Liu K.J., *A new class of biased estimate in linear regression*, Communications in Statistics-Theory and Methods, 22(2), 393-402, 1993.

[11] Liu X.-Q., Hu P., *General ridge predictors in a mixed linear model*, Statistics, 47(2), 363-378, 2013.

[12] Lukman A.F., Amin M., Kibria B.M.G., *K-L estimator for the linear mixed models: Computation and simulation*, Concurrency and Computation: Practice and Experience, 34(6), e6780, 2022.

[13] Özkale M.R., Can F., *An evaluation of ridge estimator in linear mixed models: An example from kidney failure data*, Journal of Applied Statistics, 44(12), 2251-2269, 2017.

[14] Özkale M.R., Kaçıranlar S., *The restricted and unrestricted two-parameter estimators*, Communications in Statistics-Theory and Methods, 36(15), 2707-2725, 2007.

[15] Özkale M.R., Kuran Ö., *A further prediction method in linear mixed models: Liu prediction*, Communications in Statistics - Simulation and Computation, 49(12), 3171-3195, 2020.

[16] Pereira L.N., Coelho P.S., *A small area predictor under area-level linear mixed models with restrictions*, Communications in Statistics-Theory and Methods, 41(13-14), 2524-2544, 2012.

[17] Robinson G.K., *That BLUP is a good thing: The estimation of random effects*, Statistical Science, 6(1), 15-51, 1991.

[18] Searle S.R., Matrix Algebra Useful for Statistics, John Wiley and Sons, 1982.

[19] Štulajter F., *Predictions in nonlinear regression models*, Acta Mathematica Universitatis Comenianae, 66(1), 71-81, 1997.

[20] Swindel B.F., *Good ridge estimators based on prior information*, Communications in Statistics-Theory and Methods, 5(11), 1065-1075, 1976.

[21] Yang H., Ye H., Xue K., *A further study of predictions in linear mixed models*, Communications in Statistics-Theory and Methods, 43(20), 4241-4252, 2014.