Journal of Data Science and Modeling, Vol. 3, No. 1, 95-122, December 2024

Research Manuscript

Diagnostic measures based on restricted ridge estimator in linear mixed measurement error models

Fatemeh Ghapani*[1],

[1]Department of Mathematics and Statistics, Sho.C., Islamic Azad University, Shoushtar, Iran.

Recieved: 22/06/2025 Accepted: 20/02/2026

Abstract:

This article focuses on diagnostic measures for identifying high-leverage points and influential observations in linear mixed measurement error (LMME) models. It achieves this by imposing stochastic restrictions on the parameters and incorporating the ridge estimator to address the issue of multicollinearity. To this end, generalized leverage matrices are defined using the restricted ridge estimator (RRE) to identify high-leverage observations. Additionally, analogs of Cook's distance and likelihood distance are proposed to determine influential observations through a case deletion approach. Simulation studies and real-life applications support the theoretical results.

Keywords: Case deletion model; Cook's distance; Diagnostics, Leverage points, Ridge estimator.

Mathematics Subject Classification (2010): 62J05, 62J20..

*Corresponding Author: Fatemeh.Ghapani@iau.ac.ir

1. Introduction

The linear mixed measurement error model can be expressed as:

$$\begin{aligned} y &= Z\beta + Ub + \varepsilon, \\ X &= Z + \Delta, \end{aligned} \tag{1}$$

where $y = (y_1', y_2', ..., y_l')'$ is an $n \times 1$ vector of observations, $Z = (Z_1', Z_2', ..., Z_l')'$ is an $n \times p$ matrix of regressors for the fixed effects, β is a $p \times 1$ parameter vector of fixed effects, and $U = [U_1|U_2|...|U_l]$ is an $n \times q$ known design matrix of the random effect factor, with U_i being $n \times q_i$, such that $\sum_{i=1}^{l} q_i = q$. $b' = (b_1', b_2', ..., b_l')$ is a $q \times 1$ unobservable vector of random effects from $N(0, \sigma^2\Sigma)$, where Σ is a block diagonal matrix with the ith block being $\gamma_i I_{q_i}$ for $\gamma_i = \sigma_i^2/\sigma^2$.

Furthermore, $\varepsilon = (\varepsilon_1', \varepsilon_2', ..., \varepsilon_l')'$ is an $n \times 1$ unobservable vector of random errors from $N(0, \sigma^2 I_n)$, and X is the observed version of Z with measurement error Δ, where Δ is an $n \times p$ random matrix from $MN(0, I_n \otimes \Lambda)$. Here, Λ is a $p \times p$ matrix of known values with nonnegative diagonal elements Fuller (1987). Additionally, it is assumed that b, ε, and Δ are mutually independent.

Under model (1), it follows that $y \sim MN(Z\beta, \sigma^2 V)$ with $V = I_n + U\Sigma U'$, and $b|y \sim N(\Sigma U'V^{-1}(y - Z\beta), \sigma^2\Sigma T)$, where $T = (I_q - U'V^{-1}U\Sigma) = (I_q + U'U\Sigma)^{-1}$. The corrected score estimates (CSE) of β and σ^2 are given by

$$\hat{\beta} = \left[X'V^{-1}X - \mathrm{tr}(V^{-1})\Lambda\right]^{-1} X'V^{-1}y, \quad \hat{\sigma}^2 = \frac{(y - X\hat{\beta})'V^{-1}(y - X\hat{\beta}) - \mathrm{tr}(V^{-1})\hat{\beta}'\Lambda\hat{\beta}}{n},$$

where

$$\hat{\sigma}_i^2 = \frac{\hat{b}_i'\hat{b}_i - \mathrm{tr}(\hat{D}_i'\hat{D}_i)\hat{\beta}'\Lambda\hat{\beta}}{q_i - \mathrm{tr}(T_{ii})}, \quad i = 1, ..., l,$$

with $\hat{D}_i = \hat{\gamma}_i U_i' V^{-1} = \frac{\hat{\sigma}_i^2}{\hat{\sigma}^2} U_i' V^{-1}$, and $\hat{b} = \Sigma U'V^{-1}(y - X\hat{\beta})$ is the predicted random effects. Here, T_{ij} denotes the ijth block of matrix T Zhong *et al.* (2002); Zare *et al.* (2012).

The presence of multicollinearity in linear regression models leads to higher variance and unstable parameter estimates when using ordinary least squares. To address this, several biased estimators have been developed, including the Stein estimator Stein (1956), ridge regression Hoerl and Kennard (1970), and the Liu estimator Liu (1993).

To overcome multicollinearity, Liu and Hu (2013) introduced methods, and Ozkale and Can (2017) proposed the ridge estimator and ridge predictor in linear mixed models when fixed-effect variables have no measurement error. Ganjealivand *et al.* (2021) examined stochastic restricted and unrestricted two-parameter estimators for fixed and random effects in linear mixed measurement error models. Ghapani (2022) focused on stochastic restricted Liu estimation for parameters in LMME models with multicollinearity. Yavarizadeh *et al.* (2022) introduced ridge estimation in LMME models with stochastic linear mixed restrictions. Ganjealivand and Ghapani (2024) presented a weighted two-parameter estimator for predicting fixed and random effects in LMME models using additional linear stochastic constraints.

Since influential observations and multicollinearity often co-occur in LMME models, it is important to consider them in data analysis. Various methods, including residuals and case deletion models (CDM), have been proposed for this purpose. Diagnostic measures for linear mixed models are discussed in Christensen *et al.* (1992); Banerjee and Frees (1997); Zhong and Wei (1999); Haslett and Dillane (2004); Zewotir and Galpin (2005); Li *et al.* (2009). Fung *et al.* (2003) studied estimation and influence diagnostics in LMME models, while Zare and Rasekh (2012) introduced case deletion and mean-shift outlier models using Nakamura (1990) corrected likelihood. Zare and Rasekh (2014) derived residuals and leverage in LMME models. Maksaei *et al.* (2023) focused on ridge-based diagnostic methods, and Borhani *et al.* (2023, 2024) investigated influential and outlier detection with Liu's corrected likelihood estimator under multicollinearity. To our knowledge, little attention has been given to leverage and influence diagnostics for RRE outcomes in LMME models. This paper evaluates the influence of observations on RREs of fixed effects and predicted random effects.

The remainder of this article is organized as follows. Section 2 discusses parameter estimation in LMME models via the stochastic restricted ridge method and examines asymptotic properties. Section 3 introduces influence measures for detecting influential observations in ridge LMME models with stochastic linear restrictions and employs a parametric bootstrap to generate empirical distributions of test statistics. Section 4 illustrates the proposed diagnostics through a numerical example, followed by simulation results in Section 5. Concluding remarks are presented in Section 6.

2. Stochastic Restricted Ridge Estimator

In many applications, additional or prior information about the unknown parameter vector β is available. This information may come from theoretical considerations or previous studies. Suppose r is an $m \times 1$ observable random vector, R is a known $m \times p$ matrix of rank $m < p$, and e is an $m \times 1$ vector of unobservable random errors from $N(0, \sigma^2 W)$, where W is a positive definite matrix of known elements. In addition to model (1), consider the stochastic linear restrictions:

$$r = R\beta + e. \tag{2}$$

Assume that e is stochastically independent of ε and Δ. Combining the information in (2) with model (1) leads to the following mixed model:

$$y_r = Z_r\beta + U_r b + \varepsilon_r, \tag{3}$$

or

$$\begin{bmatrix} y \\ r \end{bmatrix} = \begin{bmatrix} Z \\ R \end{bmatrix}\beta + \begin{bmatrix} U \\ 0 \end{bmatrix} b + \begin{bmatrix} \varepsilon \\ e \end{bmatrix}$$

where b and y_r are jointly distributed as $\begin{bmatrix} b \\ y_r \end{bmatrix} \sim N(\begin{bmatrix} 0 \\ Z_r\beta \end{bmatrix}, \begin{bmatrix} \sigma^2\Sigma & \sigma^2\Sigma U'_r \\ \sigma^2 U_r\Sigma & \sigma^2 V_r \end{bmatrix})$ where $V_r = \begin{bmatrix} V & 0 \\ 0 & W \end{bmatrix}$. Then the conditional distribution of b given y_r is $b|\, y_r \sim N(\Sigma U_r' V_r{}^{-1}(y_r - Z_r\beta), \sigma^2\Sigma T_r)$, where $T_r = I_q - U_r' V_r^{-1} U_r\Sigma = (I_q + U_r' U_r\Sigma)^{-1}$. According to Ghapani (2022) the corrected stochastic restricted estimates of β , σ^2 and the predictor of b are given by

$$\begin{aligned}
&\hat{\beta}_r = A_r^{-1}(X'V^{-1}y + R'W^{-1}r) \\
&\hat{\sigma}_r^2 = \tfrac{(y - X\hat{\beta}_r)'V^{-1}(y - X\hat{\beta}_r) + (r - R\hat{\beta}_r)'W^{-1}(r - R\hat{\beta}_r) - tr(V^{-1})\hat{\beta}_r'\Lambda\hat{\beta}_r}{(n+m)} \quad , \\
&\hat{b}_r = \Sigma U'V^{-1}(y - X\hat{\beta}_r)
\end{aligned}$$

where $A_r = X'V^{-1}X + R'W^{-1}R - \mathrm{tr}(V^{-1})\Lambda$. To address the multicollinearity problem, the ridge estimator is employed in LMME models. For this purpose, consider the stochastic linear restriction

$$0 = \sqrt{k} I_p\beta + \varphi, \quad \varphi \sim N(0, \sigma^2 I_p),$$

where $k > 0$ is the ridge biasing parameter.

To incorporate the ridge restriction into the parameter estimation, the log-likelihood function and the corrected log-likelihood function for the joint distribution of y_r and b are defined as

$$\begin{aligned}
&l(\theta; Z, y, r) = -\tfrac{1}{2}\left\{(n + q + m)\log(2\pi\sigma^2) + \log|\Sigma|\right\} \\
&- \tfrac{1}{2\sigma^2}\left[(y - Z\beta - Ub)'(y - Z\beta - Ub) + b'\Sigma^{-1}b\right] \\
&- \tfrac{1}{2\sigma^2}\left[(r - R\beta)'W^{-1}(r - R\beta)\right] - \tfrac{k\beta'\beta}{2\sigma^2},
\end{aligned}$$

$$\begin{aligned}
&l^*(\theta; X, y, r) = -\tfrac{1}{2}\left\{(n + q + m)\log(2\pi\sigma^2) + \log|\Sigma|\right\} \\
&- \tfrac{1}{2\sigma^2}\left[(y - X\beta - Ub)'(y - X\beta - Ub) - tr(V^{-1})\beta'\Lambda\beta + b'\Sigma^{-1}b\right] \\
&- \tfrac{1}{2\sigma^2}\left[(r - R\beta)'W^{-1}(r - R\beta)\right] - \tfrac{k\beta'\beta}{2\sigma^2},
\end{aligned}$$

where $\theta = (\beta, b, \gamma, k)$ and $\gamma' = (\gamma_1, \gamma_2, ..., \gamma_l)$. The $l^*(\theta; X, y, r)$ have the following properties (Nakamura, 1990):
$E^*\left[\frac{\partial l^*(\theta;X,y,r)}{\partial\beta}\right] = \frac{\partial l(\theta;Z,y,r)}{\partial\beta}$ and $E^*\left[\frac{\partial l^*(\theta;X,y,r)}{\partial b}\right] = \frac{\partial l(\theta;Z,y,r)}{\partial b}$
where E^* denotes the conditional expectation with respect to X given y. The corrected score restricted ridge estimates of β and b are obtained by differentiating $l^*(\theta; X, y, r)$ with respect to β and b, respectively, as follows:

$$\begin{aligned}\hat{\beta}_{rk} &= A_{rk}^{-1}\left(X'V^{-1}y + R'W^{-1}r\right),\\ \hat{b}_{rk} &= \Sigma U'V^{-1}(y - X\hat{\beta}_{rk}),\end{aligned} \tag{4}$$

where $A_{rk} = A_r + kI_p$.

The marginal corrected log-likelihood function in LMME models with the ridge condition can be written as:

$$\begin{aligned}l^*(\theta'; X, y, r) = &-\tfrac{n+m}{2}\log(2\pi\sigma^2) - \tfrac{1}{2}\log|V|\\ &-\tfrac{1}{2\sigma^2}\left[(y - X\beta)'V^{-1}(y - X\beta) - \text{tr}(V^{-1})\beta'\Lambda\beta\right]\\ &-\tfrac{1}{2\sigma^2}\left[(r - R\beta)'W^{-1}(r - R\beta)\right] - \tfrac{k\beta'\beta}{2\sigma^2},\end{aligned}$$

where $\theta' = (\sigma^2, \gamma, k)$.

Let $l_1(\theta'; Z, y, r) = l(\tilde{\beta}(\gamma), \tilde{b}(\gamma), \theta'; Z, y, r)$, where $\tilde{\beta}(\gamma)$ and $\tilde{b}(\gamma)$ are the maximum likelihood estimates of β and b. Also, define $l_1^*(\theta'; X, y, r) = l^*(\hat{\beta}(\gamma), \hat{b}(\gamma), \theta'; X, y, r)$, where $\hat{\beta}(\gamma)$ and $\hat{b}(\gamma)$ are the corrected restricted ridge likelihood estimates of β and b.

The corrected log-likelihood $l^*(\theta'; X, y, r)$ should satisfy
$E^*\left[\frac{l_1^*(\theta';X,y,r)}{\partial\sigma^2}\right] = \frac{\partial l_1(\theta';Z,y,r)}{\partial\sigma^2}$ and $E^*\left[\frac{l_1^*(\theta';X,y,r)}{\partial\gamma_i}\right] = \frac{\partial l_1(\theta';Z,y,r)}{\partial\gamma_i}$.

By solving the equation $\partial l^*(\theta'; X, y, r)/\partial\sigma^2 = 0$, the corrected restricted ridge estimator of σ^2 is given by

$$\begin{aligned}\hat{\sigma}^2_{rk} = \tfrac{1}{n+m}\Big[&(y - X\hat{\beta}_{rk})'V^{-1}(y - X\hat{\beta}_{rk}) - tr(V^{-1})\hat{\beta}'_{rk}\Lambda\hat{\beta}_{rk}\\ &+ (r - R\hat{\beta}_{rk})'W^{-1}(r - R\hat{\beta}_{rk}) + k\hat{\beta}'_{rk}\hat{\beta}_{rk}\Big].\end{aligned}$$

If the elements of γ_i are unknown, the RRE of the unknown parameters are substituted back into Σ to obtain $\hat{\beta}_{rk}$, $\hat{\sigma}^2_{rk}$, and $\hat{b}_{rk}$. According to Zare *et al.* (2012), for the estimation of the γ_i's, the corrected score estimates of $\sigma^2_{rk1}, \ldots, \sigma^2_{rkl}$ are used, as
$\hat{\sigma}^2_{rki} = \frac{\hat{b}_{rki}\hat{b}_{rki} - tr(\hat{D}'_{rki}\hat{D}_{rki})\hat{\beta}'_{rk}\Lambda\hat{\beta}_{rk}}{[q_i - tr(T_{ii})]}$; $i = 1, ..., l$ with $\hat{D}_{rki} = \hat{\gamma}_{rki}U_i'V^{-1} = \frac{\hat{\sigma}^2_{rki}}{\hat{\sigma}^2_{rk}}U_i'V^{-1}$.

2.1 Asymptotic properties of restricted ridge estimator

To investigate the asymptotic behavior of the estimators, the theory of asymptotic approximation for large samples is used to study their asymptotic distribution. It is assumed that the parameter β is identifiable, and as n tends to infinity, the limits

$n^{-1}(Z'V^{-1}Z + R'W^{-1}R)$ and $n^{-1}(Z'V^{-1}Z + R'W^{-1}R + kI_p)$ exist and E denotes the global expectation taken at the true value .

Theorem 2.1. *The asymptotic distribution of* $\sqrt{n}(\hat{\beta}_{rk} - G_r G_{rk}^{-1}\beta)$ *is normal with mean vector zero and covariance matrix* $AVar(\hat{\beta}_{rk}) = G_{rk}^{-1}(B + \sigma^2 G_r)G_{rk}^{-1}$ *, where* $B = \left[\sigma^2 tr(V^{-1}) + \beta' Z' V^{-2} Z\beta\right]\Lambda$ *,* $G_{rk} = Z'V^{-1}Z + R'W^{-1}R + kI_p$ *and* $G_r = G_{rk(k=0)}$.

Proof. Let $\xi = n^{-\frac{1}{2}}(X'V^{-1}y + R'W^{-1}r)$, so we obtain the asymptotic properties of ξ . It follows from $E(X'V^{-1}y + R'W^{-1}r) = (Z'V^{-1}Z + R'W^{-1}R)\beta$, that $E(\xi) = n^{-\frac{1}{2}}G_r\beta$. The variance of ξ can be obtained by

$$\begin{aligned} Var(\xi) &= E\left[Var(\xi\,|y)\right] + Var\left[E(\xi\,|y)\right] \\ &= n^{-1}E(y'V^{-2}y\Lambda) + n^{-1}Var(Z'V^{-1}y + R'W^{-1}r), \end{aligned}$$

since, $E(y'V^{-2}y) = \sigma^2 tr(V^{-1}) + \beta' Z'V^{-2}Z\beta$ and $Var(Z'V^{-1}y + R'W^{-1}r) = \sigma^2 G_r$, therefore, $Var(\xi) = n^{-1}(B + \sigma^2 G_r)$. Also, since the

$$E(X'V^{-1}X) = Z'V^{-1}Z + tr(V^{-1})\Lambda$$

, by Fung et al. (2003), can be written, $X'V^{-1}X = Z'V^{-1}Z + tr(V^{-1})\Lambda + O_p(n^{\frac{1}{2}})$. Then,

$$n^{-1}A_{rk} = n^{-1}(Z'V^{-1}Z + R'W^{-1}R + kI_p) + O_p(n^{-\frac{1}{2}}),$$

so, it follows from $\hat{\beta}_{rk}$

$$\begin{aligned} \sqrt{n}\hat{\beta}_{rk} &= \left[n^{-1}A_{rk} + O_p(n^{-\frac{1}{2}})\right]^{-1} n^{-\frac{1}{2}}(X'V^{-1}y + R'W^{-1}r) \\ &= \left[I_p + O_p(n^{-\frac{1}{2}})\right]^{-1}(n^{-1}G_{rk})^{-1}\xi = \left[I_p + O_p(n^{-\frac{1}{2}})\right](n^{-1}G_{rk})^{-1}\xi, \end{aligned}$$

where $\left[I_p + O_p(n^{-\frac{1}{2}})\right]^{-1} = I_p + O_p(n^{-\frac{1}{2}})$ is obtained from Taylor series expansion. Since the limit of $C = n^{-1}G_{wrk}$ exists, then can be written

$$\sqrt{n}\hat{\beta}_{rk} = C^{-1}\xi + O_p(n^{-\frac{1}{2}}),$$

Consequently, it is asymptotically concluded that

$$\sqrt{n}(\hat{\beta}_{rk} - G_{rk}^{-1}G_r\beta) = C^{-1}\left[\xi - E(\xi)\right] + O_p(n^{-\frac{1}{2}})$$

that $\sqrt{n}(\hat{\beta}_{rk} - G_{rk}^{-1}G_r\beta)$ is asymptotically normal with mean zero. Furthermore, It can be concluded that

$AVar(\sqrt{n}\hat{\beta}_{rk}) = C^{-1}Var(\xi)C^{-1}$. Thus $AVar(\hat{\beta}_{rk}) = G_{rk}^{-1}(B + \sigma^2 G_r)G_{rk}^{-1}$, this completes the proof of Theorem 2.1.

□

Corollary 2.2. *$\hat{\beta}$ has an asymptotically normal distribution with mean $AE(\hat{\beta}) = \beta$ and covariance matrix $AVar(\hat{\beta}) = (Z'V^{-1}Z)^{-1}(\mathrm{B} + \sigma^2 Z'V^{-1}Z)(Z'V^{-1}Z)^{-1}$.*

Corollary 2.3. *$\hat{\beta}_r$ is asymptotically normally distribution with mean $E(\hat{\beta}_r) = \beta$ and $AVar(\hat{\beta}_r) = G_r^{-1}(B + \sigma^2 G_r)G_r^{-1}$.*

Corollary 2.4. *$\hat{\beta}_k$ is asymptotically normally distribution with mean $E(\hat{\beta}_k) = G_k^{-1}G\beta$ and $AVar(\hat{\beta}_k) = G_k^{-1}(B + \sigma^2 G)G_k^{-1}$,where $G_k = G_{rk(R=0)}$ and $G = G_{k(k=0)}$.*

2.2 Mean Square Error Matrix Comparisons

One of the criteria used to evaluate the performance of estimators is the mean squared error (MSE) matrix criterion. The mean-square error matrix (MSEM) for any estimator $\hat{\beta}$ of β is defined as

$$MSEM(\hat{\beta}) = \mathrm{Var}(\hat{\beta}) + \mathrm{Bias}(\hat{\beta})\,\mathrm{Bias}(\hat{\beta})',$$

where $\mathrm{Bias}(\hat{\beta})$ denotes the bias vector.

Another criterion for evaluating an estimator is the MSE value, which is obtained as follows:

$$MSE(\hat{\beta}) = tr\left[MSEM(\hat{\beta})\right] = tr\left[Var(\hat{\beta})\right] + Bias(\hat{\beta})'Bias(\hat{\beta})$$

Definition 2.5. *When two estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ are given, the estimator $\hat{\beta}_2$ is said to superior to in the MSEM sense if and only if $\Delta = MSEM(\hat{\beta}_1) - MSEM(\hat{\beta}_2)$ $\Delta \geq 0$. If Δ is positive definite (p.d.) matrix, $\hat{\beta}_2$ is said to be strongly superior to $\hat{\beta}_1$, i.e. $\Delta > 0$.*

The asymptotic MSEM of the estimators $\hat{\beta}$,$\hat{\beta}_k$, $\hat{\beta}_r$ and $\hat{\beta}_{rk}$ obtained as follows:

$$MSEM(\hat{\beta}) = G^{-1}(B + \sigma^2 G)G^{-1}$$

,

$$MSEM(\hat{\beta}_k) = G_k^{-1}(B + \sigma^2 G)G_k^{-1} + b_1 b_1'$$

$$MSEM(\hat{\beta}_r) = G_r^{-1}(B + \sigma^2 G_r)G_r^{-1}$$

$$MSEM(\hat{\beta}_{rk}) = G_{rk}^{-1}(B + \sigma^2 G_r)G_{rk}^{-1} + b_2 b_2'$$

where, $b_1 = -kG_k^{-1}\beta$ and $b_2 = -kG_{rk}^{-1}\beta$.

3. Influence Measures in restricted ridge estimator

Sometimes, a small subset of data exerts a disproportionate influence on the model coefficients and properties. Therefore, the general objective is to account for influential points in data analysis. Various methods exist to identify these points. In the following subsections, some of these methods are presented.

3.1 Residuals

Model (2) can be written as

$$y = X\beta + Ub + v,$$

where $v = \varepsilon - \Delta\beta$ and v_i is the ith element of the vector v. Using the RR estimators, the conditional residuals are given by

$$\begin{aligned}\hat{v}_{rk} &= y - X\hat{\beta}_{rk} - U\hat{b}_{rk} = V^{-1}(y - X\hat{\beta}_{rk})\\ &= (V^{-1} - V^{-1}XA_{rk}^{-1}X'V^{-1})y + V^{-1}R'W^{-1}r\\ &= Py + V^{-1}XA_{rk}^{-1}R'W^{-1}r,\end{aligned}$$

where $C = V^{-1} = \begin{bmatrix} c_{ii} & c'_{i(i)} \\ c_{i(i)} & V_{[i]}^{-1} + c_{i(i)}c'_{i(i)}/c_{ii} \end{bmatrix}$, $c'_{i(i)}$ denotes the ith row of V^{-1} with the ith element removed and $P = V^{-1} - V^{-1}XA_{rk}^{-1}X'V^{-1}$. The ith standardized residual is $t_i = \hat{v}_{rki}/\sqrt{\hat{v}'_{rk}\hat{v}_{rk}}$ and $\hat{v}_{rki} = y_i - x_i'\hat{\beta}_{rki} - u_i'\hat{b}_{rki}$ is the ith element of the vector of $\hat{v}_{rk}$. Based on Zare and Rasekh (2014) the asymptotic distribution of $t_i^{*2} = \frac{(n-1)t_i^2}{1-t_i^2}$ is Fisher‘s distribution with 1 and $n-1$ degree of freedom. The ith observation may be considered an outlier if $t_i^{*2} > F(1, n-1, \alpha)$, where $F(1, n-1, \alpha)$ denotes the upper percentile of the Fisher distribution with 1 and $n-1$ degrees of freedom. The vector of fitted values is

$$\begin{aligned}\hat{y} &= X\hat{\beta}_{rk} + U\hat{b}_{rk} = y - V^{-1}y + V^{-1}XA_{rk}^{-1}X'{}_rV_r^{-1}y_r\\ &= (I_n - V^{-1} + V^{-1}XA_{rk}^{-1}X'V^{-1})y + V^{-1}XA_{rk}^{-1}R'W^{-1}r\ ,\\ &= Hy + V^{-1}XA_{rk}^{-1}R'W^{-1}r\end{aligned}$$

where, $H = I_n - P$.

3.2 Studentized residuals

Since the residuals have different variances, they are not directly comparable. Accordingly, studentized residuals are used as a more appropriate criterion for detecting outliers. The ith studentized residual of the model is defined as:

$$s_{rki} = \frac{\hat{v}_{rki}}{\hat{\sigma}_v \sqrt{p_{ii}}},$$

where

$$\hat{\sigma}_v^2 = \hat{\sigma}_{rk}^2 + \hat{\beta}'_{rk}\Lambda\hat{\beta}_{rk},$$

and p_{ii} is the ith diagonal element of P. Studentized residuals are more effective than standardized residuals for detecting outlying observations. If an observation has a large s_{rki}^2, it is considered an outlier.

3.3 Generalized Leverage

Following Wei *et al.* (1998), the generalized leverage matrix for the fixed effects in LMME models with the RR estimator is defined as the partial derivative of the marginal fitted values with respect to the response values, i.e.,

$$GL(\hat{\beta}_{rk}) = \frac{\partial X\hat{\beta}_{rk}}{\partial y'} = XA_{rk}^{-1}X'V^{-1}.$$

Therefore, the generalized leverage for the ith observation is

$$GL_{ii}(\hat{\beta}_{rk}) = x_i' A_{rk}^{-1} X' c_i,$$

where c_i' is the ith row of the matrix V^{-1}. Based on Hoaglin and Welsch (1978), when

$$GL_{ii}(\hat{\beta}_{rk}) > \frac{3 \times \mathrm{tr}(GL(\hat{\beta}_{rk}))}{n},$$

the ith observation is considered to have high leverage on the fixed effects.

Since in linear mixed models one observation can affect both the estimation of fixed effects and the predicted values of random effects, it is possible to evaluate the joint effect of each observation on both. The generalized leverage matrix in LMME models with RR estimation of fixed and random effects is defined as

$$GL(\hat{\beta}_{rk}, \hat{b}_{rk}) = \frac{\partial \hat{y}}{\partial y'} = H.$$

The ith diagonal element of H,

$$h_{ii} = 1 - c_{ii} + c_i' X A_{rk}^{-1} X' c_i,$$

where c_{ii} denotes the ith diagonal element of V^{-1}, represents the leverage of the response value y_i on the corresponding fitted value $\hat{y}_i$. According to Hoaglin and Welsch (1978), the ith observation is said to have high leverage on both β and b if

$$h_{ii} > \frac{3 \times \mathrm{tr}(H)}{n}.$$

Additionally, based on Wei *et al.* (1998), the generalized leverage matrix for the random effects with ridge estimation in LMME models is defined as

$$GL(\tilde{b}_{rk}) = I_n - V^{-1}.$$

As before, the ith observation is said to have high leverage on b if

$$GL_{ii}(\hat{b}_{rk}) > \frac{3 \times \mathrm{tr}(GL(\hat{b}_{rk}))}{n}.$$

3.4 Case Deletion Diagnostics

The aim of analyzing influential observations is to evaluate the impact of the ith observation on the estimation of parameters. There are different approaches to assess the influence of perturbations in a dataset and in the model given the estimated parameters. Case-deletion diagnostics is an example of global influence analysis, which assesses the effect of an observation by completely removing it. The matrices are rearranged so that the ith deleted case is placed in the first row. Therefore,

$$y = \begin{bmatrix} y_i \\ y_{(i)} \end{bmatrix}, X = \begin{bmatrix} x'_i \\ X_{(i)} \end{bmatrix},\ Z = \begin{bmatrix} z'_i \\ Z_{(i)} \end{bmatrix} \text{ and } U = \begin{bmatrix} u_i \\ U_{(i)} \end{bmatrix}.$$

The restricted ridge case deletion model with the ith observation deleted is defined as

$$\begin{aligned} &y_{(i)} = Z_{(i)}\beta + U_{(i)}b + \varepsilon_{(i)}, X_{(i)} = Z_{(i)} + \Delta_{(i)}, i = 1, 2, ...n., with \\ &r = R\beta + e and 0 = \sqrt{k}\mathrm{I}_p\beta + e. \end{aligned} \tag{5}$$

The corrected log-likelihood function joint of $y_{(i)}$ and b and the marginal corrected log-likelihood function of $y_{(i)}$, respectively are define as

$$\begin{aligned} &l_i^*(\theta; X, y, r) = -\tfrac{1}{2}\left\{(n+q+m-1)\log(2\pi\sigma^2) + \log|\Sigma|\right\} \\ &-\tfrac{1}{2\sigma^2}\left[(y_{(i)} - X_{(i)}\beta - U_{(i)}b)'(y_{(i)} - X_{(i)}\beta - U_{(i)}b) - tr(V_{[i]}^{-1})\beta'\Lambda\beta + b'\Sigma^{-1}b\right] \\ &-\tfrac{1}{2\sigma^2}\left[(r - R\beta)'W^{-1}(r - R\beta)\right] - \tfrac{k\beta'\beta}{2\sigma^2}, \\ &l_i^*(\theta'; X, y, r) = -\tfrac{n+m-1}{2}\log(2\pi\sigma^2) - \tfrac{1}{2}\log\left|V_{[i]}\right| \\ &-\tfrac{1}{2\sigma^2}\left[(y_{(i)} - X_{(i)}\beta)'V_{[i]}^{-1}(y_{(i)} - X_{(i)}\beta) - tr(V_{[i]}^{-1})\beta'\Lambda\beta\right] \\ &-\tfrac{1}{2\sigma^2}\left[(r - R\beta)'W^{-1}(r - R\beta)\right] - \tfrac{k\beta'\beta}{2\sigma^2}, \end{aligned}$$

The corrected score estimates of β and b will be obtained with differentiating of $l_i^*(\beta, k; X, y, b)$ with respect to β and b . Then we have

$$\begin{aligned} &\tfrac{\partial l_i^*(\beta,k;X,y,b)}{\partial\beta} = 0 \\ &A_{rk(i)}\hat{\beta}_{rk(i)} + X'_{(i)}U_{(i)}\tilde{b}_{rk(i)} = X'_{(i)}y_{(i)} + R'W^{-1}r \end{aligned}$$

where, $A_{rk(i)} = (X'_{(i)}X_{(i)} + R'W^{-1}R - tr(V_{[i]}^{-1})\Lambda + kI_p),$

$$\frac{\partial l_i^*(\beta, k; X, y, b)}{\partial b} = 0$$

$$\begin{aligned} &U'_{(i)}y_{(i)} - U'_{(i)}X_{(i)}\hat{\beta}_{rk(i)} - U'_{(i)}U_{(i)}\hat{b}_{rk(i)} - \Sigma^{-1}\hat{b}_{rk(i)} = 0 \\ &\hat{b}_{rk(i)} = (U'_{(i)}U_{(i)} + \Sigma^{-1})^{-1}U'_{(i)}(y_{(i)} - X_{(i)}\hat{\beta}_{rk(i)}), \end{aligned} \quad (6)$$

substituting $\hat{b}_{k(i)}$ in to the first equation of (6) gives

$$A_{rk(i)}\hat{\beta}_{rk(i)} + X'_{(i)}U_{(i)}(U'_{(i)}U_{(i)} + \Sigma^{-1})^{-1}U'_{(i)}\left[y_{(i)} - X_{(i)}\hat{\beta}_{rk(i)}\right] = X'_{(i)}y_{(i)},$$

using $V_{[i]}^{-1} = (I_{n-1} + U_{(i)}\Sigma U'_{(i)})^{-1} = I_{n-1} - U_{(i)}(U'_{(i)}U_{(i)} + \Sigma^{-1})^{-1}U'_{(i)}$ and

$$(U'_{(i)}U_{(i)} + \Sigma^{-1})^{-1}U'_{(i)} = \Sigma U'_{(i)}V_{[i]}^{-1},$$

we obtain

$$\hat{\beta}_{rk(i)} = A_{rk(i)}^{-1}(X'_{(i)}V_{[i]}^{-1}y_{(i)} + R'W^{-1}r)$$

$$\hat{b}_{rk(i)} = \Sigma U'_{(i)}V_{[i]}^{-1}(y_{(i)} - X_{(i)}\hat{\beta}_{rk(i)})$$

Taking the differential of $l_i^*(\sigma^2, k; X, y)$ with respect to σ^2 , it follows that

$$\begin{aligned} (n+m-1)\hat{\sigma}^2_{rk(i)} = \Big[&(y_{(i)} - X_{(i)}\hat{\beta}_{rk(i)})'V_{[i]}^{-1}(y_{(i)} - X_{(i)}\hat{\beta}_{rk(i)}) - tr(V_{[i]}^{-1})\hat{\beta}'_{rk(i)}\Lambda\hat{\beta}_{rk(i)} \\ &+ (r - R\hat{\beta}_{rk(i)})'W^{-1}(r - R\hat{\beta}_{rk(i)}) + k\hat{\beta}'_{rk(i)}\hat{\beta}_{rk(i)}\Big] \end{aligned}$$

Theorem 3.1. *For model (5), the following results are obtained:*

$$\hat{\beta}_{rk(i)} = \hat{\beta}_{rk} - A_{rk(i)}^{-1}X'c_i\frac{\hat{v}_{rki}}{p_{ii}} + O_p(n^{-1}),$$

$$\hat{\sigma}^2_{rk(i)} = \frac{n\hat{\sigma}^2_{rk} - (\hat{v}^2_{rki}/p_{ii})}{n+m-1} + O_p(n^{-1}),$$

$$\hat{b}_{rk(i)} = \Sigma U'_{(i)}V_{[i]}^{-1}\left[y_{(i)} - X_{(i)}\hat{\beta}_{rk(i)}\right] = \hat{b}_{rk} - \Sigma U'p_i\frac{\hat{v}_{rki}}{p_{ii}}$$

where, c_i' and p_i' are ith rows of V^{-1} and P , respectively. Proof. Given in Appendix.

3.5 Distance Measure: Cook's distance and Likelihood distance

3.5.1 Cook's distance for fixed effects

To assess changes in the estimated parameter vector, Cook's distance, based on Cook (1977), for the deletion of the ith observation in LMME models with ridge estimation is defined as

$$CD_i(\beta) = (\hat{\beta}_{rk} - \hat{\beta}_{rk(i)})' M (\hat{\beta}_{rk} - \hat{\beta}_{rk(i)})$$

where $M = \hat{\sigma}_{rk}^{-2} A_{rk}$. Large values of $CD_i(\beta)$ indicate that the ith observation has a substantial effect on the full-sample estimate. Therefore, it is concluded that:

$$\begin{aligned} CD_i(\beta) &= \hat{\sigma}_{rk}^{-2} c'_i X A_{rk}^{-1} X' c_i \frac{\hat{v}_{rki}^2}{p_{ii}^2} + O_p(n^{-1}) \\ &= \hat{\sigma}_{rk}^{-2} \frac{(c_{ii} - p_{ii}) \hat{v}_{rki}^2}{p_{ii}^2} + O_p(n^{-1}) \end{aligned}$$

3.5.2 Cook's distance for random effects

A convenient measure of influence for random effects in LMME models with ridge estimation is based on the difference between two estimators: one that includes the ith observation in the data set, and the other that excludes it. The Cook's distance for random effects is defined as

$$CD_i(b) = (\hat{b}_{rk} - \hat{b}_{rk(i)})' G (\hat{b}_{rk} - \hat{b}_{rk(i)})$$

where, $G = \hat{\sigma}_{rk}^{-2} (U'U + \Sigma^{-1})$.

$$\begin{aligned} CD_i(b) &= (\Sigma U' p_i \frac{\hat{v}_{rki}}{p_{ii}})' \hat{\sigma}_k^{-2} (U'U + \Sigma^{-1}) (\Sigma U' p_i \frac{\hat{v}_{rki}}{p_{ii}}) \\ &= p'_i U \Sigma (U'U + \Sigma^{-1}) \Sigma U' p_i \frac{\hat{v}_{rki}^2}{\hat{\sigma}_{rk}^2 p_{ii}^2} \\ &= p'_i U \Sigma U' (U \Sigma U' + I_n) p_i \frac{\hat{v}_{rki}^2}{\hat{\sigma}_{rk}^2 p_{ii}^2}. \end{aligned}$$

Note that $V = I_n + U\Sigma U'$, then it can written

$$\begin{aligned} CD_i(b) &= p'_i (V - I_n) V p_i \frac{\hat{\sigma}_v^2 s_{rki}^2}{\hat{\sigma}_{rk}^2 p_{ii}} \\ &= p'_i (V - I_n) V p_i (\frac{\hat{\sigma}_{rk}^2 + \hat{\beta}'_{rk} \Lambda \hat{\beta}_{rk}}{\hat{\sigma}_{rk}^2}) \frac{s_{rki}^2}{p_{ii}} \\ &= p'_i (V - I_n) V p_i (1 + \frac{\hat{\beta}'_{rk} \Lambda \hat{\beta}_{rk}}{\hat{\sigma}_{rk}^2}) \frac{s_{rki}^2}{p_{ii}}. \end{aligned}$$

3.5.3 Conditional Cook's distance

To investigate the influence of observations on the predicted values, we define the conditional Cook's distance, following Tan *et al.* (2001), for LMME models with ridge estimation as follows:

$$CD_{cond_i} = \hat{\sigma}_{rk}^{-2} (\hat{y}_{rk} - \hat{y}_{rk(i)})' (\hat{y}_{rk} - \hat{y}_{rk(i)})$$

which can be decomposed into three components

$$CD_{cond_i} = CD_{cond_i}^1 + CD_{cond_i}^2 + CD_{cond_i}^3$$

with

$$CD^1_{cond_i} = \hat{\sigma}^{-2}_{rk}(\hat{\beta}_{rk} - \hat{\beta}_{rk(i)})' X' X(\hat{\beta}_{rk} - \hat{\beta}_{rk(i)})$$

$$CD^2_{cond_i} = \hat{\sigma}^{-2}_{rk}(\hat{b}_{rk} - \hat{b}_{rk(i)})' U' U(\hat{b}_{rk} - \hat{b}_{rk(i)})$$

$$CD^3_{cond_i} = 2\hat{\sigma}^{-2}_{rk}(\hat{\beta}_{rk} - \hat{\beta}_{rk(i)})' X' U(\hat{b}_{rk} - \hat{b}_{rk(i)})$$

The $CD^1_{cond_i}$ is related to the fixed effects, $CD^2_{cond_i}$ identifies the observations that may influence the predictors of the random effects, and finally $CD^3_{cond_i}$ is related to the covariance between $\hat{\beta}$ and $\hat{b}$, which is expected to be close to zero.

3.5.4 Likelihood distance

The likelihood distance is a popular measure for assessing the influence of the ith observation on the corrected score estimate. We consider the corrected log-likelihood evaluated at $\hat{\beta}_{rk}$ and $\hat{\beta}_{rk(i)}$, and a measure of the influence of the ith observation on $\hat{\beta}_{rk}$ can be defined as

$$LD_i(\beta) = 2\left\{ l^*(\hat{\beta}_{rk}, k; X, y, b) - l^*(\hat{\beta}_{rk(i)}, k; X, y, b) \right\}.$$

A Taylor expansion of $l^*(y, \hat{\beta}_{rk(i)}, b, k)$ around $\hat{\beta}_{rk}$ gives

$$l^*(\hat{\beta}_{rk(i)}, k; X, y, b) = l^*(\hat{\beta}_{rk}, k; X, y, b) + \left[\tfrac{\partial l^*(\beta,k;X,y,b)}{\partial \beta} \Big|_{\beta=\hat{\beta}_{rk}, b=\tilde{b}_{rk}} \right]' \left[\hat{\beta}_{rk(i)} - \hat{\beta}_{rk} \right] + \tfrac{1}{2} \left[\hat{\beta}_{rk(i)} - \hat{\beta}_{rk} \right]' \left[\tfrac{\partial^2 l^*(\beta,k;X,y,b)}{\partial \beta \partial \beta'} \Big|_{\beta=\hat{\beta}_{rk}, b=\tilde{b}_{rk}} \right] \left[\hat{\beta}_{rk(i)} - \hat{\beta}_{rk} \right]$$

so,

$$LD_i(\beta) = 2\left[\tfrac{\partial l^*(\beta,k;X,y,b)}{\partial \beta} \Big|_{\beta=\hat{\beta}_{rk}, b=\tilde{b}_{rk}} \right]' \left[\hat{\beta}_{rk} - \hat{\beta}_{rk(i)} \right] + \left[\hat{\beta}_{rk} - \hat{\beta}_{rk(i)} \right]' \left[-\tfrac{\partial^2 l^*(\beta,k;X,y,b)}{\partial \beta \partial \beta'} \Big|_{\beta=\hat{\beta}_{rk}, b=\tilde{b}_{rk}} \right] \left[\hat{\beta}_{rk} - \hat{\beta}_{rk(i)} \right]$$

we have

$$\left[\frac{\partial l^*(\beta, k; X, y, b)}{\partial \beta} \Big|_{\beta=\hat{\beta}_{rk}, b=\tilde{b}_{rk}} \right] = 0$$

$$\left[-\frac{\partial^2 l^*(\beta, k; X, y, b)}{\partial \beta \partial \beta'} \Big|_{\beta=\hat{\beta}_{rk}, b=\tilde{b}_{rk}} \right] = \frac{A_{rk}}{\hat{\sigma}^2_{rk}}$$

and so

$$LD_i(\beta) = \hat{\sigma}^{-2}_{rk} \left[\hat{\beta}_{rk} - \hat{\beta}_{rk(i)} \right]' A_{rk} \left[\hat{\beta}_{rk} - \hat{\beta}_{rk(i)} \right].$$

As seen, we have $LD_i(\beta) = CD_i(\beta)$. Similarly, it can be shown that $LD_i(b) = CD_i(b)$.

3.6 Empirical distribution

To generate an empirical distribution of the test statistics under the null hypothesis that no influential observations exist in the data, the following algorithm is performed (see Lin *et al.* (1993); Rebai *et al.* (1994)):

Algorithm 3.2. *The algorithm is carried out in four steps:*

- ***Step 1.*** *Fit model (2) to the data with stochastic linear restrictions (3) and calculate the RRE of the parameters. An estimate of Z can be derived as*

$$\hat{Z}_{rk} = X + \hat{\sigma}_v^{-2} \hat{v}_{rk} \hat{\beta}'_{rk} \Lambda,$$

  (see Zare et al. (2012)).

- ***Step 2a.*** *Generate a new data vector as*

$$\begin{aligned} y^* &= \hat{Z}_{rk} \hat{\beta}_{rk} + U b^* + \varepsilon^*, \\ X^* &= \hat{Z}_{rk} + \Delta, \\ r^* &= R \hat{\beta}_{rk} + e^*, \end{aligned}$$

  where Δ is randomly generated as $MN(0, I_n \otimes \Lambda)$, $b^* \sim N(0, \hat{\sigma}^2_{1rk} I_q)$, $\varepsilon^* \sim N(0, \hat{\sigma}^2_{rk} I_n)$, $e^* \sim N(0, \hat{\sigma}^2_{rk} I_m)$, and R is a known matrix.

- ***Step 2b.*** *Compute the test statistics $CD_i(\beta)$, $CD_i(b)$, and CD_{cond_i} for $i = 1, 2, ..., n$ and save the order statistics of the set*

$$\{CD_i(\beta), CD_i(b), CD_{cond_i}\}.$$

- ***Step 3.*** *Repeat steps 2a and 2b, N times, for a reasonably large N. This generates an empirical distribution for each order statistic.*

- ***Step 4.*** *Calculate the $100(1-\alpha)$ percentile for each order statistic to be used as a threshold for the test statistic from the original analysis. If the ith largest values of the test statistic from the original data exceed their respective thresholds, then these observations are identified as influential.*

4. Simulation Studies

In this section, three simulation studies are presented to evaluate the performance of the estimators and the credibility of the proposed diagnostic measures, providing further evidence of the good performance of the RRE in linear mixed measurement error models.

4.1 Simulation Study One

To further investigate the behavior of the estimators, a Monte Carlo simulation study was conducted to compare the performances of the estimators mentioned above. For this purpose, the ith data set was generated from the simulated data as follows:

$$\begin{aligned} &y_i = Z\beta + Ub_i + \varepsilon_i, \quad i = 1, ..., 1000, \\ &X = Z + \Delta \quad and \quad r_i = R\beta + e_i, \end{aligned} \tag{7}$$

where

$$y_i = (y_{11i}, ..., y_{1n_1i}, y_{21i}, ..., y_{2n_2i}, ..., y_{l1i}, ..., y_{ln_li}), \quad b_i = (b_{1i}, b_{2i}, ..., b_{li})',$$

$$Z = (z^{(1)}, ..., z^{(p)}), \quad z^{(t)} = (z_{11}^{(t)}, ..., z_{1n_1}^{(t)}, z_{21}^{(t)}, ..., z_{2n_2}^{(t)}, ..., z_{l1}^{(t)}, ..., z_{ln_l}^{(t)})', \quad t = 1, ..., p,$$

and ε_i is rewritten in accordance with y_i. Here, l represents the number of independent groups, n_i is the size of group i, and the total sample size is $n = \sum_{i=1}^{l} n_i$.

The matrix $U = 1_{n_1} \oplus 1_{n_2} \oplus ... \oplus 1_{n_l}$ is an $n \times q$ matrix, where 1_{n_i} is an $n_i \times 1$ vector of ones. Furthermore,

$$r_i = (r_{1i}, ..., r_{mi})', \quad R = (R^{(1)}, ..., R^{(p)}), \quad R^{(t)} = (R_{1j}, ..., R_{mj})', \quad e_i \sim N(0, \sigma^2 I_m).$$

To achieve different degrees of collinearity, following McDonald and Galarneau (1975), the fixed effects variables are computed as

$$z_{it} = \sqrt{1-\rho^2}\, w_{it} + \rho\, w_{i,p+1}, \quad i = 1, ..., n, \ t = 1, ..., p,$$

where w_{it} are independent standard normal pseudo-random numbers and ρ^2 represents the correlation between any two fixed effects. Three different values of ρ were considered: 0.70, 0.80, and 0.90. For each set of explanatory variables, the parameter vector was chosen as the eigenvector corresponding to the largest eigenvalue of $Z'V^{-1}Z$.

The simulation study was carried out using R software (the R codes are available from the author upon request). The following combinations of parameters were considered: $n = 50$ or $n = 100$, $p = 3$, $\varepsilon_{ij} \sim N(0, \sigma^2)$, $b_{ij} \sim N(0, \sigma_1^2)$ for $i = 1, ..., q$, with $(\sigma_1^2, \sigma^2) = (0.1, 0.5)$ or $(0.3, 0.4)$, $\Lambda = \text{diag}(0, 0.05, 0.05, 0.05)$, $m = 2$, and $R_{ij}^{(t)} \sim N(0, 1)$.

The simulation was replicated 1000 times for each combination of parameters, generating new error terms for each replicate. For each replicate, the mean squared error (MSE) of the estimators was computed as

$$\text{MSE}(\tilde{\beta}) = \frac{1}{1000} \sum_{j=1}^{1000} \sum_{l=1}^{3} (\tilde{\beta}_{lj} - \beta_l)^2,$$

Table 1: Estimated MSE values of the mentioned estimators with $n = 50$.

ρ	σ^2	σ_1^2	$\hat{\beta}$	$\hat{\beta}_r$	$\hat{\beta}_k$	$\hat{\beta}_{rk}$
0.70	0.5	0.1	0.0727	0.0695	0.0696	0.0667
	0.4	0.3	0.0573	0.0549	0.0552	0.0529
0.80	0.5	0.1	0.1052	0.0988	0.0993	0.0935
	0.4	0.3	0.0817	0.0767	0.0777	0.0732
0.90	0.5	0.1	0.2206	0.1952	0.2021	0.1802
	0.4	0.3	0.1659	0.1470	0.1537	0.1370

Table 2: Estimated MSE values of the mentioned estimators with $n = 100$.

ρ	σ^2	σ_1^2	$\hat{\beta}$	$\hat{\beta}_r$	$\hat{\beta}_k$	$\hat{\beta}_{rk}$
0.70	0.5	0.1	0.0311	0.0290	0.0305	0.0285
	0.4	0.3	0.0255	0.0237	0.0251	0.0233
0.80	0.5	0.1	0.0472	0.0425	0.0460	0.0415
	0.4	0.3	0.0377	0.0338	0.0369	0.0331
0.90	0.5	0.1	0.1078	0.0889	0.1033	0.0855
	0.4	0.3	0.0818	0.0668	0.0787	0.0646

where $\tilde{\beta}_{lj}$ denotes the estimate of the lth parameter in the jth replication, and β represents the true parameter values. The results are presented in Tables 1-2.

Based on Tables 1-2, the following conclusions can be drawn:

- The estimated MSE values of all estimators increase with the level of collinearity.
- The estimated MSE values of the estimators decrease as n increases.
- In all cases, $\hat{\beta}_{rk}$ has smaller estimated MSE values than the other existing estimators. Therefore, it is concluded that the proposed estimator performs better than the other estimators based on estimated MSE values.

4.2 Simulation Study Two

To investigate the behavior of $GL(\hat{\beta}_{rk})$, $GL(\hat{b}_{rk})$, and $GL(\hat{\beta}_{rk}, \hat{b}_{rk})$ in linear mixed measurement error models, we generated the ith set of simulated data according to model (7). In this simulation study, the value of ρ was set to 0.90, $n = 50$, and $(\sigma_1^2, \sigma^2) = (0.1, 0.5)$.

To create observations with high leverage, we considered $w_{it} \sim N(5, 2)$ for $i = 1$ and $i = 8$, intended to generate two high-influence observations (observations 1 and 8). To visualize the performance of $GL(\hat{\beta}_{rk})$, $GL(\tilde{b}_{rk})$, and $GL(\hat{\beta}_{rk}, \tilde{b}_{rk})$, their values were computed for 1000 simulated datasets by generating new error terms and then averaging over the simulation runs.

Figures 1–2 present the generalized leverage plots from the simulated data. All dotted lines correspond to three times the mean leverages. The figures show that observations 1 and 8 have high generalized leverage on the fixed effects and on the combined fixed and random effects.

4.3 Simulation Study Three

In this section, a parametric bootstrap simulation is performed to evaluate the performance of Cook's distance in terms of type I error and power of the test. The jth simulated dataset was generated according to model (7), using the parameter combinations from Simulation Study One. For each dataset, diagnostic measures were calculated for the first observation (arbitrarily chosen).

To generate an empirical distribution of the test statistics under the null hypothesis, the datasets were simulated as

$$\begin{aligned} y_{jh}^* &= \hat{Z}_{rkj}\hat{\beta}_{rkj} + Ub_{jh}^* + \varepsilon_{jh}^*, \quad h = 1, ..., 1000, \\ X_{jh}^* &= \hat{Z}_{rkj} + \Delta, \quad r_{jh}^* = R\hat{\beta}_{rkj} + e_{jh}^*, \end{aligned}$$

where $\varepsilon_{jh}^*, b_{jh}^* \sim N(0, \hat{\sigma}_{rkj}^2 I_n)$ and $N(0, \hat{\sigma}_{1rkj}^2 I_q)$, respectively, and $\Delta \sim N(0, I_n \otimes \Lambda)$. Additionally, $\hat{\beta}_{rkj}$, $\hat{Z}_{rkj}$, $\hat{\sigma}_{rkj}^2$, and $\hat{\sigma}_{1rkj}^2$ are the RRE estimates of parameters from model (7).

The diagnostic measures were performed for the first observation of each simulated dataset, and the $100(1-\alpha)$ percentile from the empirical distribution was used as the threshold value. The estimated probability of type I error for different test statistics at $\alpha = 0.05$ was calculated as the proportion of datasets for which the test statistic exceeded the $100(1-\alpha)$ percentile of the empirical distribution.

To assess the power of Cook's distance, two high-influence cases for the first observation were considered:

(i) $w_{1t} \sim N(3, 1),\ t = 1, 2, 3$ with $\varepsilon_1 = 1.5$,

(ii) $w_{1t} \sim N(5, 2),\ t = 1, 2, 3$ with $\varepsilon_1 = 2$.

For each combination of parameters, 1000 datasets were generated according to

$$\begin{aligned} y_i &= Z\beta + Ub_i + \varepsilon_i, \quad i = 1, ..., 1000, \\ X_i &= Z + \Delta, \quad r_i = R\beta + e_i, \end{aligned}$$

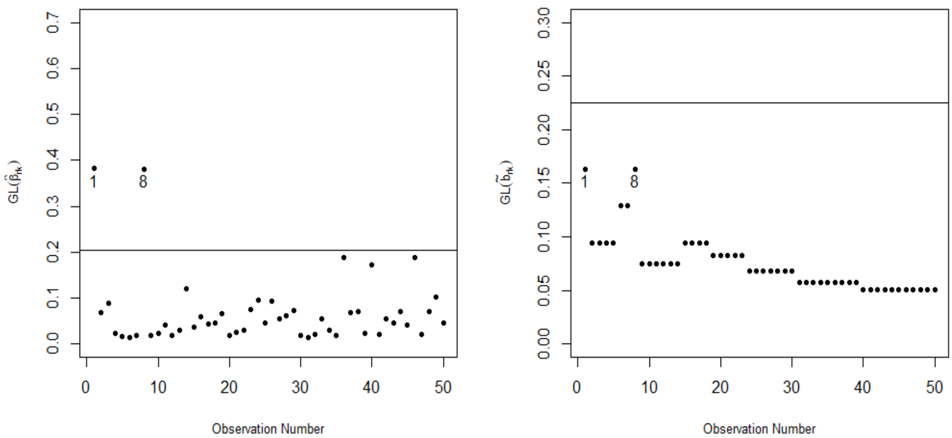

Figure 1: Generalized leverage plots on fixed effects and random effects

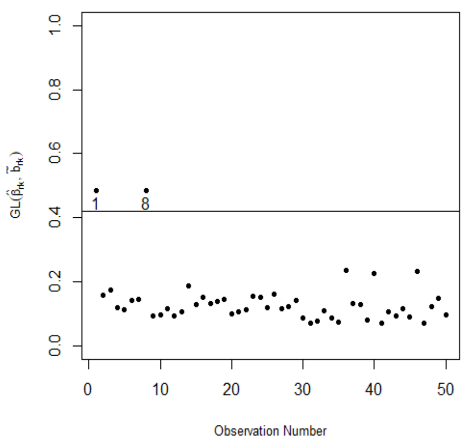

Figure 2: Generalized leverage plot on combined fixed and random effects

with the same parameters as in the type I error evaluation. For each dataset, RRE estimates and Cook's distances for the first observation were calculated. The power was computed as the proportion of datasets for which Cook's distances exceeded the $100(1-\alpha)$ percentile of the empirical distribution.

Based on Tables 3-4, the following conclusions are drawn:

(i) The type I error of the different Cook's distances is close to the nominal level of 0.05.

(ii) The power of the different Cook's distances increases as the sample size increases.

(iii) In case (ii), the power of Cook's distances generally increased compared to case (i).

Table 3: Type I error of Cook's distance with different combination of parameters

n	ρ	σ^2	σ_1^2	P-value		
				$CD_i(\beta)$	$CD_i(b)$	CD_{condi}
50	0.70	0.5	0.10	0.038	0.052	0.053
		0.40	0.30	0.046	0.055	0.053
	0.80	0.50	0.10	0.037	0.052	0.052
		0.40	0.30	0.046	0.052	0.053
	0.90	0.50	0.10	0.041	0.051	0.054
		0.40	0.30	0.045	0.049	0.047
100	0.70	0.5	0.10	0.033	0.045	0.043
		0.40	0.30	0.038	0.043	0.045
	0.80	0.50	0.10	0.028	0.045	0.045
		0.40	0.30	0.035	0.049	0.053
	0.90	0.50	0.10	0.033	0.046	0.044
		0.40	0.30	0.032	0.044	0.040

5. Real Data Analysis:

6. Real Data Example

In this section, the historical market dataset for real estate valuation from Sindian District, New Taipei City, Taiwan, is used to evaluate the performance of the proposed diagnostic criteria. The dataset is publicly available online at the UCI Machine Learning Repository (https://archive.ics.uci.edu/ml/datasets.html)

Table 4: Power of Cook's distance with different combination of parameters

				Power					
n	ρ	σ^2	σ_1^2	$CD_i(\beta)$		$CD_i(b)$		CD_{condi}	
				case(i)	case(ii)	case(i)	case(ii)	case(i)	case(ii)
50	0.70	0.5	0.10	0.837	0.888	0.124	0.144	0.743	0.854
		0.40	0.30	0.779	0.874	0.167	0.173	0.610	0.778
	0.80	0.50	0.10	0.824	0.8892	0.130	0.146	0.727	0.838
		0.40	0.30	0.751	0.857	0.173	0.187	0.542	0.746
	0.90	0.50	0.10	0.775	0.877	0.122	0.154	0.677	0.810
		0.40	0.30	0.696	0.831	0.158	0.186	0.520	0.723
100	0.70	0.5	0.10	0.938	0.964	0.251	0.311	0.850	0.926
		0.40	0.30	0.947	0.970	0.342	0.413	0.768	0.903
	0.80	0.50	0.10	0.934	0.960	0.259	0.318	0.846	0.926
		0.40	0.30	0.973	0.982	0.402	0.479	0.833	0.933
	0.90	0.50	0.10	0.925	0.952	0.262	0.330	0.837	0.916
		0.40	0.30	0.828	0.958	0.344	0.427	0.770	0.904

and is labeled "Real Estate Valuation Data Set." The input variables are as follows:

$$\begin{aligned} Y &= \text{house price per unit area,} \\ X_1 &= \text{house age,} \\ X_2 &= \text{distance to the nearest MRT station,} \\ X_3 &= \text{number of convenience stores in the living circle on foot,} \\ X_4 &= \text{latitude,} \\ X_5 &= \text{longitude,} \\ &\text{transaction date (2012 or 2013).} \end{aligned}$$

Measurement errors are considered in some variables due to rounding of the observed values. The dataset was fitted using a linear mixed measurement error (LMME) model of the form

$$y = X\beta + Ub + \varepsilon,$$

where y is a 414×1 vector of response variables, X is a 414×5 regression matrix, and U is a 414×2 design matrix. The transaction date (2012 or 2013) is modeled as a random effect. The initial values for the variance components were set as $\sigma^2 = 0.5$ and $\sigma_1^2 = 0.1$.

The condition number of $\hat{Z}'V^{-1}\hat{Z}$ is 137,976.4, indicating severe multicollinearity among the fixed effects variables. The 114th element of y and the 114th row of X were taken as r and R, respectively, for stochastic restriction purposes. The parameter estimates for the LMME model are presented in Table 5.

Table 5: Parameter estimates for the Real estate valuation data set.

Parameter	corrected score	Restricted score	Restricted ridge score
β_1	-0.2558	-0.2537	-0.2503
β_2	-0.0055	-0.0055	-0.0056
β_3	1.3125	1.2973	1.3253
β_4	-25.4877	-26.5538	-72.5817
β_5	5.5868	5.8014	15.2575
b_1	-1.2264	-0.8049	-0.8027
b_2	1.2263	1.7112	1.7216
σ^2	83.81	86.88	88.02
σ_1^2	1.8369	2.1843	2.18

Generalized leverage plots of the observations are shown in Figures 3-4. All the dotted lines correspond to three times the mean leverages. Figure 3 shows that observations 9, 117, 250, 256, and 348 have high generalized leverage on the fixed effects based on the generalized marginal leverage $GL(\hat{\beta}_{rk})$. Figure 4 shows that these observations also have high generalized leverage on the fixed and random effects.

Additionally, the Cook's distance for fixed and random effects, as described in previous sections, was calculated for each observation. Figures 5-6 show plots of Cook's distance for fixed effects, random effects, and the Conditional Cook's distance, respectively. A glance at Figures 5-6 indicates that observation 271 has a substantial influence on the fixed effects, random effects, and predicted values.

7. Discussion

In this paper, to overcome the problem of multicollinearity, the restricted ridge estimator (RRE) for the parameter vector using Nakamura's corrected score function is presented for LMME models. The performance of the estimator was evaluated using the asymptotic MSEM criterion, and it was shown that, under this criterion, the proposed estimator is more efficient than existing estimators.

Furthermore, diagnostic methods based on the RRE were presented to identify high-leverage points and influential observations in the proposed model. Using parametric bootstrap simulations, different Cook's distances were studied in terms of type I error and test power. It was found that the type I error of the test statistics for different parameter combinations is close to the nominal value α, and

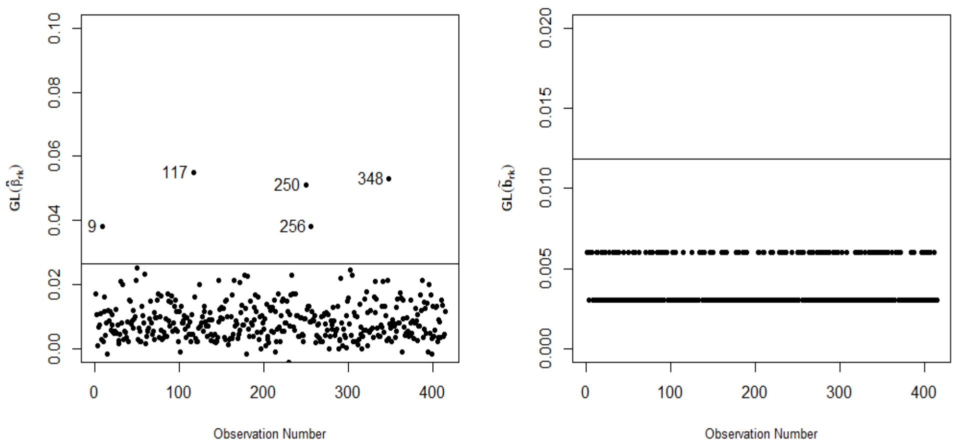

Figure 3: Plot of generalized leverage for fixed effects and random effects

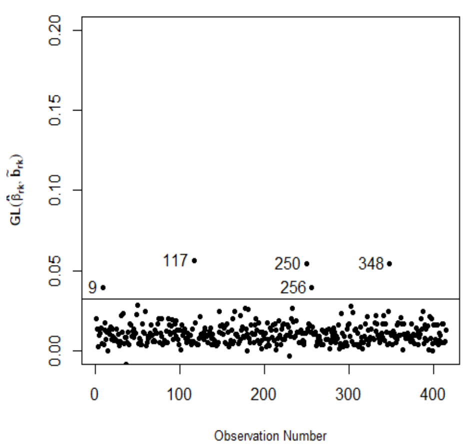

Figure 4: Plot of generalized leverage for fixed and random effects

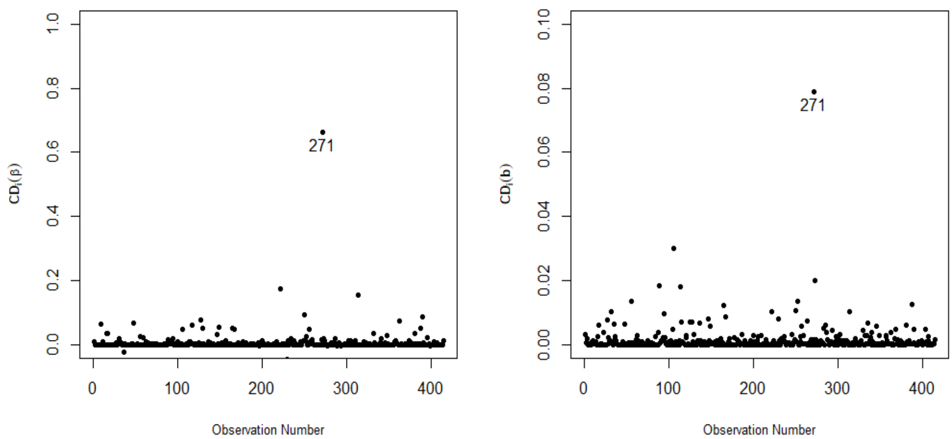

Figure 5: Plot of Cook's distance for fixed effects and random effects

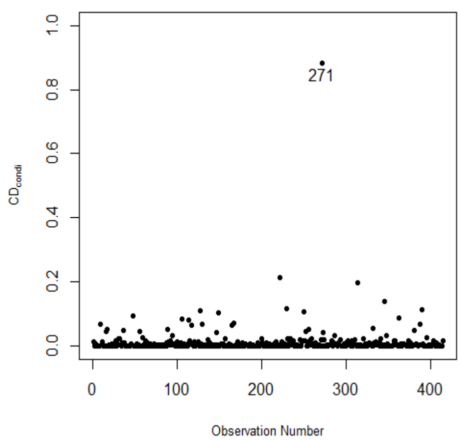

Figure 6: Plot of Conditional Cook's distance

that the power of the test statistics increases with the sample size. Both simulation studies and real data analysis demonstrate that the proposed diagnostic measures perform very well in correctly identifying influential observations in LMME models with RRE.

Acknowledgement

The author would like to express gratitude to the editor and the anonymous referees for their valuable comments and suggestions, which have significantly improved the presentation of this paper.

References

Armstrong, A. (1985). Measurement error in the generalised linear model. *Communications in Statistics - Simulation and Computation,* **14**(3), 529-544.

Banerjee, M., and Edward W. Frees. (1997). Influence Diagnostics for Linear Longitudinal Models. *Journal of the American Statistical Association,* **92**, 999–1005.

Borhani, Sh., F. Ghapani, and R. Jafaraghaie. (2023). Detection of outliers and influential observations in linear mixed measurement error models with Liu estimation. *Journal of Statistical Modelling: Theory and Applications,* **4**(2), 1-16.

Borhani, Sh., F. Ghapani, and R. Jafaraghaie. (2024). Detecting Influential Observations Using Liu Estimator in Linear Mixed Measurement Error Models. *Journal of Statistical Theory and Practice,* **5**(2), 137-152. doi:10.1007/s42519-023-00363-2.

Carroll, R. J., D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective,* **105**, Chapman and Hall/CRC.

Cook, R. D. (1977). Detection of influential observations in linear regression. *Technometrics,* **19**, 15–18.

Christensen, R., L. M. Pearson, and W. Johnson. (1992). Case-Deletion Diagnostics for Mixed-Models. *Technometrics,* **34**, 38–45.

Feller, W. (1972). *An Introduction to Probability Theory and Its Applications*, New York, John Wiley.

Cui, H., K. W. Ng, and L. Zhu. (2004). Estimation in mixed effects model with errors in variables. *Journal of Multivariate Analysis,* **91**(1), 53–73.

Davidian, M., and D. M. Giltinan. (1995). *Nonlinear Models for Repeated Measurement Data,* London: Chapman and Hall.

Fuller, W. A. (1987). *Measurement Error Models,* Wiley, New York.

Fung, W. K., X. P. Zhong, and B. C. Wei. (2003). On estimation and influence diagnostics in linear mixed measurement error models. *American Journal of Mathematical and Management Sciences,* **23**, 37–59.

Ganjealivand, N., F. Ghapani, A. Zaherzadeh, and F. Hormozinejad. (2021). Performance of stochastic restricted and unrestricted two-parameter estimators in linear mixed models. *Revista Internacional de Métodos Numéricos para Cálculo y Diseño en Ingeniería,* **37**(2). doi:10.23967/j.rimni.2021.06.001.

Ganjealivand, N., and F. Ghapani. (2024). The weighted two-parameter estimation for linear mixed models with measurement error under stochastic linear mixed restrictions. *Journal of Statistical Modelling: Theory and Applications,* **5**(2), 137–152.

Ghapani, F. (2022). Stochastic restricted Liu estimator in linear mixed measurement error models. *Communications in Statistics - Simulation and Computation,* **51**, 1220–1233.

Gilmour, A. R., B. R. Cullis, S. J. Welhman, B. J. Gogel, and R. Thompson. (2004). An efficient computing strategy for predicting in mixed linear models. *Computational Statistics and Data Analysis,* **44**(4), 571–586.

Gumedze, F. N., S. J. Welham, B. J. Gogel, and R. Thompson. (2010). A variance shift model for detection of outliers in the linear mixed model. *Computational Statistics and Data Analysis,* **54**(9), 2128–2144.

Harrison, D., and D. L. Rubinfeld. (1978). Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management,* **5**(1), 81–102.

Haslett, J., and D. Dillane. (2004). Application of 'Delete = Replace' to Deletion Diagnostics for Variance Component Estimation in the Linear Mixed Model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* **66**(1), 131–143.

Henderson, C. R., O. Kempthorne, S. R. Searle, and C. M. von Krosigk. (1959). The estimation of environmental and genetic trends from records subject to culling. *Biometrics,* **15**(2), 192–218.

Hoaglin, D. C., and R. E. Welsch. (1978). The hat matrix in regression and ANOVA. *The American Statistician,* **32**(1), 17–22.

Hoerl, A. E., and R. W. Kennard. (1970). Ridge regression: biased estimation for non-orthogonal problems. *Technometrics,* **12**(1), 55–67.

Kuran, O., and M. R. Ozkale. (2016). Gilmour's approach to mixed and stochastic restricted ridge predictions in linear mixed models. *Linear Algebra and its Applications,* **508**, 22–47.

Laird, N. M., and J. H. Ware. (1982). Random-effects models for longitudinal data. *Biometrics,* **38**(4), 963–974.

Li, Z., W. Xu, and L. Zhu. (2009). Influence diagnostics and outlier tests for varying coefficient mixed models. *Journal of Multivariate Analysis,* **100**, 2002–2017.

Lin, D. Y., L. J. Wei, and Z. Ying. (1993). Checking the Cox model using cumulative sums of martingale-based residuals. *Biometrika,* **80**(3), 557–572.

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

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

Maksaei, N., A. Rasekh, and B. Babadi. (2023). Influence measures and outliers detection in linear mixed measurement error models with Ridge estimation. *Communications in Statistics - Simulation and Computation,* **52**(7), 3062–3078.

McDonald, C., and D. A. Galarneau. (1975). A Monte Carlo Evaluation of some Ridge-Type Estimators. *Journal of the American Statistical Association,* **70**, 407–416.

Nakamura, T. (1990). Corrected score function for errors-in-variables models: Methodology and application to generalized linear models. *Biometrika,* **77**, 127–137.

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

Rao, C. R., H. Toutenburg, C. Shalabh, and Heumann. (2008). *Linear Models and Generalizations,* Springer, Berlin.

Rebai, A., Goffinet, B., and B. Mangin. (1994). Approximate thresholds of interval mapping tests for QTL detection. *Genetics,* **138**(1), 235–240.

Stein, C. (1956). Inadmissibility of the usual estimator for mean of multivariate normal distribution. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability,* **1**, 197–206.

Stefanski, L. A., and R. J. Carroll. (1991). Deconvolution-based score tests in measurement error models. *Annals of Statistics,* **19**(1), 249–259.

Tan, F. E. S., M. J. N. Ouwens, and M. P. F. Berger. (2001). Detection of influential observations in longitudinal mixed effects regression models. *Journal of the Royal Statistical Society: Series D (The Statistician),* **50**(3), 271–284.

Wei, B. C., Y. Q. Hu, and W. K. Fung. (1998). Generalized leverage and its applications. *Scandinavian Journal of Statistics,* **25**, 25–37.

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

Yavarizadeh, B., A. R. Rasekh, S. E. Ahmed, and B. Babadi. (2022). Ridge estimation in linear mixed measurement error models with stochastic linear mixed restrictions. *Communications in Statistics - Simulation and Computation,* **51**(6), 3037–3053.

Zare, K., and A. Rasekh. (2012). Diagnostic measures for linear mixed measurement error models. *SORT,* **35**, 125–144.

Zare, K., and A. Rasekh. (2014). Residuals and leverages in the linear mixed measurement error models. *Journal of Statistical Computation and Simulation,* **84**, 1427–1443.

Zare, K., A. Rasekh, and A. Rasekhi. (2012). Estimation of variance components in linear mixed measurement error models. *Statistica Papers,* **53**, 849–863.

Zewotir, T., and J. S. Galpin. (2005). Influence diagnostics for linear mixed models. *Journal of Data Science,* **3**, 153–177.

Zhong, X. P., and B. C. Wei. (1999). Influence analysis on linear models with random effects. *Applied Mathematics Journal, Chinese University, Series B,* **14**, 169–176.

Zhong, X. P., W. K. Fung, and B. C. Wei. (2002). Estimation in linear models with random effects and errors-in-variables. *Annals of the Institute of Statistical Mathematics,* **54**, 595–606.

8. Appendix

Lemma 8.1. *Assume that square matrices A and C are not singular and B and D are matrices with proper orders; then $(A+BCD)^{-1} = (A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)DA^{-1}$. (Rao et al. (2008), Theorem A. 18)*

Proof. Proof of Theorem 3.1

$$\hat{\beta}_{rk(i)} = \left[X'_{(i)}V_{[i]}^{-1}X_{(i)} + R'W^{-1}R - tr(V_{[i]}^{-1})\Lambda + kI_p\right]^{-1}(X'_{(i)}V_{[i]}^{-1}y_{(i)} + R'W^{-1}r)$$
$$= \left[A_{rk} - X'c_ic'_iX/c_{ii} + (c_{ii} + \tfrac{1}{c_{ii}}c'_ic_i)\Lambda\right]^{-1} \times (X'V^{-1}y - X'c_ic'_iy/c_{ii} + R'W^{-1}r)$$
$$= [A_{rk} - X'c_ic'_iX/c_{ii}]^{-1} \times (X'V^{-1}y - X'c_ic'_iy/c_{ii} + R'W^{-1}r) + O_p(n^{-1})$$

Using Lemma 8.1, it can be written

$$(A_{rk} - X'c_ic'_iX/c_{ii})^{-1} = A_{rk}^{-1} + A_{rk}^{-1}X'c_i(1 - \frac{c'_iXA_{rk}^{-1^{-1}}X'c_i}{c_{ii}})^{-1}A_{rk}^{-1}$$

With substituting the above expression in $\hat{\beta}_{k(i)}$, it is concluded that

$$\hat{\beta}_{rk(i)} = \hat{\beta}_{rk} - A_{rk}^{-1} X' c_i \frac{\hat{v}_{rki}}{p_{ii}} + O_p(n^{-1})$$

$$\begin{aligned}
&\hat{b}_{rk(i)} = \Sigma U'_{(i)} V_{[i]}^{-1} \left[y_{(i)} - X_{(i)} \hat{\beta}_{k(i)} \right] = \Sigma U'_{(i)} V_{[i]}^{-1} y_{(i)} - \Sigma U'_{(i)} V_{[i]}^{-1} X_{(i)} \hat{\beta}_{rk(i)} \\
&= \Sigma (U' V^{-1} y - \tfrac{U' c_i c'_i y}{c_{ii}}) \\
&- \Sigma (U' V^{-1} X - \tfrac{U' c_i c'_i X}{c_{ii}}) \left[\hat{\beta}_{rk} - A_{rk}^{-1} X' c_i \tfrac{\hat{v}_{rki}}{p_{ii}} + O_p(n^{-1}) \right] \\
&\simeq \Sigma U' V^{-1} y - \Sigma \tfrac{U' c_i c'_i y}{c_{ii}} - \Sigma U' V^{-1} X \hat{\beta}_{rk} + \Sigma U' V^{-1} X A_{rk}^{-1} X' c_i \tfrac{\hat{v}_{rki}}{p_{ii}} \\
&+ \Sigma \tfrac{U' c_i c'_i X}{c_{ii}} \hat{\beta}_{rk} - \Sigma U' c_i \tfrac{c'_i X A_{rk}^{-1} X' c_i}{c_{ii}} \tfrac{\hat{v}_{rki}}{p_{ii}} \\
&= \hat{b}_{rk} - \Sigma U' p_i \tfrac{\hat{v}_{rki}}{p_{ii}},
\end{aligned}$$

and

$$\begin{aligned}
&(n+m-1)\hat{\sigma}^2_{rk(i)} = (y_{(i)} - X_{(i)} \hat{\beta}_{rk(i)})' V_{[i]}^{-1} (y_{(i)} - X_{(i)} \hat{\beta}_{rk(i)}) \\
&- tr(V_{[i]}^{-1}) \hat{\beta}'_{rk(i)} \Lambda \hat{\beta}_{rk(i)} + k \hat{\beta}'_{rk(i)} \hat{\beta}_{rk(i)} \\
&= y'_{(i)} V_{[i]}^{-1} y_{(i)} - \hat{\beta}'_{rk(i)} X'_{(i)} V_{[i]}^{-1} y_{(i)} \\
&= y'_{(i)} V_{[i]}^{-1} y_{(i)} - \left\{ \hat{\beta}_{rk} - A_{rk}^{-1} X' c_i \tfrac{\hat{v}_{rki}}{p_{ii}} + O_p(n^{-1}) \right\}' X'_{(i)} V_{[i]}^{-1} y_{(i)} \\
&= y' V^{-1} y - y' c_i c'_i y / c_{ii} - \left[\hat{\beta}_{rk} - A_{rk}^{-1} X' c_i \tfrac{\hat{v}_i}{p_{ii}} + O_p(n^{-1}) \right] \\
&\times (X' V^{-1} y - X' c_i c'_i y / c_{ii}) \\
&= y' V^{-1} y - y' c_i c'_i y / c_{ii} - \hat{\beta}'_{rk} X' V^{-1} y + \hat{\beta}'_{rk} X' c_i c'_i y / c_{ii} \\
&+ c_i' X A_{rk}^{-1} X' V^{-1} y \tfrac{\hat{v}_{rki}}{p_{ii}} - c_i' X A_{rk}^{-1} X' c_i c'_i y \tfrac{\hat{v}_{rki}}{c_{ii} p_{ii}} + O_p(1) \\
&= n \hat{\sigma}^2_{rk} - \tfrac{\hat{v}^2_{rki}}{p_{ii}} + O_p(1) \Rightarrow \hat{\sigma}^2_{rk(i)} = \tfrac{n \hat{\sigma}^2_{rk} - \frac{\hat{v}^2_{rki}}{p_{ii}}}{n+m-1} + O_p(n^{-1}).
\end{aligned}$$

□