https://jaslda.journals.pnu.ac.ir

Payame Noor University

Statistical Learning & Data Analysis
Volume 1, Issue 1, Summer- Autumn 2026, Pages 89–106
doi:10.30473/jaslda.2026.79273.1009

OPEN ACCESS

JASLDA
Journal of Advances in Statistical Learning and Data Analysis

Research Article

Sensitivity Analysis of Diagnostic Measures to the Liu Biasing Parameter and Error Distribution in Linear Mixed Measurement Error Models

Fatemeh Ghapani[1], Razieh Jafaraghaie[1,*]

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

Article Info

Article history:
Received August 29, 2026;
Accepted October 16, 2026;
Available online October 16, 2026.

Abstract. In this paper, we investigate the sensitivity of diagnostic measures to the Liu biasing parameter and error distribution assumptions in linear mixed measurement error (LMME) models when multicollinearity is present. We evaluate the performance of Cook's distance for fixed effects, random effects, and conditional responses under different error distributions (Normal, heavy-tailed, and skewed). A parametric bootstrap procedure is used to obtain empirical distribution of the test statistic and a simulation study has been given to show the performance diagnostic criteria. The simulation study shows that the proposed methods maintain the Type I error rate close to the nominal level and exhibit high power, even under non-normal error distributions. The real data analysis, utilizing the Boston Housing dataset, demonstrates that the identification of influential points depends heavily on the Liu biasing parameter, emphasizing the need for sensitivity analysis in practical applications.

How to Cite
Ghapani F., Jafaraghaie R. (2026). "Sensitivity Analysis of Diagnostic Measures to the Liu Biasing Parameter and Error Distribution in Linear Mixed Measurement Error Models". *Journal of Advances in Statistical Learning and Data Analysis*, 1(1), 89–106. https://doi.org/10.30473/jaslda.2026.79273.1009

Keywords. Case deletion model; Cook's distance; Liu estimator; Measurement error; Multicollinearity; Parametric bootstrap.

MSC. 62J05, 62J07, 62F40 .

* Corresponding author. **E-mail:** r.jafaraghaee@iau-shoushtar.ac.ir
©2026 by the authors. Licensee PNU, Tehran, Iran. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 International (CC BY4.0) (http:/creativecommons.org/licenses/by/4.0)

1 Introduction

Linear mixed models have tremendous flexibility to fit statistical models and are used to analyze data over a wide range of topics. These models are commonly used to analyze longitudinal and repeated measures data in economic, social, and medical studies. For more details, studying the works of [9, 11, 20] is recommended. However, in many practical applications, the explanatory variables are subject to measurement errors. The maximum likelihood method for estimating linear model parameters with normal measurement error leads to incompatibility and biased estimates [5, 6, 12].

It is a very common problem to deal with highly intercorrelated explanatory variables (multicollinearity). There are some results indicating multicollinearity problems, such as having inflated variance scores and unstable estimations of the parameters. There are several ways to overcome this problem, such as the Stein estimator [24], ridge regression [16], and the Liu estimator [18]. [18] proposed a new estimator to improve the Ridge estimator by combining it with the Stein estimator. The characteristics of this estimator have been investigated by [17] and [1]. In order to overcome the problems of multicollinearity, [19] and [23] introduced the ridge estimator and predictor in linear mixed models. More recently, [25] introduced the Ridge estimator in the presence of multicollinearity in linear mixed models with measurement error. [14] concentrated on the stochastic restricted Liu estimation of parameters in linear mixed measurement error models and indicated that there exists a Liu biasing parameter d, so that the Liu estimator dominates the corrected likelihood estimator (CLE).

Since the simultaneous presence of influential observations and multicollinearity in LMME models is not unusual, it is important to consider influential points in data analysis. Various methods have been proposed for this purpose. These include residual analysis and case deletion models (CDM) [2, 7, 29]. In LMME models, the available work in the field of diagnostic methods is mainly related to [13, 26, 28].

However, to the best of our knowledge, no attention has been paid in the literature concerning how the choice of the Liu biasing parameter d and the violation of the normality assumption affect the outcome of outlier and influential detection. In this paper, by integrating and extending existing diagnostic methods, we investigate how changes in the parameter d and the violation of the normality assumption of errors affect the control of Type I error and the power of the test in identifying outlier and influential observations.

As demonstrated in previous studies [4, 14], the identification of influential observations (such as observations 9 and 36 in the Boston Housing dataset) is highly sensitive to the choice of the Liu biasing parameter d. However, the statistical performance of these diagnostic measures, specifically in terms of Type I error rate and power of the test under various error distributions and different values of d, has not been systematically investigated in the literature. This paper aims to fill this gap by conducting a comprehensive sensitivity analysis and simulation study.

The rest of this paper is organized as follows. In Section 2, the Liu estimator in linear mixed measurement error models is introduced. In Section 3, we study diagnostic measures, including residuals, case deletion models, and Cook's distance, and introduce a parametric bootstrap procedure. In Section 4, a simulation study is conducted to examine the performance of the influence measures under different error distributions. In Section 5, the proposed diagnostic methods are demonstrated with the analysis of the Boston housing data. Finally, concluding remarks are given in Section 6.

2 Model and Estimation

Consider the following linear mixed model with measurement errors:

$$y = X\beta + Zb + \epsilon, \quad x = X + \Delta \tag{1}$$

In this model, y is an $n \times 1$ vector of observations, and β is a $p \times 1$ vector of unknown regression coefficients (fixed effects). X and Z are $n \times p$ and $n \times q$ matrices of regressors, respectively, where Z has a block structure $Z = [Z_1|Z_2|\dots|Z_l]$, where Z_i is an $n_i \times q_i$ known design matrix of the ith random effect factor, such that $n = \sum_{i=1}^{l} n_i$ and $q = \sum_{i=1}^{l} q_i$. b is a $q \times 1$ vector of unobservable random effects from $N(0, \sigma^2\Sigma_b)$, independent of $\epsilon \sim N(0, \sigma^2 I_n)$, which Σ_b is a block diagonal matrix with the ith block being $\gamma_i I_{q_i}$ for $\gamma_i = \frac{\sigma_i^2}{\sigma^2}$ and ϵ is an $n \times 1$ vector of unobservable random errors. The matrix x is the observed value of X with the measurement error Δ, where Δ is an $n \times p$ random matrix with $E(\Delta) = 0$ and $Cov(vec(\Delta)) = I_n \otimes \Sigma_\Delta$, and Σ_Δ is a known nonnegative definite matrix [12].

Special Cases of Covariance Matrices: It is worth noting specific forms of the measurement error covariance matrix Σ_Δ. If $\Sigma_\Delta \to 0$ (no measurement error), the model reduces to a standard linear mixed model. If Σ_Δ is a diagonal matrix, it implies that measurement errors in different predictors are uncorrelated, which is a common and practical assumption in many applied settings.

Under model (1), y has a multivariate normal distribution with $\mathrm{E}(y) = X\beta$ and $\mathrm{Var}(y) = \sigma^2 V$, where $V = Z\Sigma_b Z' + I_n$. The conditional distribution of $b\,|y$ is $b\,|y \sim N(\Sigma_b Z' V^{-1}(y - X\beta), \sigma^2\Sigma_b T)$, where $T = (I_q - Z'V^{-1}Z\Sigma_b) = (I_q + Z'Z\Sigma_b)^{-1}$. Using the corrected score method proposed by [22], the Corrected Likelihood Estimates (CLE) of β, b and σ^2 are given by $\hat{\beta} = A^{-1}C$, in which $C = x'V^{-1}y$, $\hat{b} = \Sigma_b Z' V^{-1}(y - x\hat{\beta})$, $\hat{\sigma}^2 = \frac{1}{n}(y'V^{-1}y - \hat{\beta}'x'V^{-1}y)$, where $A = x'V^{-1}x - tr(V^{-1})\Sigma_\Delta$ and $\hat{\sigma}_i^2 = \frac{[\hat{b}_i'\hat{b}_i - tr(\hat{D}_i'\hat{D}_i)\hat{\beta}'\Sigma_\Delta\hat{\beta}]}{q_i - tr(T_{ii})}$ in which $\hat{D}_i = \hat{\gamma}_i Z_i' V^{-1}$ for T_{ij} is ijth block of matrix T (see [27, 30]). If the γ_i's are unknown, the corrected score estimates are substituted back into Σ_b (and/or V) to obtain $\hat{\beta}$, $\hat{\sigma}^2$ and $\hat{b}$.

To overcome the destructive effects of multicollinearity, we use the Liu estimator. Based on [18], we consider the linear restriction $d\hat{\beta} = \beta + e, e \sim N(0, \sigma^2 I_p)$, where $d \in (0, 1)$ is the Liu biasing parameter. The Liu Corrected Likelihood Estimator (LCLE) of β, b and σ^2 are obtained as [14]:

$$\hat{\beta}_d = (A + I_p)^{-1}(A + dI_p)\hat{\beta}, \quad \hat{b}_d = \Sigma_b Z' V^{-1}(y - x\hat{\beta}_d) \tag{2}$$

$$\hat{\sigma}_d^2 = n^{-1}\left[(y - x\hat{\beta}_d)'V^{-1}(y - x\hat{\beta}_d) + (d\hat{\beta} - \hat{\beta}_d)'(d\hat{\beta} - \hat{\beta}_d) - tr(V^{-1})\hat{\beta}_d'\Sigma_\Delta\hat{\beta}_d\right] \tag{3}$$

As demonstrated in previous studies [3, 14], there exists a biasing parameter $d \in (0, 1)$ such that the LCLE dominates the CLE in the sense of the Mean Squared Error Matrix (MSEM).

3 Diagnostic Methods

To determine the observations that behave abnormally, we study diagnostic measures in the context of LMME models under the LCLE. We rearrange the matrices so that the ith deleted case to be in the first row. Then

$$y = \begin{bmatrix} y_i \\ y_{(i)} \end{bmatrix}, x = \begin{bmatrix} x'_i \\ x_{(i)} \end{bmatrix}, X = \begin{bmatrix} X'_i \\ X_{(i)} \end{bmatrix}, Z = \begin{bmatrix} Z'_i \\ Z_{(i)} \end{bmatrix},$$

and

$$C = V^{-1} = \begin{bmatrix} c_{ii} & c'_{i(i)} \\ c_{i(i)} & V_{[i]}^{-1} + c_{i(i)}c'_{i(i)} \big/ c_{ii} \end{bmatrix}$$

For notational simplicity x'_i denotes the ith row of x , $x_{(i)}$ denotes matrix x with the ith row removed, c_{ii} the ith elements of the diagonal matrix V^{-1} , $c_{(i)}$ denotes the ith row of V^{-1} , $c_{(i)}$ denotes vector c with the ith element removed and c_i the ith element of c , $V_{[i]}$ denotes matrix V without the ith row and the ith column.

3.1 Residuals

Residual analysis is a fundamental tool for evaluating the appropriateness of any statistical model and detecting potential outliers. In the context of mixed models, we distinguish between marginal and conditional residuals. Using the Liu estimators, the conditional residuals of the model are defined as $\hat{v}_d = y - x\hat{\beta}_d - Z\hat{b}_d$. These residuals account for both the fixed effects and the predicted random effects, making them highly sensitive to observations that deviate from the expected conditional mean. Large absolute values of standardized conditional residuals often indicate the presence of outlying responses that warrant further investigation. Also, the ith studentized residual of the model is defined as $t_i = \frac{\hat{v}_{di}}{\hat{\sigma}_v\sqrt{r_{pii}}}$, where r_{pii} denotes the ith diagonal elements of $R_p = V^{-1} - V^{-1}x(A + I_p)^{-1}x'V^{-1}$. The studentized residuals are going to be more effective for detecting outlying observations than standardized residuals. If an observation has a large t_i^2, then it is an outlier observation.

3.2 Case Deletion Diagnostic

To quantify the effects of deleting the ith observation, the case deletion model based on the Liu restriction with the ith observation deleted is defined as:

$$y_{(i)} = X_{(i)}\beta + Z_{(i)}b + \varepsilon_{(i)}, \quad x_{(i)} = X_{(i)} + \Delta_{(i)}, i = 1, 2, ...n, \quad \text{and} \quad d\hat{\beta} = \beta + e. \tag{4}$$

The corrected log-likelihoods of y and the conditional corrected log-likelihood of $b\,|y$ for case deletion model are defined as:

$$\begin{aligned}
{l_{ci}}^*(\theta, x, y) = & -\frac{n-1}{2}\log(2\pi\sigma^2) - \frac{1}{2}\log\left|V_{[i]}\right| \\
& - \frac{1}{2\sigma^2}\left[(y_{(i)} - x_{(i)}\beta)' V^{-1}(y_{(i)} - x_{(i)}\beta) - tr({V_{[i]}}^{-1})\beta'\Sigma_\Delta\beta\right] \\
& - \frac{1}{2\sigma^2}\left[(d\hat{\beta} - \beta)'(d\hat{\beta} - \beta)\right], \\
l^*_{bci}(\theta; x, y) = & -\frac{q}{2}\log(2\pi\sigma^2) - \frac{1}{2}\log\left(\left|{Z'}_{(i)}Z_{(i)} + \Sigma^{-1}\right|^{-1}\right) \\
& - \frac{1}{2\sigma^2}\left[b'[{Z'}_{(i)}Z_{(i)} + \Sigma^{-1}]\, b\right. \\
& -2b'[{Z'}_{(i)}Z_{(i)} + \Sigma^{-1}]^{-1}\Sigma\,[{Z'}_{(i)}V^{-1}_{[i]}(y_{(i)} - x_{(i)}\beta) \\
& \left.+(y_{(i)} - x_{(i)}\beta)'(I - {V_{[i]}}^{-1})(y_{(i)} - x_{(i)}\beta) - tr(I - {V_{[i]}}^{-1})\beta'\Sigma_\Delta\beta\right]
\end{aligned}$$

where, $\theta = (\beta, \sigma^2, d, \gamma)$. With differentiating $l^*_{ci}(\theta; x, y)$ with respect to β and σ^2 and $l^*_{bci}(\theta; x, y)$ with respect to b, respectively, the LCLE without the ith case can be approximated as [3]:

$$\hat{\beta}_{d(i)} \simeq \hat{\beta}_d - A_p^{-1}\frac{x' c_i \hat{v}_{di}}{r_{pii}}, \quad \hat{b}_{d(i)} \simeq \hat{b}_d - \Sigma_b Z' r_{pi}\frac{\hat{v}_{di}}{r_{pii}} \tag{5}$$

$$\hat{\sigma}^2_{d(i)} \simeq \frac{n}{n-1}\hat{\sigma}^2_d - \frac{\hat{v}^2_{di}}{(n-1)r_{pii}} \tag{6}$$

where ${c_i}'$ and r'_{pi} are ith rows of V^{-1} and $R_p = V^{-1} - V^{-1}X(A + I_p)^{-1}X'V^{-1}$, respectively.

3.3 Generalized Cook's Distance

Cook's distance is a convenient measure of influence [8]. For the fixed effects under LCLE, it is defined as:

$$CD_i(\beta) = \frac{(\hat{\beta}_d - \hat{\beta}_{d(i)})'\left[x'V^{-1}x - tr(V^{-1})\Sigma_\Delta + \mathrm{I}_p\right](\hat{\beta}_d - \hat{\beta}_{d(i)})}{\hat{\sigma}^2_d} \tag{7}$$

we have approximately $CD_i(\beta) = \frac{(c_{ii} - r_{pii})\hat{v}^2_{di}}{\hat{\sigma}^2_d r^2_{pii}} + O_p(n^{-1})$. Large values of $CD_i(\beta)$ indicate that observation i has a substantial effect on the full sample estimate. Similarly, the Cook's distance for random effects is defined as:

$$CD_i(b) = \frac{(\hat{b}_d - \hat{b}_{d(i)})'(Z'Z + \Sigma_b^{-1})(\hat{b}_d - \hat{b}_{d(i)})}{\hat{\sigma}^2_d} \tag{8}$$

To examine the impact on predicted values, the conditional Cook's distance, incorporating a combination of fixed and random effects, is defined as follows:

$$CD_{cond_i} = \hat{\sigma}_d^{-2}(\hat{y}_d - \hat{y}_{d(i)})'(\hat{y}_d - \hat{y}_{d(i)}) \tag{9}$$

Large values of CD_{cond_i} indicate that the subset of the observations indexed by i are jointly influential on fixed effects and random effects. Since test statistics in CDM are function of a multiple of the square of studentized residuals of model and their joint distribution are independent, it is difficult to calculate them. Therefore, a parametric bootstrap procedure can be used to obtain the Empirical distributions of the Cook's distance statistics.

3.4 Parametric Bootstrap Algorithm

To generate an empirical distribution of the test statistics under the null hypothesis that no outlier (influential) observations exist in the data, we perform the following algorithm. We chose the parametric bootstrap over the non-parametric approach because the latter may violate the specific measurement error structure (i.e., resampling raw data does not preserve the $X = Z + \Delta$ relationship). Furthermore, the exact finite-sample distribution of these diagnostic statistics under the corrected likelihood framework is analytically intractable.

The number of bootstrap replications is set to $N = 1000$. This criterion is selected based on standard Monte Carlo practices [10], which ensure that the empirical 95th percentile is estimated with a margin of error of approximately $1/\sqrt{N} \approx 0.03$, providing a stable and reliable threshold.

1. Fit the model to the data and calculate LCLE of parameters.
2. Generate a new data vector y^* using the estimated parameters and randomly generated errors.
3. Compute the test statistics (Cook's distances) for the new data and save the order statistics.
4. Repeat steps 2 and 3, N times, for a reasonably large N. This procedure yields an empirical distribution of size N for each order statistic. Calculate the 95th percentile as a threshold.

4 Simulation Study

In this section, a simulation study is performed to evaluate the performance of the proposed diagnostic measures in terms of Type I error and power of the test under different error distributions and sample sizes. The fixed effects variables are generated following [21] to achieve different degrees of collinearity ($\rho = 0.75$). To investigate the robustness of the proposed method, the error terms are generated from three different distributions: Normal, t-distribution with 4 degrees of freedom (heavy-tailed), and Skew-Normal distribution (with a shape parameter of $\alpha = 5$). The Liu biasing parameter d is considered at five different values: 0.1, 0.3, 0.5, 0.7, and 0.9. Two sample sizes are considered: $n = 50$ (with $q = 10$ clusters of size 5) and $n = 100$ (with $q = 10$ clusters of size 10). The simulation study was conducted using the R software. For each combination of parameters, 1000 data sets are generated. The Type I error rate is calculated as the proportion of times the test statistic exceeds the 95th percentile of the empirical distribution under the null hypothesis. To evaluate the power of the test, an outlier is introduced in the first observation, and the power is calculated as the proportion of correct detections.

Comparison with Existing Methods: To rigorously evaluate the performance of the proposed LCLE, we compare it with the standard Corrected Likelihood Estimator (CLE), which corresponds to the Liu estimator with $d = 1$. Under high multicollinearity ($\rho = 0.75$), the parametric bootstrap procedure effectively controls the Type I error rate for both methods. However, a closer inspection reveals that the proposed LCLE ($d = 0.5$) exhibits a more conservative and stable behavior, particularly under non-normal error distributions. For instance, under the Skew-Normal distribution with $n = 50$, the LCLE achieves a Type I error rate of 0.010, compared to

0.028 for the existing CLE, indicating a lower rate of false positives. Furthermore, under heavy-tailed distributions (t-distribution with 4 degrees of freedom), the proposed LCLE demonstrates slightly higher power (0.900) compared to the existing CLE (0.894) for $n = 50$. This confirms that the proposed LCLE provides a more robust diagnostic framework by effectively mitigating variance inflation, leading to comparable or superior detection power without sacrificing error control.

The results of the simulation study are listed in Table 2 and visualized in Figures 1 and 2. It appears that in general, the Type I error rates of the proposed diagnostic measures for different combinations of parameters and error distributions are close to the nominal value of 0.05, indicating that the parametric bootstrap procedure successfully controls the false positive rate. As shown in Figure 1, the Normal and t-distribution exhibit similar patterns, while the Skew-Normal distribution shows slightly more conservative behavior with Type I error rates below 0.03. The results demonstrate that the diagnostic measures maintain better control of Type I error with the larger sample size ($n = 100$) compared to $n = 50$.

Furthermore, the power analysis reveals excellent performance of the proposed diagnostic measures. According to Table 2 and Figure 2, the power values are high across all scenarios, ranging from 0.85 to 1.00. Notably, the method maintains high power even when the normality assumption is violated. The Skew-Normal distribution consistently achieves the highest power (approximately 0.98-1.00), followed by the Normal distribution (0.89-0.95) and the t-distribution (0.85-0.97). All three types of Cook's distance show similar power performance, with the fixed effects measure exhibiting slightly higher power in most cases. The sensitivity analysis with respect to the Liu biasing parameter d indicates that the diagnostic measures maintain stable performance across different values of d, with no systematic pattern suggesting that one particular value of d is universally superior. This demonstrates that the proposed diagnostic measures based on LCLE are robust against deviations from the normality assumption.

To provide a more intuitive visualization of these results, Figures 3 and 4 present heatmaps of the Type I error rates and power values, respectively, for both sample sizes. These heatmaps clearly illustrate that the Skew-Normal distribution consistently yields the most favorable combination of conservative Type I error control and exceptionally high power across all values of the Liu biasing parameter d, particularly for the larger sample size.

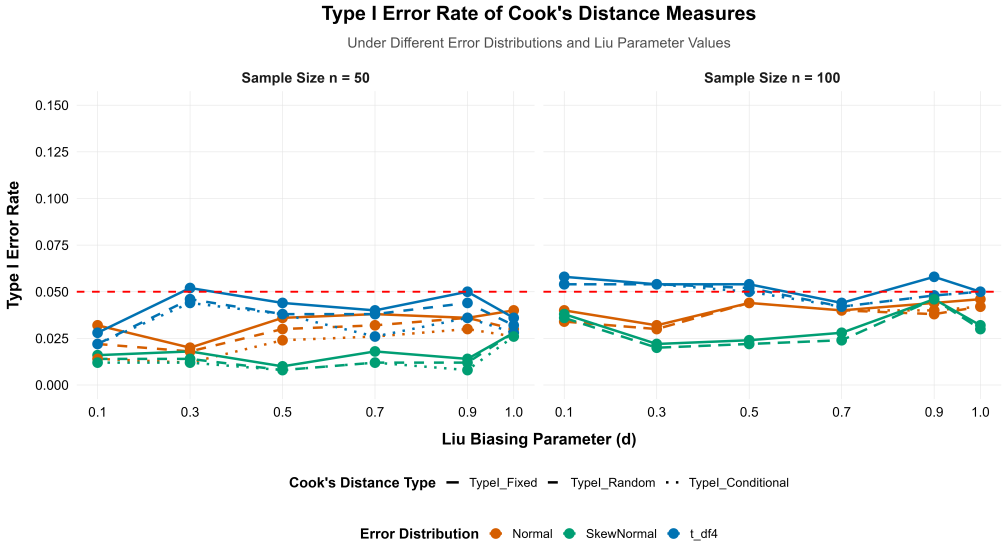


Figure 1: Type I Error Rate of Cook's Distance Measures under Different Error Distributions and Liu Parameter Values for sample sizes $n = 50$ and $n = 100$. The red dashed line represents the nominal level of 0.05.

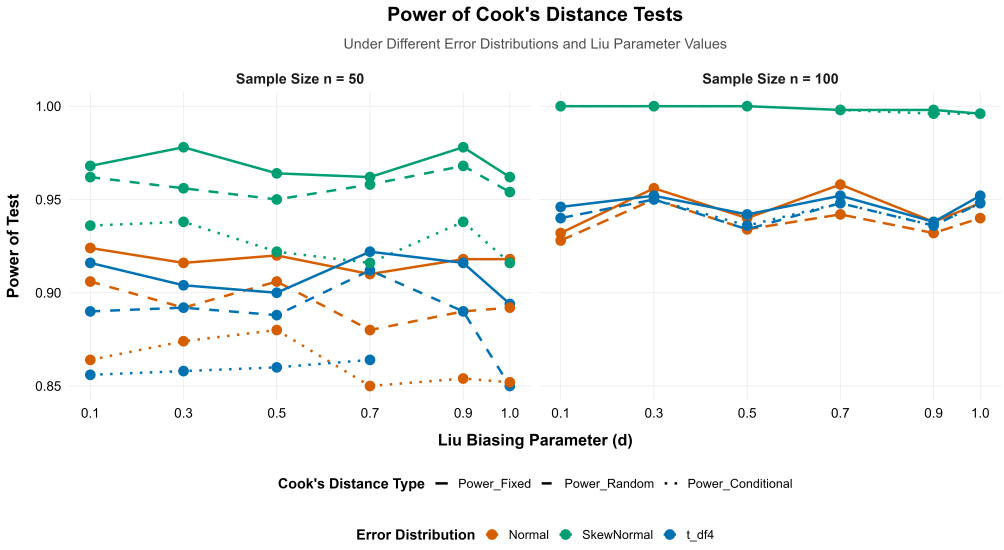


Figure 2: Power of Cook's Distance Tests under Different Error Distributions and Liu Parameter Values for sample sizes $n = 50$ and $n = 100$.



Figure 3: Heatmap of Type I Error Rates for different error distributions and Liu biasing parameter values, for sample sizes $n = 50$ and $n = 100$.

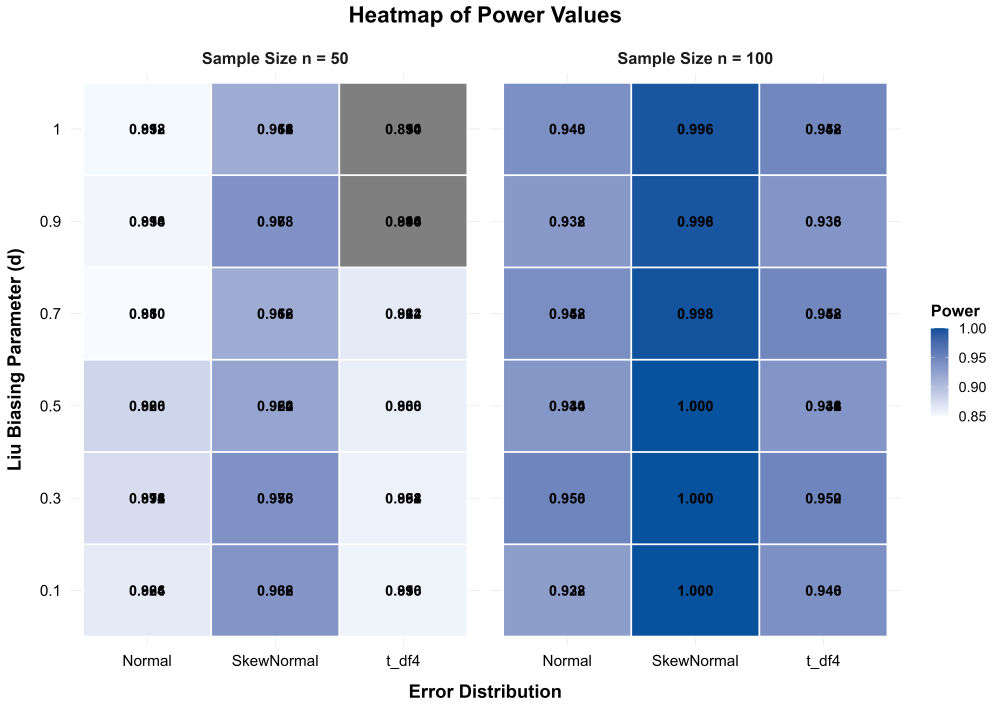


Figure 4: Heatmap of Power Values for different error distributions and Liu biasing parameter values, for sample sizes $n = 50$ and $n = 100$.



Figure 5: Comparison of Type I Error Rate and Power between the proposed LCLE ($d = 0.5$) and the existing CLE ($d = 1$) under high multicollinearity ($\rho = 0.75$). The red dashed line in the top panel represents the nominal level of 0.05.

Table 1: Comparison of Type I error rates and Power between the proposed LCLE ($d = 0.5$) and the existing CLE ($d = 1$) under high multicollinearity ($\rho = 0.75$).

n	Error Dist.	Method	Type I Error		Power	
			Fixed	Random	Fixed	Random
50	Normal	LCLE ($d = 0.5$)	0.036	0.030	0.920	0.906
		CLE ($d = 1$)	0.040	0.030	0.918	0.892
	t (df=4)	LCLE ($d = 0.5$)	0.044	0.038	0.900	0.888
		CLE ($d = 1$)	0.036	0.032	0.894	0.850
	Skew-Normal	LCLE ($d = 0.5$)	0.010	0.008	0.964	0.950
		CLE ($d = 1$)	0.028	0.028	0.962	0.954
100	Normal	LCLE ($d = 0.5$)	0.044	0.044	0.940	0.934
		CLE ($d = 1$)	0.046	0.042	0.948	0.940
	t (df=4)	LCLE ($d = 0.5$)	0.054	0.052	0.942	0.934
		CLE ($d = 1$)	0.050	0.050	0.952	0.948
	Skew-Normal	LCLE ($d = 0.5$)	0.024	0.022	1.000	1.000
		CLE ($d = 1$)	0.032	0.030	0.996	0.996

Table 2: Type I error rates and Power of Cook's distance measures for different combinations of parameters, error distributions, and sample sizes.

n	Error Dist.	d	Type I Error			Power		
			Fixed	Random	Cond.	Fixed	Random	Cond.
50	Normal	0.1	0.032	0.022	0.014	0.924	0.906	0.864
		0.3	0.020	0.018	0.012	0.916	0.892	0.874
		0.5	0.036	0.030	0.024	0.920	0.906	0.880
		0.7	0.038	0.032	0.026	0.910	0.880	0.850
		0.9	0.036	0.036	0.030	0.918	0.890	0.854
		1.0	0.040	0.030	0.026	0.918	0.892	0.852
	t (df=4)	0.1	0.028	0.022	0.022	0.916	0.890	0.856
		0.3	0.052	0.046	0.044	0.904	0.892	0.858
		0.5	0.044	0.038	0.038	0.900	0.888	0.860
		0.7	0.040	0.038	0.026	0.922	0.912	0.864
		0.9	0.050	0.044	0.036	0.916	0.890	0.844
		1.0	0.036	0.032	0.028	0.894	0.850	0.814
	Skew-Normal	0.1	0.016	0.014	0.012	0.968	0.962	0.936
		0.3	0.018	0.014	0.012	0.978	0.956	0.938
		0.5	0.010	0.008	0.008	0.964	0.950	0.922
		0.7	0.018	0.012	0.012	0.962	0.958	0.916
		0.9	0.014	0.012	0.008	0.978	0.968	0.938
		1.0	0.028	0.028	0.026	0.962	0.954	0.916
100	Normal	0.1	0.040	0.034	0.034	0.932	0.928	0.928
		0.3	0.032	0.030	0.030	0.956	0.950	0.950
		0.5	0.044	0.044	0.044	0.940	0.934	0.934
		0.7	0.040	0.040	0.040	0.958	0.942	0.942
		0.9	0.044	0.038	0.040	0.938	0.932	0.932
		1.0	0.046	0.042	0.042	0.948	0.940	0.940
	t (df=4)	0.1	0.058	0.054	0.054	0.946	0.940	0.940
		0.3	0.054	0.054	0.054	0.952	0.950	0.950
		0.5	0.054	0.052	0.050	0.942	0.934	0.936
		0.7	0.044	0.042	0.042	0.952	0.948	0.948
		0.9	0.058	0.048	0.048	0.938	0.936	0.936
		1.0	0.050	0.050	0.050	0.952	0.948	0.948
	Skew-Normal	0.1	0.038	0.036	0.036	1.000	1.000	1.000
		0.3	0.022	0.020	0.020	1.000	1.000	1.000
		0.5	0.024	0.022	0.022	1.000	1.000	1.000
		0.7	0.028	0.024	0.024	0.998	0.998	0.998
		0.9	0.046	0.046	0.046	0.998	0.998	0.996
		1.0	0.032	0.030	0.030	0.996	0.996	0.996

5 Real Data Analysis

To illustrate the performance of the proposed diagnostic measures, we consider the Boston Housing data set, which is widely used in the literature [3, 4, 15, 26, 30]. This data set contains 132 census tracts within 15 districts of the Boston city. The response variable is the median value

of owner-occupied homes, and the predictors include structural and neighborhood characteristics. Following previous studies, the nitrogen oxide level (NOX) is assumed to be measured with error.

We apply the proposed diagnostic measures to this data set under the Liu restriction. Table 3 presents the values of Cook's distance for fixed effects, random effects, and conditional responses for observation 9 (which is identified as a high-leverage point) under different values of the Liu biasing parameter d.

The results in Table 3 indicate that observation 9 has a substantial influence on the parameter estimates, with Cook's distance values ranging from approximately 5.2 to 10.3. Notably, the detection of this influential observation is highly sensitive to the choice of the Liu biasing parameter d. The influence measures increase as d increases from 0.1 to 0.7, reaching a peak at $d = 0.7$ (where $CD_{Fixed} \approx 7.78$ and $CD_{Random} \approx 9.60$), and then decrease at $d = 0.9$ and further drop for the standard CLE ($d = 1.0$). This finding highlights that while the observation remains influential across different values of d, the magnitude of its detected influence varies significantly, emphasizing the critical importance of sensitivity analysis when selecting the biasing parameter in the presence of multicollinearity.

Figure 6 displays the studentized residuals t_i^2 for all 132 observations using $d = 0.5$. The red dashed line represents the 95% threshold at 3.91. Observations with t_i^2 values exceeding this threshold are identified as potential outliers. The plot clearly shows that observation 9 has the largest studentized residual ($t_9^2 = 11.48$), confirming its status as an influential outlier. Additionally, observations 12, 15, 16, 17, 36, 49, 54, 57, 110, and 118 also exceed the threshold, indicating potential outlying behavior in the response variable.

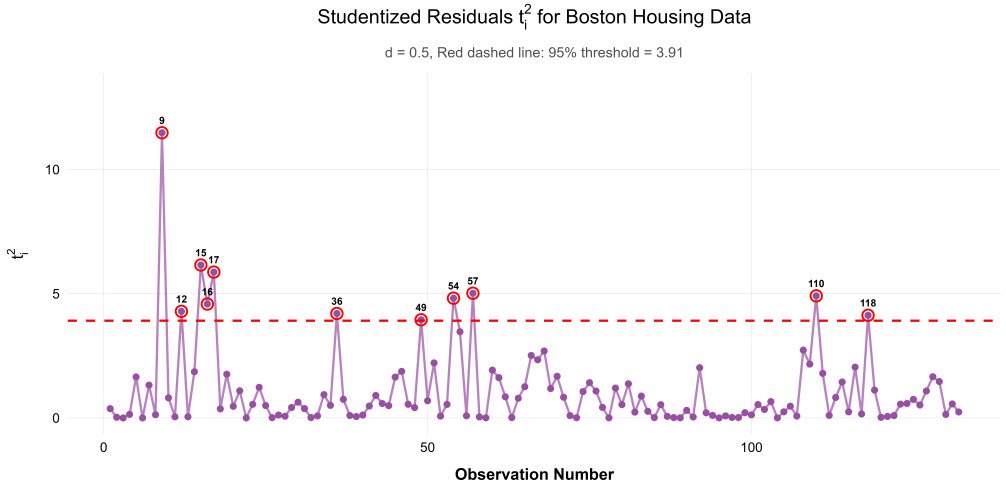


Figure 6: Studentized Residuals t_i^2 for Boston Housing Data with $d = 0.5$. The red dashed line represents the 95% threshold ($F_{0.95}(1, 131) \approx 3.91$). Observations exceeding this threshold are labeled with their observation numbers.

Figure 7 displays the Cook's distance measures for observations 1-30 using $d = 0.5$. Observation 9 clearly stands out as the most influential point across all three diagnostic measures, with Cook's distance values exceeding 7.5 for fixed effects and reaching 10.2 for conditional responses. Additionally, observations 15, 16, and 17 exhibit relatively higher influence compared to the rest of the data, with Cook's distance values ranging from 1.0 to 3.5. These findings are consistent with previous studies [3, 4, 26], which also identified observations 9, 15, and 16 as potentially influential. To further investigate the sensitivity of the most influential point (observation 9) to

the choice of d, we present a detailed analysis in Table 3 and Figure 8.

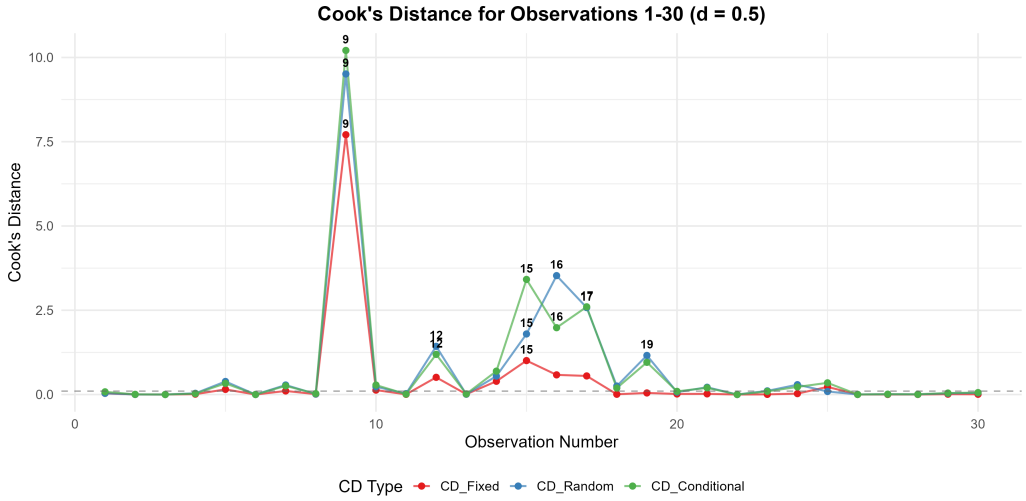


Figure 7: Cook's distance measures for observations 1-30 using $d = 0.5$.

Table 3: Cook's distance measures for observation 9 in the Boston Housing data under different values of the Liu biasing parameter d.

d	CD_Fixed	CD_Random	CD_Conditional
0.1	6.5402	8.0697	8.6619
0.3	7.1437	8.8143	9.4612
0.5	7.7080	9.5106	10.2085
0.7	7.7838	9.6042	10.3090
0.9	6.4832	7.9993	8.5864
1.0	5.2018	6.4184	6.8894

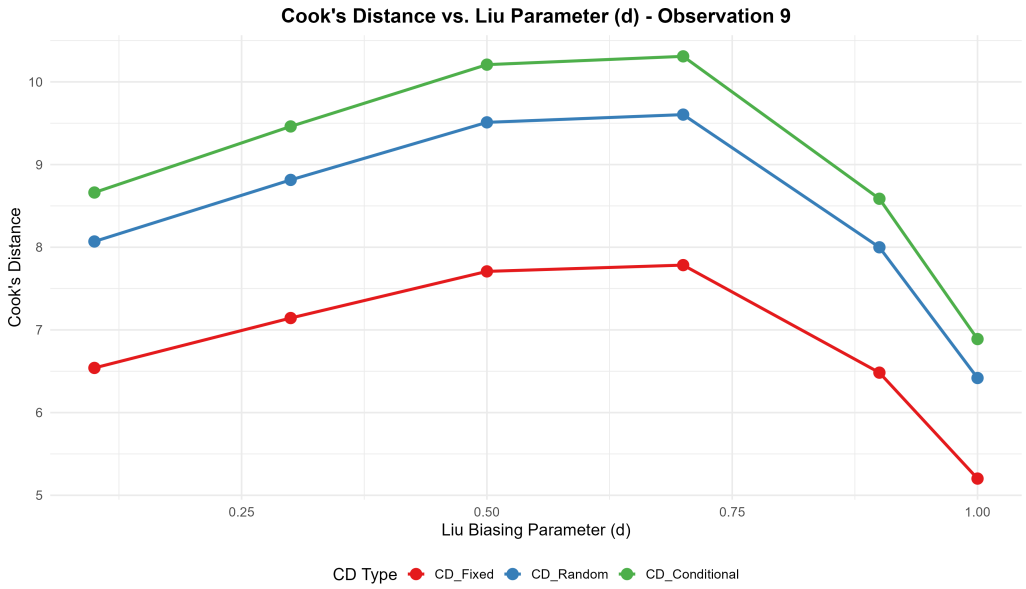


Figure 8: Sensitivity of Cook's Distance measures to the Liu biasing parameter (d) for observation 9 in the Boston Housing data.

6 Conclusion

In this paper, we extended the diagnostic measures for detecting influential observations in linear mixed measurement error models under the Liu estimator. We derived Cook's distance for fixed effects, random effects, and conditional responses. A parametric bootstrap procedure was used to obtain the empirical distribution of the test statistics. The simulation study showed that the proposed methods maintain the Type I error rate close to the nominal level and exhibit high power, even under non-normal error distributions. The real data analysis demonstrated that the identification of influential points depends on the Liu biasing parameter, emphasizing the need for sensitivity analysis in practical applications.

Novelty and Contributions: The primary novelty of this work lies in systematically evaluating how the Liu biasing parameter d interacts with non-normal error distributions (heavy-tailed and skewed) in the context of measurement error models. While previous studies focused on standard estimators, we demonstrated that the LCLE-based diagnostic measures maintain robust performance and high detection power even when the strict normality assumption is violated.

Limitations: Despite its strengths, this study has some limitations. First, the corrected score method assumes that the measurement error covariance matrix (Σ_Δ) is known, which may not always hold in practice. Second, the parametric bootstrap procedure, while highly accurate, can be computationally intensive for very large datasets or complex hierarchical structures.

Future Research: Future research could address these limitations by developing methods for the simultaneous estimation of Σ_Δ alongside the model parameters. Additionally, extending this diagnostic framework to Generalized Linear Mixed Measurement Error Models (GLMMME) for non-continuous responses (e.g., binary or count data) represents a promising and valuable

direction for future work.

Declarations

Author Contributions

Conceptualization: F. Ghapani, R. Jafaraghaie; Methodology: F. Ghapani, R. Jafaraghaie; Software and Validation: R. Jafaraghaie; Formal Analysis: F. Ghapani, R. Jafaraghaie; Investigation: R. Jafaraghaie; Resources: F. Ghapani, R. Jafaraghaie; Data Curation: R. Jafaraghaie; Writing – Original Draft: F. Ghapani, R. Jafaraghaie; Writing – Review & Editing: F. Ghapani, R. Jafaraghaie; Visualization: R. Jafaraghaie; Supervision: R. Jafaraghaie; Project Administration: R. Jafaraghaie; Funding Acquisition: None.

Acknowledgments

The authors received no specific funding for this research. The authors would like to thank all colleagues and reviewers whose constructive comments significantly improved the quality of this manuscript.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Artificial Intelligence Statement

Artificial intelligence (AI) tools, including large language models, were used solely for language editing and improving readability. AI tools were not used for generating ideas, performing analyses, interpreting results, or writing the scientific content. All scientific conclusions and intellectual contributions were made exclusively by the authors.

Publisher's Note

The publisher remains neutral regarding jurisdictional claims in published maps and institutional affiliations.

References

[1] Akdeniz, F., and Erol, H. (2003). Mean squared error matrix comparisons of some biased estimators in linear regression. *Communications in Statistics - Theory and Methods*, 32(12), 2389–2413, doi: https://doi.org/10.1081/STA-120025385.

[2] Banerjee, M., and Frees, E. W. (1997). Influence Diagnostics for Linear Longitudinal Models. *Journal of the American Statistical Association*, 92, 999–1005, doi: https://doi.org/10.1080/01621459.1997.10474055.

[3] Borhani, S., Ghapani, F., and Jafaraghaee, R. (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, doi: https://doi.org/10.22034/jsmta.2024.20632.1115.

[4] Borhani, S., Ghapani, F., and Jafaraghaee, R. (2024). Detecting influential observations using Liu estimator in linear mixed measurement error models. *Journal of Statistical Theory and Practice*, 18(2), 18, doi: https://doi.org/10.1007/s42519-023-00363-2.

[5] Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models*. Chapman and Hall, London.

[6] Cheng, C. L., and Van Ness, J. W. (1999). *Statistical Regression with Measurement Error*. Arnold, London.

[7] Christensen, R., Pearson, L. M., and Johnson, W. (1992). Case-Deletion Diagnostics for Mixed-Models. *Technometrics*, 34, 38–45, doi: https://doi.org/10.2307/1269550.

[8] Cook, R. D. (1977). Detection of influential observations in linear regression. *Technometrics*, 19, 15–18, doi: https://doi.org/10.2307/1268249.

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

[10] Davison, A. C., and Hinkley, D. V. (1997). *Bootstrap Methods and their Application*. Cambridge University Press.

[11] Diggle, P. J., Liang, K. Y., and Zeger, S. L. (1994). *Analysis of Longitudinal Data*. Clarendon Press, Oxford.

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

[13] Fung, W. K., Zhong, X. P., and Wei, B. C. (2003). On estimation and influence diagnostics in linear mixed measurement error models. *American Journal of Mathematical and Management Sciences*, 23, 37–59, doi: https://doi.org/10.1080/01966324.2003.10737603.

[14] Ghapani, F. (2022). Stochastic restricted Liu estimator in linear mixed measurement error models. *Communications in Statistics - Simulation and Computation*, 51(3), 1220–1233, doi: https://doi.org/10.1080/03610918.2019.1664581.

[15] Harrison, D., and Rubinfeld, D. L. (1978). Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management*, 5, 81–102, doi: https://doi.org/10.1016/0095-0696(78)90006-2.

[16] Hoerl, A. E., and Kennard, R. W. (1970). Ridge regression: biased estimation for non-orthogonal problems. *Technometrics*, 12, 55–67, doi: https://doi.org/10.1080/00401706.1970.10488634.

[17] Kaciranlar, S., Sakallioglu, S., Akdeniz, F., Styan, G. P. H., and Werner, H. J. (1999). A new biased estimator in linear regression and a detailed analysis of the widely-analysed data set on Portland Cement. *Sankhya: The Indian Journal of Statistics, Series A*, 61, 443–459, doi: https://www.jstor.org/stable/25053104.

[18] Liu, K. (1993). A new class of biased estimate in linear regression. *Communications in Statistics - Theory and Methods*, 22, 393–402, doi: https://doi.org/10.1080/03610929308831027.

[19] Liu, X. Q., and Hu, P. (2013). General ridge predictors in a mixed linear model. *Journal of Theoretical and Applied Statistics*, 47, 363–378, doi: https://doi.org/10.1080/02331888.2011.592190.

[20] McCulloch, C. E., Searle, S. R., and Neuhaus, J. M. (2008). *Generalized, Linear and Mixed Models*. Wiley, New York.

[21] McDonald, G. C., and Galarneau, D. I. (1975). A Monte Carlo evaluation of some ridge-type estimators. *Journal of the American Statistical Association*, 70(350), 407–416, doi: https://doi.org/10.2307/2285832.

[22] Nakamura, T. (1990). Corrected score function for errors-in-variables models: Methodology and application to generalized linear models. *Biometrika*, 77, 127–137, doi: https://doi.org/10.2307/2336055.

[23] Ozkale, M. R., and Can, F. (2017). An evaluation of ridge estimator in linear mixed models: an example from kidney failure data. *Journal of Applied Statistics*, 44(12), 2251–2269, doi: https://doi.org/10.1080/02664763.2016.1252732.

[24] 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, doi: https://doi.org/10.1525/9780520313880-018.

[25] Yavarizadeh, B., Rasekh, A., Ahmed, S. E., and Babadi, B. (2019). Ridge estimation in linear mixed measurement error models with stochastic linear mixed restrictions. *Communications in Statistics - Simulation and Computation*, 51(6), 3037–3053, doi: https://doi.org/10.1080/03610918.2019.1705974.

[26] Zare, K., and Rasekh, A. (2011). Diagnostic measures for linear mixed measurement error models. *SORT - Statistics and Operations Research Transactions*, 35, 125–144, doi: https://raco.cat/index.php/SORT/article/view/248346.

[27] Zare, K., and Rasekh, A. (2012). Estimation of variance components in linear mixed measurement error models. *Statistical Papers*, 53(4), 849–863, doi: https://doi.org/10.1007/s00362-011-0387-0.

[28] Zare, K., and Rasekh, A. (2014). Residuals and leverages in the linear mixed measurement error models. *Journal of Statistical Computation and Simulation*, 84, 1427–1443, doi: https://doi.org/10.1080/00949655.2012.747203.

[29] Zewotir, T., and Galpin, J. S. (2007). A unified approach on residuals, leverages and outliers in the linear mixed model. *Test*, 16, 58–75, doi: https://doi.org/10.1007/s11749-006-0001-2.

[30] Zhong, X. P., Fung, W. K., and Wei, B. C. (2002). Estimation in linear models with random effects and errors-in-variables. *Annals of the Institute of Statistical Mathematics*, 54, 595–606, doi: https://doi.org/10.1023/A:1022467212133.

Author(s) Bio-sketches

Fatemeh Ghapani received a Ph.D. in Statistics from Shahid Chamran University, Ahvaz, Iran. Her research interests include linear mixed models, measurement error models,

diagnostic measures, and biased estimation methods. She is currently affiliated with the Department of Mathematics and Statistics, Shoushtar Branch, Islamic Azad University, Shoushtar, Iran, as a faculty member. fatemeh.ghapani@iau.ac.ir

Razieh Jafaraghaie received a Ph.D. in Statistics from Allameh Tabataba'i University, Tehran, Iran. Her research interests include computational statistics, robust estimation, multicollinearity, and sensitivity analysis in mixed measurement error models. She is currently affiliated with the Department of Mathematics and Statistics, Shoushtar Branch, Islamic Azad University, Shoushtar, Iran, as a faculty member. Corresponding author: r.jafaraghaee@iau-shoushtar.ac.ir