Robust Estimation Methods in Random Intercept Regression Model: A Comparative Study

Hend A. Auda, Yahia S. El-Horbaty and Eman M. Hanafy
Department of Mathematics, Insurance and Applied Statistics, Helwan University, Egypt

Abstract

Random intercept regression models are used in modeling grouped data where the observations are correlated in each group. This paper presents a comparative simulation study between parametric and two robust estimation methods to assess the influence of the violation of the normality of the error distributions on the efficiency of the model parameter estimates. The asymptotic relative efficiency is used under various factors including the number of groups, the group size and the interclass correlation coefficient. The methods under consideration are applied to data on the faculty of social work's math-achievement at Helwan University.

Keywords: random intercept regression model; restricted maximum likelihood estimation method; robust estimation methods.

Abbreviations: ML: Maximum Likelihood; REML: Restricted Maximum Likelihood; LS: Least Squares; GLS: Generalized Least Squares; MINQUE: Minimum Norm Quadratic Estimation; JR: Robust Joint-Rank; RP: Robust Parametric; ARE: Asymptotic Relative Efficiency; ICC: Intraclass Correlation Coefficient; MSE: Mean Squared Errors; NG: Number of Groups; GS: Group Size; Q-Q: Quantile-Quantile plot.

Introduction

The random intercept regression model is necessary to handle clustered or grouped data. The parameters to be estimated in this model are the fixed coefficients that represent the fixed part of the model and the variances components. In order to estimate unknown parameters, several parametric procedures are well known in literature, but the commonly used methods are maximum likelihood (ML) and restricted maximum likelihood (REML) (Hox, 2002).

A fundamental assumption for tests of significance is normality of the error components distributions involved. However, as any real-life data, data modeled by random intercept regression model might contain outliers, or any other contamination. Even small departures can drive the classic estimates far away from what they would be without the contamination. Robust estimation methods aim to solve these types of problems to provide estimates where contamination has only little influence. A simulation study by Mass and Hox (2003) shows that the non-normality distribution for the residuals at the group level leads to biased estimates of the group level standard errors. That the standard errors of the variances for the group level residuals are highly inaccurate; however the robust estimates do better performance than REML estimates.

Depending on robust rank-based analysis, Mckean et al. (2004) investigated the robust approach introduced by Hettmansperger and Mckean (1978) for linear regression models. The study compared the asymptotic relative efficiency (ARE), as a principal comparison measure, of the robust rank-based estimation method and least square (LS) estimators in terms of their asymptotic variances. The authors concluded that, under normality the robust estimator losses 4.5% in efficiency since it provides 95.5% as efficient as LS estimators. However, the ARE is usually greater than 1 if the true distribution has tails heavier than those under a normal distribution, when errors have contaminated normal distributions or the data are corrupted by outliers.

Under linear mixed models, which represent the general representation of random intercept regression models, Kloke et al. (2009) extended the robust approach from simple linear models to linear mixed models with covariates using general score functions. The authors compared between the LS and robust rank-based estimation method using Wilcoxon score function by ARE as an assessment measure and they concluded that the robust analysis is more efficient than LS in the presence of outliers.

Besides, Mckean and Kloke (2014) proposed a family of optimal score functions under contaminated normal distributions of the error terms in both linear and nonlinear models. In this study, they compared robust rank-based estimators with LS and ML estimation methods in terms of their asymptotic variances by using the ARE to conclude about the efficiency and validity of robust rank-based method over skewed normal and symmetric contaminated normal distributions. Mckean and Kloke (2014) illustrated how to apply robust non-parametric statistical methods in linear, nonlinear and mixed regression models using the R-package.

Furthermore, Auda et al. (2018) provided a simulation study to compare between the REML and the robust rank-based procedure denoted by joint-rank (JR) under different situations of error distributions including normality, contaminated normal distribution, skewed contaminated normal distribution and when data corrupted by two types of outliers. The study concluded that the JR fit is more efficient and powerful than REML under all non-normal cases. However, it loses little of its power and efficiency in normal errors.

In this paper we extend the investigation simulation study produced by Auda et al. (2018) to cover another robust estimation method beside JR called robust parametric (RP) method. Accordingly, we organize the rest of this paper as follows. In Section 2, we describe the random intercept regression model. In section 3, we discuss the REML and robust estimation methods for random intercept regression model. In Section 4, we offer a simulation study that confirms the validity of our analysis (robustness of parameter estimates) under different situations of error distributions and the results of the study are presented in Section 5. In Section 6, we examine a real dataset obtained from the faculty of Social Work that contains several outliers and violate normality assumption; where the traditional REML analysis is sensitive to these outliers, whereas the other rank-based methods are robust. And finally in Section 7 we discuss the issues raised by the results of this study and draw some conclusions about this analysis.

1. Random Intercept Regression Model

Assume that we have data from m groups, with a different number of respondents n_j , where $j = 1,2,,\dots,m$ and $n = \sum_{j=1}^{m} n_j$. Then the data can be modeled using random intercept regression model as

$$Y_j = \alpha \mathbf{1}_{n_j} + X_j \beta + \delta_j \mathbf{1}_{n_j} + \varepsilon_j \qquad , j = 1,2,,\dots,m \qquad [1]$$

where Y_j is $n_j \times 1$ vector of response variables, X_j is $n_j \times p$ design matrix of explanatory variables, α is the intercept parameter, β is $p \times 1$ vector of regression coefficient of fixed effects, δ_j represents the random effect of group j where $\mathbf{1}_{n_j}$ is n_j vector of ones and ε_j is a vector of length n_j representing the residuals on the same group level. It is further assumed that $\delta_j \sim N(0, \sigma_0^2)$ and is uncorrelated with ε_j such that $\varepsilon_j \sim N(\mathbf{0}, \sigma_e^2 I_{n_j})$. The previous model can also reformulated as

$$Y_j = \alpha \mathbf{1}_{n_j} + X_j \beta + e_j \qquad , j = 1,2,,\dots,m \qquad [2]$$

where $e_j = \delta_j \mathbf{1}_{n_j} + \varepsilon_j$ represents the vector of all random errors in group j. Combining model [1] for all groups yields

$$Y = \alpha \mathbf{1}_n + X \beta + \delta \mathbf{1}_n + \varepsilon = \alpha \mathbf{1}_n + X \beta + e \qquad [3]$$

where $Y = (Y_1^T, \dots, Y_m^T)^T$ is $n \times 1$ vector of responses, $X = (X_1^T, \dots, X_m^T)^T$ is $n \times p$ the design matrix of explanatory variables. Because an intercept parameter is in the model, we assume that X is centered and has full column rank. $\varepsilon = (\varepsilon_1^T, \dots, \varepsilon_m^T)^T$ is $n \times 1$ vector of residual errors, δ is $m \times 1$ vector of random effects and $\delta \sim N(\mathbf{0}, \sigma_0^2 \mathbf{1}_n)$, where $\mathbf{1}_n$ is $n \times n$ matrix of ones. So under normality of residual errors, the response variable has normal distribution with mean $(\alpha \mathbf{1}_n + X \beta)$ and variance covariance matrix Σ_Y, that

$$\Sigma_Y = \sigma_0^2 \mathbf{1}_n + \sigma_e^2 I_n \qquad [4]$$

Because of the hierarchical structure of the multilevel data in random intercept regression models, the observations violate the assumption of independence within groups. The amount of dependence can be measured using intraclass correlation coefficient (ICC) which can be expressed under random intercept models as follows,

$$\text{ICC} = \rho = (\sigma_0^2 / (\sigma_0^2 + \sigma_e^2)) = \sigma_0^2 / \sigma^2 \qquad [5]$$

where $\sigma^2 = \sigma_0^2 + \sigma_e^2$, then we can reformulate [4] according to [5] as follows;

$$\Sigma_Y = \sigma^2[\rho \mathbf{1}_n + (1-\rho)\boldsymbol{I}_n] \qquad [6]$$

In the next section, we present a parametric and two robust estimation methods that will be used in the simulation study.

2. Estimation Methods

Once the model has been formulated, methods are needed to estimate the model parameters. Several estimation methods, as generalized least square (GLS) and Henderson's mixed model equations (Henderson, 1950), have been produced for estimating fixed and random effects simultaneously. These methods assume that the variance components of the model parameters are known, however in practice they are usually unknown.

In order to estimate unknown parameters, namely the fixed effects and the variance components, several procedures for variance parameter estimation are discussed in Searle et al. (1992). These methods include the ANOVA method for balanced data which uses the expected mean squares approach. On the other hand, Rao (1971) proposed the minimum norm quadratic estimation (MINQUE) for estimating variance parameters in case of unbalanced data. However, for both balanced and unbalanced data we can use maximum likelihood ML and restricted maximum likelihood REML estimation methods a. Below we will describe REML as a parametric method for estimate variance parameters.

Restricted Maximum Likelihood

ML and REML are common estimation methods for estimating fixed effects parameters, as well as estimating the unknown variance components. ML estimators of the variance parameters are always biased because they do not take into account the degrees of freedom lost in the estimation of the fixed effects (Lin and Allister, 1984; Swallow et al, 1984). This problem can be overcome by REML (Anderson, 1952; Patterson, 1971) since it takes into account the degrees of freedom lost in estimating the fixed effects. Hence we depend on REML estimation of the variance parameters instead of ML estimation. REML estimation includes no procedure for estimating fixed effects. However, maximum likelihood estimation of fixed effects parameters are achieved by replacing variance components by their REML estimates.

An important assumption underlying REML estimation method is the normality of the error distributions. When the residual errors are not normally distributed, the parameter estimates produced by REML are asymptotically unbiased. However, the asymptotic standard errors are incorrect and significance tests and confidence intervals using those standard errors cannot be trusted (Goldstein, 2010). This problem does not completely vanish when the sample gets larger. Accordingly, we discuss two of robust estimation methods called robust parametric (RP) fit and joint rank-based JR fit Kloke et al. (2009).

Robust Rank-Based Estimation Method

Rank-based fitting of linear models offers an alternative estimation method to least squares and maximum likelihood estimation methods. These methods (rank-based) need the selection of a score function $\varphi(u)$. If the form of the underlying error distribution is known, we can obtain an optimal score function which minimizes the variance of the estimator. For example, if the error distribution is normal, then the optimal score function is the normal scores which defined as; $\varphi_{ns}(u) = \Phi^{-1}(u)$, where $\Phi(u)$ denotes the standard normal cumulative distribution function, while Laplace distributed errors produces the sign scores$\left(\varphi_{sgn}(u) = sgn\,(u - 1/2)\right)$. However, if there is little or no information about the error distribution, then Wilcoxon score function is used $\left(\varphi_w(u) = \sqrt{12}\,(u - 1/2) \text{ and } 0 < u < 1\right)$.

The geometry of rank-based fit is similar to that of least squares estimation methods in linear regression models. Since we replace Euclidean-norm by the pseudo-norm as

$$\|C\|_{\varphi} = \sum_{i=1}^{n} a\left(R(c_i)\right) c_i \qquad , C \in R^n \qquad [7]$$

where the scores are generated as $a(i) = \varphi(\mathrm{i}/(\mathrm{n}+1)$ where $\varphi(u)$ is defined as non-decreasing square-integrable score function defined on the interval $(0, 1)$, where, without loss of generality, standardize as $\int_0^1 \varphi(u)\,du = 0$ and $\int_0^1 \varphi^2(u)\,du = 1$, and $R(c_i)$ represents the rank of c_i among $c_1, c_2, \ldots, c_n$. Then we can define the rank-based estimate of fixed coefficient parameters $\boldsymbol{\beta}$ as follows

$$\widehat{\boldsymbol{\beta}}_{\varphi} = Argmin_{\,\beta\in\mathbb{R}}\, \|Y - X\boldsymbol{\beta}\|_{\varphi} \qquad [8]$$

then the asymptotic distribution of the rank based estimator is given by

$$\widehat{\boldsymbol{\beta}}_{\varphi} \sim N\left(\boldsymbol{\beta},\ \tau_{\varphi}^2 (X^T X)^{-1}\right) \qquad , \qquad \tau_{\varphi} = \left(\int \varphi(u)\varphi_f(u)du\right)^{-1}$$

where $\varphi(u)$ is the score function and $\varphi_f(u) = \frac{f'\left(F^{-1}(u)\right)}{f\left(F^{-1}(u)\right)}$. The τ_{φ} is the scale parameter of random error term that under the Wilcoxon scores we can simplify the scale parameter τ_{φ} to;

$$\tau_{\varphi} = \tau_W = \left[\sqrt{12} \int f^2(t)dt\right]^{-1} \qquad [9]$$

$$\tau_s = \left[2 f(0)\right]^{-1} \qquad [10]$$

where τ_s is the scale parameter of the estimate of intercept parameter $\hat{\alpha}_s$, which defined as the location estimate based on the residuals. For LS, the arithmetic mean is used while for the rank-based estimates the median is used $\left(\hat{\alpha}_s = med_{1\le i\le n}\left(y_i - x_i^T \widehat{\boldsymbol{\beta}}_{\varphi}\right)\right)$.

Based on the previous rank-based regression methodology, Kloke et al. (2009) extend it to include random effects in mixed models. In this section we introduce joint rank-based estimation method.

Robust Joint Ranking Method

This rank-based analysis has been extended for nested mixed models by Kloke et al. (2009). A full development of the rank-based analysis can be found in Chapters 3-5 of the monograph by Hettmansperger and McKean (2011). The idea of this method is to fit fixed effects of the model at first then the variance-covariance of the rank-based fit and the variance components are robustly estimated based on the residuals of the fixed effects fit (Kloke et al., 2009 and Auda et al., 2018).

By following the algorithm of rank-based estimation method as shown in Kloke et al. (2009), we first find the estimation of fixed effects in the JR method using the dispersion function as in the independent linear model (see [8]) and their variances as appear in [9] and [10], so the asymptotic distribution of $\widehat{\boldsymbol{\beta}}_{\varphi}$ is normal with mean $\boldsymbol{\beta}$ and covariance matrix $\boldsymbol{V}_{\varphi}$ which is defined in general mixed models as:

$$V_{\varphi} = \tau_{\varphi}^2 (X^T X)^{-1} \left(\sum_{j=1}^{m} X_j^T \Sigma_{\varphi,j} X_j\right) (X^T X)^{-1} \qquad [11]$$

where $\Sigma_{\varphi,j} = cov\left(\varphi\left(F(e_j)\right)\right)$ and $F(e_j)$ denotes the distribution function of errors. Then, as linear models, once $\boldsymbol{\beta}$ is estimated, we estimate the intercept α as the median of the residuals that, $\hat{\alpha}_s = med_{ij}(y_{ij} - x_{ij}^T \widehat{\boldsymbol{\beta}}_{\varphi})$. Letting $\tau_s = 1/2f(0)$ (see [10]), $\hat{\alpha}_s$ is asymptotically normal with mean α and variance

$$\sigma_1^2 = \tau_s^2 \frac{1}{n} \sum_{j=1}^{m} \left[\sum_{i=1}^{n_j} var\left(sgn(e_{ij})\right) + \sum_{i \neq i} cov\left(sgn(e_{ij}), sgn(e_{i'j})\right)\right] \qquad [12]$$

Also to conduct inference, we need an estimate of the covariance matrix of $\widehat{\boldsymbol{\beta}}_{\varphi}$. And as we said in the previous section, we robustly estimate the variance-covariance of the rank-based fit and the variance components based on the residuals of the fixed effects, then by define the residuals of the JR fit we obtain

$$\hat{e}_{JR} = Y - \hat{\alpha}_s \mathbf{1}_n - X \widehat{\boldsymbol{\beta}}_{\varphi} \qquad [13]$$

Using these residuals, we can estimate the parameter τ_{φ} and τ_s by their estimators as proposed by Koul et al. (1987). Next, a nonparametric estimate of $\Sigma_{\varphi,j}$ is obtained by replacing the distribution function $F(e_j)$ by the empirical distribution function of the residuals.

However, for the simulation study in this paper depended on random intercept regression model as defined in [3] which represents a special case of linear mixed models. So the asymptotic variance-covariance matrix defined in [11] can be simplified as

$$V_{\varphi} = \tau_{\varphi}^{2}(X^{T}X)^{-1}\left(\sum_{j=1}^{m} X_j^{T}\Sigma_{\varphi,j}X_j\right)(X^{T}X)^{-1} \quad , \quad \Sigma_{\varphi,j} = \Sigma_{Y_j} = \sigma^{2}\left[\rho_{\varphi}1_{n} + \left(1-\rho_{\varphi}\right)I_{n}\right] \quad [14]$$

as defined in [6] for group j, and the intraclass correlation coefficient for each two residuals defined as; $\rho_{\varphi} = cov\left\{\varphi\left(F(e_{11})\right), \varphi\left(F(e_{21})\right)\right\} = E\left\{\varphi\left(F(e_{11})\right), \varphi\left(F(e_{21})\right)\right\}$. Similarly, for the asymptotic variance of the intercept in [12] can simplify at

$$\sigma_1^2 = \tau_s^2 \frac{1}{n} \left(1 + n^{*}\rho_s^{*}\right) \quad [15]$$

for ρ_s^{*} defined for each two residuals defined as; $\rho_s^{*} = cov\left\{sgn(e_{11}), sgn(e_{21})\right\}$ and $n^{*} = n^{-1}\sum_{j=1}^{m} n_j\left(n_j - 1\right)$. Let $M = \sum_{j=1}^{m}\binom{n_j}{2} - p$, then the simple moment estimators of ρ_{φ} and ρ_s^{*} are

$$\hat{\rho}_{\varphi} = M^{-1}\sum_{j=1}^{m}\sum_{i>i} a\left[R\left(\hat{e}_{ij}\right)\right]a\left[R\left(\hat{e}_{ij}\right)\right] \quad [16]$$

$$\hat{\rho}_s^{*} = M^{-1}\sum_{j=1}^{m}\sum_{i>i} sgn\left(\hat{e}_{ij}\right)\, sgn\left(\hat{e}_{ij}\right) \quad [17]$$

Plugging this into [14] and using the estimate of τ_{φ} discussed in previous section, we have an estimate of the asymptotic covariance matrix of the JR estimators.

Robust Parametric method

Robust parametric RP method is based on huberization of likelihood estimation method (Koller, 2016); it is depend on the random effects contamination model. The estimation method does not make any assumption on the data's grouping structure except that the model parameters are estimable and it supports hierarchical and non-hierarchical grouping structures. The robustness of the estimates and their asymptotic efficiency is fully controlled through the function interface. Individual parts (fixed effects and variance components) can be tuned independently. A full development for how to fit robust linear mixed-effects models using RP can be found in Koller (2016).

3. Simulation Design and Procedure

The purpose of this paper is to provide a comparative simulation study to investigate the performance of specific parametric and robust estimation methods based on the accuracy of the parameter estimates and their standard errors when some fundamental assumptions are violated.

We use a simple two-level random intercept regression model with one explanatory variable at the individual level, conforming to the following combined equation:

$$y_{ij} = \alpha + \beta_1 x_{ij} + \delta_j + \varepsilon_{ij} \quad [18]$$

that the first part $\left(\alpha + \beta_1 x_{ij}\right)$ in [18] contains fixed coefficients; it is the fixed part of the model. For $\left(\delta_j + \varepsilon_{ij}\right)$ in [18] contains random error terms; it is the random part of the model. The performance is investigated under the following factors: (1) Number of groups, (2) Group size and (3) Interclass correlation coefficient.

A former simulation study of Van der Leeden et al. (1997) showed that a large number of groups is more important for the efficiency of the parameter estimates than the large number of individuals per groups, so that the highest number of groups should be sufficient. Furthermore, the simulation in Mass and Hox (2003) showed only a small sample size at the group level leads to biased estimates of the group-level standard errors. According to that three conditions of number of groups (NG) are varied (NG = 10, 20 and 30).

Similarly, for the factor of group size (GS), three sizes are used in the simulation (GS = 5, 10 and 15). The group sizes are chosen so that the highest number should be sufficient.

A recent simulation study of Auda et al. (2018) suggests that the size of intraclass correlation coefficient ICC also affects the accuracy of the parameter estimates and their standard errors. Actually, what is at issue in multilevel modeling (in general) is not so much the ICC, but the design effect, which indicates how much the standard errors are underestimated (Kish, 1965). In group data, the design effect is approximately equal to 1 + (average group size – 1) * ICC. Thus, the values for the ICC and group size are determined to make the design effect larger than two. Three values of ICC are used (0.1, 0.2 and 0.3).

For each condition, we generate 1000 simulated data sets, assuming normally distributed residuals. The model assumes that the explanatory variables are fixed in repeated samples, and are randomly generated from uniform distribution. The regression coefficients values are 1.00 for the intercept and 0.3 for the slope. Different combinations of σ_e^2 and σ_0^2 have been chosen such that according to [5] the resulting ICC becomes 0.1, 0.2 and 0.3 respectively.

In the next section, we summarize the results of empirical AREs and the relative bias of the parameter estimates in many situations: the standard case of normal errors, symmetric contaminated normal errors with three levels of contamination (10%, 20% and 30%) and the ratio of the contaminated standard deviation to uncontaminated standard deviation was set at 10, skewed contaminated normal distribution with also three levels of contamination (10%, 20% and 30%) and skewnees parameter set at 10, and finally we investigate the influence of outliers on the parameter estimates and the model efficiency. For last situation, we corrupted the normal errors with two-types of outliers as follows: first we replaced 5% of the random errors with those drawn from the normal distribution with mean 10 and variance 15^2, and then replaced random effects that belong to a specific group with random errors drawn from the normal distribution with mean 10 and variance 15^2.

For each case of previous we compared REML, RP and JR estimation methods using the R statistical package. The function `lmer` from the `lme4` package (Bates et al., 2015) is used to run REML analysis since it doesn't make assumption on grouping structure and efficiently deals with correlated and uncorrelated random effects within levels. However, we use the function `rlmer` to compute the RP analysis. Finally, we use `jrfit` function from `jrfit` package to compute JR analysis based on Wilcoxon score functions Kloke et al. (2009).

4. Results

In order to assess the importance of the normality assumption of the error distributions involved in the model, the section is divided into two parts. In first part; the accuracy of the parameter estimate are investigated using the average bias, and then in the second part, efficient of the model has been investigated using asymptotic relative efficiency.

Average Bias

The relative bias is calculated for across all cases discussed in the previous section, and we calculated it as follows; for each case of situations we generate 1000 simulated data sets and defined the average bias as follows:

$$Average\ bias = \frac{1}{1000} \sum_{i=1}^{1000} \left(E(\hat{\theta}) - \theta\right) \qquad [19]$$

where θ denote the population parameter and $\hat{\theta}$ is the parameter estimate.

The results showed that, for normal and contamination situations presented in Tables (1, 2 and 3) are nearly to zero and almost the same for REML, RP and JR estimators. Also it is important to say that we obtain the lowest values of bias for the parameter estimate by using robust JR approach in contamination situation as appeared in both Tables (2 and 3).

Table (1): Bias for fixed effect estimator (Slope) with normal errors

ICC	Number of groups	Group sizes 5			10			15		
		REML	RP	JR	REML	RP	JR	REML	RP	JR
0.1	10	0.0497	0.0517	0.0473	0.007	0.013	0.017	0.031	0.0285	0.0353
	20	0.0028	0.0062	0.0075	0.0031	0.0037	0.0015	0.0209	0.0204	0.026
	30	0.0321	0.0252	0.0377	0.0315	0.0378	0.0388	0.0199	0.0227	0.0237
0.2	10	0.0551	0.0576	0.0393	0.0075	0.012	0.0207	0.0294	0.0269	0.038
	20	0.0039	0.0059	0.0029	0.0017	0.0027	0.0011	0.0175	0.0163	0.0253
	30	0.0174	0.0154	0.0303	0.0189	0.0219	0.0238	0.0115	0.0127	0.0161
0.3	10	0.0568	0.0581	0.0386	0.0065	0.0111	0.0217	0.0273	0.0252	0.0416
	20	0.0047	0.0073	0.0028	0.0014	0.0025	0.0010	0.0176	0.0163	0.0278
	30	0.0138	0.0129	0.0329	0.0177	0.0201	0.0237	0.0106	0.0117	0.0161

Table (2): Bias for fixed effect estimator (Slope) with contaminated normal error

Level of Contamination	Number of groups	Group sizes								
		5			10			15		
		REML	RP	JR	REML	RP	JR	REML	RP	JR
10%	10	0.083	0.051	**0.038**	0.056	0.0233	**0.0231**	0.056	0.025	**0.018**
	20	0.054	0.034	**0.023**	0.0274	0.0012	**0.0003**	0.006	**0.002**	**0.002**
	30	0.078	0.025	**0.018**	0.0052	0.0053	**0.0026**	0.043	0.0078	**0.0076**
20%	10	0.08	0.06	**0.04**	**0.001**	0.033	0.027	0.076	0.043	**0.027**
	20	**0.0014**	0.0252	0.0271	0.034	0.003	**9.5E-05**	0.05	0.018	**0.007**
	30	0.076	0.041	**0.027**	0.05	0.03	**0.007**	0.021	0.009	**0.004**
30%	10	0.047	0.078	0.043	**0.013**	0.044	0.035	0.09	0.051	**0.036**
	20	**0.016**	0.037	0.035	0.056	0.014	**0.009**	0.062	**0.01**	**0.01**
	30	0.09	0.065	**0.035**	0.060	0.019	**0.0108**	0.008	0.011	**0.005**

Table (3): Bias for fixed effect estimator (Slope) with skewed contaminated normal error

Level of Contamination	Number of groups	Group sizes								
		5			10			15		
		REML	RP	JR	REML	RP	JR	REML	RP	JR
10%	10	0.016	0.021	**0.0021**	0.0092	**0.0006**	0.0092	0.009	0.0015	**0.0006**
	20	0.008	**0.001**	0.0091	0.014	0.0064	**0.0053**	0.031	0.0074	**0.0014**
	30	0.01	0.0034	**0.0006**	0.029	0.003	**0.0014**	0.009	0.007	**0.004**
20%	10	0.04	0.038	**0.018**	0.021	0.019	**0.015**	0.0296	0.0219	**0.002**
	20	0.026	**0.0029**	0.015	0.0067	0.0048	**0.0012**	0.037	0.0135	**0.006**
	30	0.025	**0.001**	0.002	0.037	0.016	**0.006**	0.019	0.0031	**0.0025**
30%	10	0.08	0.068	**0.03**	**0.025**	**0.001**	0.007	0.031	0.009	**0.000**
	20	0.02	0.014	**0.007**	0.012	**0.004**	0.007	0.0195	0.024	**0.008**
	30	0.034	0.021	**0.0004**	0.019	0.010	**0.008**	0.033	0.008	**0.002**

Although when the data corrupted by outliers, the bias results begin to differ from zero and the parameter estimates far away from what it would be without outliers. In Table (4) we presented the bias results for *Type-one outliers*: where 5% of random errors replaced with random errors drawn from the normal distribution with mean 10 and variance15^2, and the results showed that the performance of robust estimation methods is better than the performance of REML across all factors (the number of groups, group sizes and interclass correlation coefficients). In addition, by comparing the two approaches of robust estimation methods, JR approach appears to give the better performance with a little difference than RP approach in the worst conditions with small sample sizes and an ICC = 0.1.

Table (4): Bias for fixed effect estimator (Slope) with data corrupted by Type-one outliers

ICC	Number of groups	Group sizes								
		5			10			15		
		REML	RP	JR	REML	RP	JR	REML	RP	JR
0.1	10	0.657	0.112	**0.079**	0.448	0.098	**0.076**	0.47	0.011	**0.005**
	20	0.469	0.0896	**0.0894**	0.721	0.169	**0.152**	0.722	0.149	**0.131**
	30	0.056	**0.023**	**0.023**	0.732	**0.158**	**0.158**	0.397	0.089	**0.082**
0.2	10	0.696	0.133	**0.01**	0.44	0.101	**0.095**	0.064	0.017	**0.0098**
	20	0.432	**0.079**	0.111	0.734	0.176	**0.174**	0.755	0.157	**0.147**
	30	0.078	0.028	**0.022**	0.734	**0.163**	0.175	0.427	0.098	**0.092**
0.3	10	0.74	0.163	**0.119**	0.426	**0.101**	0.108	0.083	0.019	**0.013**
	20	0.382	**0.07**	0.126	0.745	**0.178**	0.191	0.78	0.17	**0.158**
	30	0.103	0.044	**0.021**	0.736	**0.162**	0.191	0.45	0.101	**0.100**

Table (5): Bias for fixed effect estimator (Slope) with data corrupted by Type-two outliers

ICC	Number of groups	Group sizes								
		5			10			15		
		REML	RP	JR	REML	RP	JR	REML	RP	JR
0.1	10	0.901	0.158	**0.041**	1.065	0.199	**0.099**	1.00	0.193	**0.142**
	20	0.478	0.033	**0.019**	0.48	0.032	**0.024**	0.464	**0.020**	0.0498
	30	0.286	**0.016**	0.035	0.301	0.026	**0.025**	0.364	**0.016**	0.045
0.2	10	0.897	0.169	**0.038**	1.06	0.211	**0.128**	1.006	0.212	**0.174**
	20	0.474	0.042	**0.021**	0.475	0.038	**0.027**	0.46	**0.026**	0.055
	30	0.287	**0.023**	0.043	0.299	0.031	**0.025**	0.359	**0.021**	0.048
0.3	10	0.895	0.184	**0.035**	1.06	0.225	**0.140**	1.01	0.232	**0.201**
	20	0.472	0.050	**0.017**	0.471	0.045	**0.023**	0.456	**0.033**	0.055
	30	0.29	**0.031**	0.055	0.296	0.037	**0.030**	0.355	**0.027**	0.049

Also, the same conclusion is obtained in Table (5) for the data corrupted by ***Type-two outliers*** (where 5% of random effects replaced with random effects drawn from the normal distribution with mean 10 and variance15^2) that the worst estimation results with biggest vales of bias are obtained by using REML, however the almost the same accurate results are obtained by using the other two robust estimation methods with prefer to JR approach results especially in small sample sizes with 5 and 10 observations per group. For all previous situations we have a general conclusion that the average of empirical bias decreases as the sample size and ICC increase.

Actually, the results show that non-normal errors of the random intercept regression model have a little or no effects on the parameter estimates since we focus on investigating the parameter estimates involved in the fixed part of the model and not their standard errors or variance components. So to assess the accuracy of the model efficiency with respect to its variance, we compare between the different estimation methods using the asymptotic relative efficiencies AREs in terms of the asymptotic variances as presented in next section.

Asymptotic Relative Efficiency

To compare between estimation methods presented in this paper the asymptotic relative efficiency ARE is used since it defined as one of the principal comparison measures. And we calculated it as follows; for each situation we generate 1000 simulated data sets and defined ARE between every two different estimation methods as the ratio of their asymptotic mean square errors(MSE) as follows:

$$ARE = MSE_r / MSE_{\acute{r}} \quad , \; r \neq \acute{r} \qquad [20]$$

where $MSE_r = \frac{1}{1000} \sum_{i=1}^{1000} \left((\hat{\theta}_{ri} - \theta)^2 \right)$ represent the mean square error for r estimation method, θ denote the population parameter and $\hat{\theta}_{ri}$ is the parameter estimate using r estimation method on the ith simulation data set. So if ARE is less than 1, then the estimators given by r estimation method are more efficient than estimators given by $\acute{r}$ and vice versa. In this paper, we have three values of ARE in each case study that compare between REML, RP and JR as follows; $ARE(1) = MSE_{REML}/MSE_{RP}$, $ARE(2) = MSE_{REML}/MSE_{JR}$ and $ARE(3) = MSE_{RP}/MSE_{JR}$.

Table (6): Asymptomatic relative efficiency for normal errors

ICC	Number of groups	Group sizes								
		5			10			15		
		ARE(1)	ARE(2)	ARE(3)	ARE(1)	ARE(2)	ARE(3)	ARE(1)	ARE(2)	ARE(3)
0.1	10	0.96	0.97	1.01	0.977	0.99	1.02	0.981	0.98	0.999
	20	0.977	0.938	0.96	0.989	0.987	1.096	0.993	0.97	0.977
	30	0.979	0.921	0.94	0.987	0.984	0.998	0.922	0.95	1.03
0.2	10	0.962	0.943	1.04	0.977	1.001	1.02	0.971	0.998	1.028
	20	0.997	0.95	0.953	0.986	0.99	1.004	0.984	0.98	0.996
	30	0.983	0.94	0.956	0.987	0.98	0.992	0.992	0.96	0.968
0.3	10	0.96	0.98	1.02	0.924	1.002	1.084	0.965	0.99	1.026
	20	0.981	0.964	0.983	0.984	0.995	1.011	0.984	0.987	1.003
	30	0.98	0.964	0.985	0.986	0.99	1.004	0.989	0.97	0.981

where $ARE(1) = MSE_{REML}/MSE_{RP}$, $ARE(2) = MSE_{REML}/MSE_{JR}$ and $ARE(3) = MSE_{RP}/MSE_{JR}$

Using this measure, Table (6) provides the empirical AREs of REML and the two robust approaches estimation methods in normal errors situation, the results show that ARE (1) and ARE (2) values for all factors conditions are less than 1 which indicate that the REML estimates of variance components are more efficient than corresponding estimators produced by robust methods where the REML has the lowest MSE.

However, in Table (7) we present the ARE results of symmetric contaminated normal distribution in which errors are contaminated by three levels of contamination (10%, 20% and 30%) and the ratio of the contaminated standard deviation to uncontaminated standard deviation was set at 10 as illustrated in Section 3. This situation has been discussed across all factors conditions (group size, number of groups and interclass correlation coefficients) as shown in the following table.

Table (7): Asymptomatic relative efficiency for the three levels of contaminated normal distribution (10% - 20% -30%)

Level	ICC	Number of groups	Group sizes 5			10			15		
			ARE(1)	ARE(2)	ARE(3)	ARE(1)	ARE(2)	ARE(3)	ARE(1)	ARE(2)	ARE(3)
10%	0.1	10	**68.6**	51.5	0.75	**46.3**	32.2	0.695	**44.1**	33.6	0.762
		20	**43.16**	25.45	0.59	**33.57**	22.82	0.679	**30.98**	22.76	0.74
		30	**4.54**	3.57	0.79	**3.61**	3.048	0.84	**3.399**	2.981	0.877
	0.2	10	**62.6**	35.29	0.564	**42.89**	27.93	0.65	**36.62**	26.023	0.71
		20	**13.16**	9.365	0.712	**9.91**	7.76	0.78	**9.03**	7.45	0.82
		30	**2.183**	1.807	0.828	**1.66**	1.45	0.875	**1.54**	1.386	0.902
	0.3	10	**20.15**	13.69	0.679	**13.12**	9.96	0.759	**10.89**	8.784	0.81
		20	**5.59**	4.36	0.779	**4.019**	3.36	0.837	**3.60**	3.13	0.869
		30	**1.156**	1.88	0.855	**9.07**	7.187	0.792	**8.603**	7.181	0.83
20%	0.1	10	64.29	**78.79**	1.213	108.12	**120.6**	1.285	107.6	**161.7**	1.51
		20	**49.9**	17.46	0.3499	**65.63**	57.23	0.26	**82.3**	73.8	0.89
		30	**38.28**	21.36	0.558	**40.07**	22.73	0.567	**39.43**	23.83	0.604
	0.2	10	132.80	**168.29**	1.267	**213.4**	180.4	0.847	**237.47**	187.37	0.789
		20	**94.033**	65.12	0.693	**117.15**	50.79	0.43	**114.74**	53.15	0.46
		30	**12.42**	13.25	1.07	**18.55**	12.085	0.651	**113.68**	53.92	0.474
	0.3	10	**103.624**	68.36	0.659	**141.37**	59.37	0.419	**129.99**	58.82	0.453
		20	**53.23**	25.79	0.484	**41.14**	23.05	0.56	**39.70**	23.54	0.59
		30	**12.12**	7.92	0.654	**9.986**	7.077	0.709	**18.35**	12.56	0.684
30%	0.1	10	18.46	**60.37**	3.35	35.09	**122.50**	3.84	58.53	**193.34**	3.327
		20	42.617	**83.39**	1.995	107.38	**185.96**	1.182	142.86	**163.661**	1.156
		30	**152.4**	101.68	0.602	**162.92**	89.90	0.548	**166.05**	93.54	0.652
	0.2	10	20.99	**70.61**	3.53	44.56	**184.48**	18.86	86.22	**163.98**	1.98
		20	53.34	**132.02**	2.49	175.56	**229.45**	1.307	**272.34**	238.71	0.877
		30	**138.16**	47.14	0.34	**116.61**	45.285	0.387	**159.017**	46.59	0.179
	0.3	10	23.62	**67.77**	2.869	55.88	**249.81**	4.47	127.91	**253.62**	1.98
		20	62.94	**98.49**	1.565	**131.57**	90.93	0.687	**270.79**	191.756	0707
		30	**89.93**	27.629	0.307	**98.71**	26.21	0.266	**101.98**	86.78	0.863

where $ARE(1) = MSE_{REML}/MSE_{RP}$, $ARE(2) = MSE_{REML}/MSE_{JR}$ and $ARE(3) = MSE_{RP}/MSE_{JR}$

From previous table the results showed that both robust estimation methods appeared to provide better performance with small MSEs than REML, also as the level of contamination increases, the MSE of REML increases, and then the corresponding values of ARE (1) and ARE (2) increase which reflect the efficiency of both RP and JR approaches (for example, suppose the condition of number of groups = 30, group size = 5 and ICC = 0.1 we obtain the following pairs of ARE (1) and ARE (2); (4.54 , 3.57), (38.28 , 21.36) and (152.4 , 101.68) according to the three levels of contaminations 10%, 20% and 30% respectively).

Moreover, there are effects of the number of groups and of the group size. With respect to group size, the larger group size leads to closer in empirical MSEs of the estimated variances of the three methods and then lower in AREs values as appear in ARE (1) and ARE (2) (for example, suppose the case of number of groups = 10 and ICC = 0.1 we obtain the following pairs of ARE (1) and ARE (2); (68.6 , 51.5), (46.3 , 32.2) and (44.1 , 33.6) corresponding to the three conditions of the group sizes 5, 10 and 15 respectively with 10% level of contamination). The factor of number of groups has more effect with the same conclusion of ARE values than the group size (for example, suppose the condition of group size = 5 and ICC = 0.1 we obtain the following pairs of ARE (1) and ARE (2); (68.6 , 51.5), (43.16 , 25.45) and (4.54 , 3.57) according to the three conditions of the number of groups 10, 20 and 30 respectively and 10% level of contamination).

Although in comparing between RP and JR robust estimation methods in Table (7), we found that, almost all the values of ARE (3) are less than 1 which reflect the efficient performance of RP for all for the three levels of contaminated normal distribution.

In Table (8) we provide the ARE results of skewed contaminated normal distribution where the errors are contaminated by skewed normal distribution with three levels of contamination (10%, 20% and 30%) and skewnees parameter set at 10. And also this situation has been discussed across all condition factors included group size, number of groups and interclass correlation coefficients as shown in the following table.

And as Table (7), we obtain the same conclusion results in Table (8), that in comparing REML with the two robust approaches and the effects of group size and of the number of groups. However, in comparing between RP and JR robust estimation methods with respect to ARE (3) values we obtain a bit different conclusion, that at a small level of skewed contamination (10%), the values of ARE (3) are larger than 1 which reflect the efficient performance of JR fit since it provides smaller MSEs than RP fit, however the higher level of skewed contaminated normal distributions lead to better performance in RP fit especially with large data sample sizes.

Table (8): Asymptomatic relative efficiency for the three level of skewed contaminated normal (10% - 20% -30%)

Level	ICC	Number of groups	Group sizes 5 ARE(1)	ARE(2)	ARE(3)	10 ARE(1)	ARE(2)	ARE(3)	15 ARE(1)	ARE(2)	ARE(3)
10%	0.1	10	42.16	8.66	0.21	**34.97**	16.86	0.48	**29.5**	21.56	0.73
		20	9.476	**14.96**	1.58	8.2	**31.59**	3.85	6.79	**48.13**	7.084
		30	1.35	**4.82**	3.583	1.035	**4.103**	3.97	0.92	**3.63**	3.95
	0.2	10	**12.19**	9.30	0.763	9.395	**20.4**	2.18	7.786	**26.89**	3.45
		20	3.72	**11.26**	3.03	3.092	**15.3**	4.95	2.46	**15.30**	6.21
		30	0.677	**1.982**	2.927	0.469	**1.38**	2.939	0.397	**1.149**	2.894
	0.3	10	5.11	**7.782**	1.523	3.698	**12.61**	3.41	2.97	**12.627**	4.258
		20	1.76	**5.59**	3.185	1.374	**5.101**	3.713	1.035	**4.061**	3.925
		30	2.64	**1.06**	2.477	4.35	**1.79**	2.414	5.523	**2.347**	2.37
20%	0.1	10	48.81	**124.73**	2.56	**140.86**	125.68	0.892	**202.36**	120.29	0.594
		20	**67.81**	34.479	0.51	**81.17**	35.39	0.437	**70.414**	32.618	0.463
		30	**12.175**	7.305	0.60	**11.156**	6.894	0.618	**10.67**	6.76	0.634
	0.2	10	**39.97**	38.87	0.972	**67.51**	36.65	0.543	**71.03**	34.73	0.49
		20	**28.41**	15.10	0.531	**29.37**	15.095	0.514	**24.21**	13.77	0.569
		30	**6.349**	4.14	0.652	**5.66**	3.826	0.676	**5.397**	3.73	0.691
	0.3	10	**24.72**	17.35	0.702	**27.743**	15.781	0.569	**26.263**	14.814	0.564
		20	**13.4**	7.865	0.587	**13.072**	7.692	0.589	**10.75**	6.93	0.645
		30	**3.59**	2.49	0.694	**3.12**	**2.25**	0.72	**2.97**	2.18	0.73
30%	0.1	10	10.36	**282.7**	27.3	16.6	**445.6**	26.8	25.2	**489.1**	19.41
		20	18.9	**103.2**	5.46	32.84	**118.01**	3.59	47.4	**105.16**	2.23
		30	**37.28**	20.63	0.55	**75.22**	20.69	0.28	**98.62**	19.95	0.22
	0.2	10	11.498	**100.33**	8.73	20.24	**119.2**	5.89	31.88	**110.39**	3.46
		20	21.76	**42.07**	1.93	44.38	**45.32**	1.02	**69.53**	40.85	0.59
		30	**33.64**	11.84	0.35	**55.56**	11.63	0.21	**60.95**	11.23	0.185
	0.3	10	12.31	**44.64**	3.625	23.31	**47.34**	2.03	39.98	**43.39**	1.085
		20	**23.47**	21.43	0.91	**50.41**	22.37	0.444	**75.703**	20.26	0.27
		30	**24.35**	7.327	0.301	**31.18**	7.08	0.23	**30.63**	6.84	0.223

where $ARE(1) = MSE_{REML}/MSE_{RP}$, $ARE(2) = MSE_{REML}/MSE_{JR}$ and $ARE(3) = MSE_{RP}/MSE_{JR}$

Finally we present the ARE results when data corrupted by the two types of outliers in order to investigate the influence of these outliers on the parameter estimates, their standard errors and the model variance. For this situation, we corrupted the normal errors with two-types of outliers as follows: **First**; 5% of the random errors have been replaced with those drawn from the normal distribution with mean 10 and variance 15^2, **Then**; replaced random effects that belong to a specific group with another drawn from the normal distribution with mean 10 and variance 15^2.

Table (9): Asymptomatic relative when data corrupted by type one and type two outliers

Types of outliers	ICC	Number of groups	Group sizes								
			5			10			15		
			ARE(1)	ARE(2)	ARE(3)	ARE(1)	ARE(2)	ARE(3)	ARE(1)	ARE(2)	ARE(3)
Type-1	0.1	10	602.88	575.65	0.955	936.90	1012.3	1.080	1111.71	1126.96	1.014
		20	405.11	418.39	1.033	518.93	645.24	1.24	711.81	584.65	0.82
		30	176.53	174.33	0.988	229.12	213.42	0.931	318.3	327.68	1.03
	0.2	10	305.51	238.80	0.78	549.52	339.18	0.617	776.44	387.73	0.499
		20	287.92	210.54	0.731	489.04	264.59	0.541	716.06	236.98	0.331
		30	164.573	105.22	0.639	264.65	126.67	0.479	378.069	166.58	0.441
	0.3	10	192.72	118.91	0.617	378.95	142.71	0.377	510.87	159.58	0.312
		20	242.09	121.46	0.502	493.30	129.297	0.262	385.19	112.82	0.293
		30	176.32	70.82	0.402	210.59	76.47	0.363	235.88	85.61	0.363
Type-2	0.1	10	56.68	12.02	0.212	61.73	12.43	0.201	62.32	13.14	0.211
		20	71.79	33.92	0.473	587.88	29.67	0.05	612.26	29.98	0.05
		30	174.29	198.50	1.14	350.91	451.72	1.29	566.9	801.3	1.41
	0.2	10	50.88	166.93	3.29	57.44	435.46	7.58	57.48	401.6	6.99
		20	228.302	551.16	2.41	374.11	584.7	1.563	369.4	533.2	1.44
		30	118.45	128.63	1.086	205.08	257.09	1.254	316.32	476.13	1.51
	0.3	10	44.73	757.01	16.93	52.03	307.3	5.91	51.3	254.6	4.97
		20	155.24	308.90	1.99	236.06	629.14	2.665	219.39	511.33	2.331
		30	82.55	87.569	1.061	125.30	144.91	1.156	183.62	235.97	1.29

where $ARE(1) = MSE_{REML}/MSE_{RP}$, $ARE(2) = MSE_{REML}/MSE_{JR}$ and $ARE(3) = MSE_{RP}/MSE_{JR}$

From the results obtained in Table (9), there is a general comment that the traditional REML estimation method is very sensitive to outliers especially in small sample size data. So for both types of corruption, the robust estimation methods (RP and JR) are more efficient than REML. In comparing between the two approaches of robust estimation methods we concluded that, for type one outliers except 5 of 27 conditions, RP appears to be more efficient than JR fit and vice versa for type two outliers since ARE (3) results are larger than 1 which reflect the high performance of JR fit.

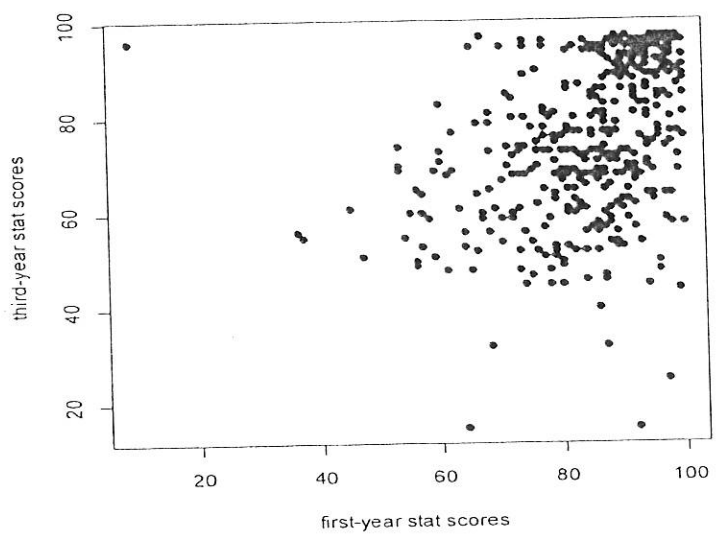
From the previous empirical analysis we can conclude that, the classical REML estimation method is not reliable for non-normal distribution of errors or data corrupted by outliers especially when the data has a small sample size. In such cases robust rank-based estimation methods would be preferred. In the next section we provide a small descriptive analysis using a real data from Helwan University.

5. Application

Consider the following real data that consists of 433 students in 14 sections in the faculty of Social Work, Helwan University. We consider two measurements occasions: the first is when the students were in the first year (2015) of their undergraduate study, and two years later in the third year (2017). We use the score in the statistics module administered on these two occasions together as well as the student's gender.

Figure 1 is a scatterplot of the third year statistic test score by the first year test score.

Figure 1 Scatterplot of 3rd year by 1st year statistic test scores



From Figure (1), there is no distinguishing between the sections to which students belong. In addition, there is a trend with increasing first year statistics test scores associated with increasing third year statistics test scores.

In Figure (2) we repeat the previous scatterplot for the different fourteen sections. The plot shows that the slopes are not the same since sections 1, 8, 12, 13 and 14 show a steeper slope compared to other sections. So the way we are going to deal with this data is to add a random section's effect. This allows us to resolve the non-independence between student's scores in the same section.

Figure 2 Scatterplot of 3rd year by 1st year statistic test scores for fourteen sections

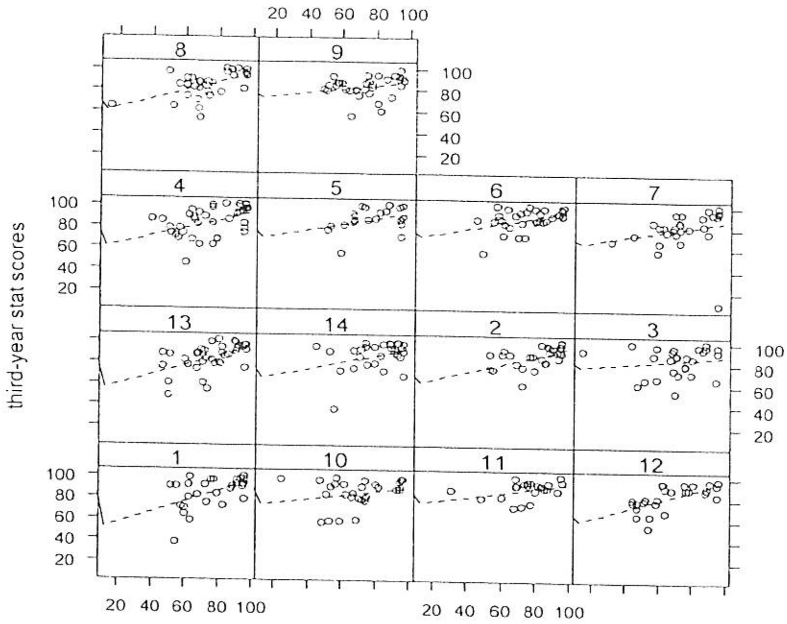


Figure 3 boxplots of 3rd year statistic scores across fourteen sections

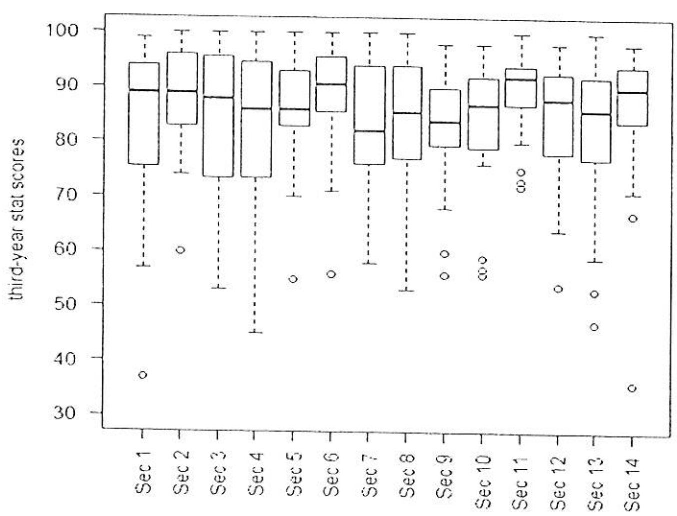


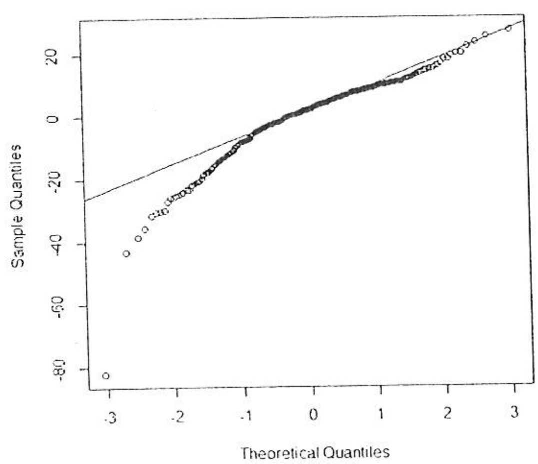
Figure (3) provided a visual depiction of how this looks like using boxplots, we immediately see that on the average the pupil's test scores on their third year are greater than or equal to eighty and sections 6, 11 and 14 appear to have the highest average of scores. But on top of that, within sections, we also see lots of individual scores variation, where some students having relatively lower scores for their section score average which reflect the presence of outliers in the data.

Now we can model these individual differences by assuming different random intercepts for each section using the following random intercept regression model:

$$third\ level.scores_{ij} = \alpha + \beta_1 first\ level.scores_{ij} + \beta_2\ gender_{ij} + \delta_{0j} + \varepsilon_{ij} \quad , \ j = 1,2,\dots,14 \ , \ i = 1,2,\dots,n_j \qquad [21]$$

where j refer to section and n_j represent the number of statistic test scores in section j. To determine if data differ from the normal distribution or not we use shapiro–wilk test where the null hypothesis of this test is that the data is normally distributed. The value of test statistic is 0.898 with p-value less than nominal level α (set $\alpha = 0.05$), then the null hypothesis is rejected and there is evidence that the data tested are not from a normally distributed population. Additional in investigating the normality we use Quantile-Quantile (Q-Q) plot, which represents a graphical method for comparing two probability distributions (sample distribution and theoretical distribution) by plotting their residual quantiles against each other. The Q-Q plot of the data is given in Figure (4).

Figure 4 Normal Q-Q plot.



The above Q-Q plot shows that most of the data points are on or near the straight reference line but there are still some points at the bottom deviate from the reference line. Then, we can conclude that there is obvious violation of the normality assumption of the error distributions.

Table (10): REML, GR and JR estimates and their SEs

Method	$\hat{\beta}_1$	$SE\ (\hat{\beta}_1)$	$Var\ (\hat{\sigma}^2)$
REML	0.30	0.033	122.83
RP	0.31	0.028	79.53
JR	0.30	0.027	64.51

To examine the parameters involved in this model, Table (10) displays the estimated of fixed and random effects and their standard errors for the REML, RP and JR analyses, where $\hat{\sigma}^2 = \hat{\sigma}_0^2(\text{between sections}) + \hat{\sigma}_e^2(\text{within sections})$. The results show that non-normal residuals have little or no effects on the parameter estimates that they are almost the same. However, for the random part in the model, there are major differences between the REML and RP and JR robust estimation methods in the results of the variance of the estimated variance in the model. The maximum value of variance is obtained by using REML which reflect the sensitive of this classical method to the outliers included in the data which appeared in Figure (3) and also affected by the violation in the normality assumption of the errors, although the robust methods (RP and JR) do perform better than REML and the minimum variance obtained by using JR fit.

Now we can say that, even little contamination (as violation in normality assumption or data contains outliers) can drive REML estimates far away from what they would be without the contamination. In these cases robust estimation methods would be preferable since they are less sensitive to outliers, violation assumptions or any other contamination.

6. Conclusion

In this paper we extend the simulation study presented by Auda et al. (2018) to involve another robust approach RP method that we compared the traditional restricted maximum likelihood method REML, robust joint ranking method JR and robust parametric method RP. For the fixed part in the model and according to empirical average bias results the study showed that non-normal situations of error distributions have almost no effects on the regression coefficients, that the estimates of regression coefficients are unbiased for the three methods.

However, for the random part in the model and according to asymptotic relative efficiency results among the three methods, the empirical validity and efficiency for the fixed effects for REML is reported to be poorer than the other two robust methods which are less sensitive to outliers and protected from the violation in normality assumption more than REML. However, in comparing between the robust estimation methods we concluded that in the cases of generated errors from skewed contaminated normal distribution and corrupted the vector of random effects by outliers, the JR method provide better efficient performance than RP fit. However, when errors generated from contaminated normal distribution the RP method would be preferred, also the large sample sizes especially in the number of groups leads to much more accurate results.

We also illustrated the robustness of the RP and JR procedures on practical using a real data set from the faculty of social work's math-achievement Helwan University that contained several outliers and violate normality assumption. The robust procedures were much less sensitive to the effect of the outliers than REML and the minimum variance is obtained by using JR fit.

Reference

1) Anderson, R. L. and Bancroft, T. A. (1952), *Statistical Theory in Research*, McGraw-Hill, New York.

2) Auda, H. A., Mckean, J., Kloke, J. and Sadek, M. M. (2018), Monte Carlo Study of REML and Robust Rank-Based Analyses for the Random Intercept Mixed Model, *Communications in Statistics - Simulation and Computation*, Accepted.

3) Bates, D. M., Machler, M., Bolker, B. and Walker, S. (2015), Fitting Linear Mixed-Effects Models Using lme4, *Journal of Statistical Software*, 67, 1-48.

4) Bilgic, T. K., McKean, J.W., Kloke, J., and Abebe, A. (2013), Iteratively reweighted generalized rank-based method in mixed models, *Manuscript*, 73-74.

5) Bilgic, Y. and Susmann, H. (2014), rlme: An R Package for Rank-Based Estimation and Prediction in Random Effects Nested Models, *The R Journal*, 5, 71-79.

6) Goldstein, H. (1989), Restricted Unbiased Iterative Generalized Least-Square Estimation, *Biometrika*, 76, 662-663.

7) Goldstein, H. (2010). *Multilevel Statistical Models. 4th Edition* . London : Institute of Education.

8) Henderson, C.R. (1950), Estimation of genetic parameters (abstract), *Ann Math*. 21, 309-310.

9) Hettmansperger, T. P. and Mckean, J. W. (1978), A Robust Analysis of the General Linear Models Based on One Step R-Estimate , *Biometrika*, 65, 571-579 .

10) Hettmansperger, T. P. and Mckean, J. W. (2011), Robust Nonparametric Statistical Methods, *Chapman & Hall*, 71-75.

11) Hox, J.J. (2002), *Multilevel Analysis: Techniques and Application*. London.

12) Kish, L. (1965). *Survey sampling*. New York, NY, John Wiley and Sons.

13) Kloke, J. D., Mckean, J. W. and Rashid, M. M. (2009), Rank-Based Estimation and Associated Inference for Linear Models with Cluster Correlated Errors , *Journal of the American Statistical Associated*, 104, 383-390.

14) Koller, M. (2015), robustlmm: An R Package for Robust Estimation of Linear Mixed-Effects Models, *Journal of Statistical Software*, 75, 1-24.

15) Koller, M. (2016), robustlmm: An R Package for Robust Estimation of Linear Mixed-Effects Models. *Journal of Statistical Software*, 75, 1-24.

16) Koul, H. L., Sievers, G. L., and McKean, J. W. (1987), An estimator of the scale parameter for the rank analysis of linear models under general score functions, *Scandinavian Journal of Statistics*, 14, 131-141.

17) Kreft, L. G. G. (1993), Using multilevel analyses to assess school effectiveness: A study of Dutch secondary schools. *Sociology of Education*, 66, 104-129.

18) Lin, C. Y. and Allister, A. J. (1984), Monte Carlo comparisons of four methods for estimation of genetic parameters in the univariate case, *J. Dairy Sci*, 67, 2389-2398.

19) Lindquist, M. A., Spicer, J., Asllani, I. and Wager, T. D. (2012), Estimating and Testing Variance Components in a Multilevel GlM , *NeuroImage*, 59, 490-501 .

20) Mass, J. M. C. and Hox, J. J. (2003), The Influence of Violations of Assumptions on Multilevel Parameter Estimates and Their Standard Errors , *Computational statistics & Data analysis*, 46, 427–440 .

21) Mckean, J. W. (2004), Robust Analysis of Linear Models, *Statistical Science*, 19, 562-570.

22) Mckean, J. W. and Kloke, J. D. (2009), Computational Rank-Based Statistics , *John Wiley & Sons*, 1, 132-140.

23) Mckean, J. w. and Kloke, J. D. (2014). Efficient and Adaptive Rank- Based Fits for Linear Models with Skew-Normal Errors. *Journal of Statistical Distributions and Applications*, 1-18.

24) Patterson, H. D. and Thompson, W. A. (1971), Recovery of Inter-Block Information When Block Sizes are Unequal, *Biometrica*, 58, 545-554.

25) Perry, A. (1908). *The management of a city school.* New York: Macmillan.

26) Purkey, S., and Smith, M. (1983). Effective schools: a review. *The Elementary School Journal*, 83, 427-452.

27) Rao, C. R. (1971). Estimation of variance and covariance components – MINQUE theory. *J. Multivariate Anal*, 1, 257-275.

28) Searle, S. R. Casella, G. and McCulloch, C. E. (1992), *Variance Components*, JohnWiley and Sons, New York.

29) Shapiro, S. S. and Wilk, M. B. (1965), An analysis of variance test for normality (complete samples), *Biometrika*, 52, 591-611.

30) Snijders, T. A. B. and Bosker, R. J. (1999), *An Introduction to basic and advanced multilevel modeling.* London.

31) Snijders, T. A. and Bosker, R. J. (2002), *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*, London: New Delhi.

32) Van der Leeden. R., Vrijburg. K. and De Leeuw. J. (1996). A review of two different approaches for the analysis of growth data using longitudinal mixed linear models: Comparing hierarchical linear regression (ML3, HLM) and repeated measures designs with structured covariance matrices (BMDP5V). *Computational Statistics and Data Analysis*, 21, 583-605.