Punjab University Journal of Mathematics
(ISSN 1016-2526)
Vol. 52(9)(2020) pp. 47-63

Selecting and Estimating Rank Score Functions Based on Residuals for Linear Mixed Models

Sehar Saleem
Department of Statistics,
University of Punjab, Pakistan,
Email: seharsaleem87@gmail.com

Rehan Ahmad Khan Sherwani
Department of Statistics,
University of Punjab, Pakistan,
Email: rehan.stat@pu.edu.pk

Received: 29 May, 2020 / Accepted: 21 August, 2020 / Published online:07 September, 2020

Abstract.: The rank-based method is a well-known robust estimation technique in analyzing linear models, it serves as an alternative to Restricted Maximum Likelihood Estimation (REML) for non-normal error distribution. It is based on minimizing a pseudo-norm and can be upgraded by selecting a suitable score function according to the probability distribution of the error term. Some generic score functions are recommended in the literature for specific shapes of the error distributions in linear models. In this study, the efficiency of score functions is examined through simulations for various level-1 and level-2 sample sizes applied on a random intercept multilevel model for symmetric, asymmetric, and light-tailed to heavy-tailed error distributions. Score functions like wscores, nscores, Bentscores1, and Bentscores4 show minimum SE only when the level-2 sample size is 10 or more. Bentscores1 and Bentscores3 are more suitable than other score functions even for the smallest sample size and their magnitudes reduce as sample size increases for right-skewed and left-skewed error distributions, respectively. Another selection criterion based on Hogg type adaptive scheme is also applied for the same class of error distribution. The efficiency rank-based fit with the selected score function is compared with the Wilcoxon score based on minimum standard error (SE).For the case of right-skewed, moderately heavy-tailed and light-tailed distribution, selected fit from the adaptive scheme is more precise than Wilcoxon fit. For contaminated normal distribution selected fit is more precise in small sample sizes only. In group size 30 or more, the selection of score function does not make a significant change in SE.

AMS (MOS) Subject Classification Codes: 35S29; 40S70; 25U09
Key Words: Adaptive Scheme, Linear Model, Rank Based Method, Score Function, Skewness, Tail Heaviness.

1. INTRODUCTION

Traditional statistical procedures based on the least squares fitting are widely used. These can easily be impaired due to the presence of outliers, e.g., a single contaminated value can spoil the validity of the methodology based on the least squares fit [1]. This issue provokes the need for robust estimation methods that are less sensitive to outliers. Wilcoxon proposed nonparametric methods for simple location problems comprising test statistics based on the ranks of data [2]. These test statistics are distribution-free and efficient like the traditional methods when the error term follows a normal distribution or even non-normal distribution [3]. Nonparametric methods are robust and powerful to provide a unified methodology in parallel with traditional methods [1]. Restricted Maximum Likelihood (REML) is a traditional method to estimate multilevel models but produces biased estimates, in case of non-normality of error terms [4]-[5]. One of the most commonly used robust methods is a rank-based method developed for linear models and provides an attractive alternative to ordinary least squares (OLS) and maximum likelihood estimation techniques (ML). Kloke et al. extended this rank theory for mixed models and developed the asymptotic theory of under the assumption that marginal distributions of residuals are the same [6]. This theory only requires the assumption that the distribution of the error term is continuous and the random errors have finite Fisher Information [1]. Additional methodologies like inferences, testing of hypotheses, diagnostic checks are developed by and summarized in [7]-[8]. Rank-based fit replaces raw scores of the dependent variable with their ranks based on non-decreasing score function [9]. It is generally highly efficient but the optimal selection of score function leads to more powerful and efficient analysis [10]. Ahmad et al. defined nine methods to define new distributions for modeling of heavy-right tailed data to obtain estimates of parameters of a special sub-model by maximum likelihood [11]. This efficiency can be boosted up according to the information available about the distribution of error term, e.g., the shape of the underlying probability density function (pdf) f is known. For example, if the random error term in the mixed-effect model follows Laplace (double exponential) distribution, the corresponding suitable score function in the rank-based method would be sign score leading towards the efficient analysis. In short, this optimality can be achieved when the shape of the underlying pdf f is known (often in practice is unknown). The rank-based analysis turns out to be more accurate and efficient when the selected score function is close to the form of f. If the probability density function (pdf) of the error term u is known as $f(u)$, then the expression for the optimal score function $\phi_f(u)$ can be mathematically derived. For this purpose, several computational procedures are available in the literature in the R language. Kloke et al. developed an R package named Rfit. This package includes the functions for complete rank-based analysis of general linear models, including the computation of confidence interval, testing of hypothesis and many more [12]. It also includes a library of score functions developed specifically for some particular distribution shapes of the error term, for example, normal, asymmetric, heavy-tailed and light-tailed distributions. In this study, the efficiency of these score functions is examined in the context of the random intercept multilevel model with the help of Monte Carlo simulations. The estimation of fixed effect, random effect parameters and their standard error (SE) for different level-1 and level-2 sample sizes are discussed. Hogg et al. presented another adaptive scheme for the two-sample location problem in linear models [13]. Kloke et al. modified Hogg's adaptive scheme and developed an R function named adaptor to select a score function from a class of suitable score functions according to the shape of the error distribution [14]. A simulation study is performed to compare the efficiency of the selected score from the adaptor function with the Wilcoxon score function on the multilevel model over a wide range of level-1 and level-2 sample sizes. The error term in the model is generated from normal, asymmetric, from heavy-tailed to light-tailed distributions. A value of precision is calculated to compare the variances of fixed effects obtained from the rank-based fit with selected scores and the Wilcoxon score. The results are discussed in the context of level-1 and level-2 sample size because in multilevel models level-2 sample size is important for consideration.

2. METHODOLOGY

2.1. General Linear Model. Al-Shomrani presented the linear regression model as shown [10]:

$$Y = \alpha_1 + X\beta + \epsilon$$

where Y is the dependent variable, X is the design matrix, β is the vector of slope parameters, α_1 is intercept and ϵ is an $n \times 1$ vector of errors that are $i.i.d.$ The density function of the error term is denoted as f and the distribution function is denoted as F. The design matrix is assumed to have a full column rank and centered as the model includes the intercept. Rank-based estimates are invariant to the intercept.

2.2. Random Intercept Model. The general form of the random intercept multilevel model in combined form is presented in Eq (2) [15]:

$$y_{ij} = \gamma_{00} + \beta_1 x_{ij} + \epsilon_{ij} + u_{0j}$$

where y_{ij} is a dependent variable measured at level-1 (i subscript refers to individual-level variation and j refers to group-level variation), x_{ij} is an explanatory variable measured at individual level, β_1 is the fixed slope parameter for each group. The term ϵ is a random error obtained from the level-1 regression equation.γ_{00} the overall intercept is a fixed effect, invariant over the clusters, i.e., it shows a common component across clusters. Also is a random error that shows the deviation of group-level slopes from the overall slope. It is assumed to be normally and independently distributed across individuals with density function f, and distribution function F.

Kloke et al. extended the rank-based fit for linear models with cluster-correlated errors [6]. Hence it is valid to apply this robust fit on multilevel models as it generates cluster-correlated errors due to the hierarchical structure in data. Rank-based fit is quite similar to the least squares method with the only difference that the Euclidean norm in the least squares is replaced with another pseudo norm (distance function) known as Jaeckel's dispersion function; see [1] for details. For the model in Eq (2) the robust estimate of β_ϕ can be defined as [6]:

$$\beta_\phi = argmin||Y - X\beta||_\phi$$

This joint rank-based estimator of β is called Jaeckel's dispersion function [16]. This is a convex function of β as it is defined in terms of the norm. It is efficient and robust in Y-space [12]. Al-Shomrani defined this pseudo norm by [10]:

$$||Y - X\beta||_\phi = \sum_{i=1}^{N}[R(y_{ij} - x'_{ki}\beta)](y_{ij} - x'_{ki}\beta)$$

where R denotes the rank of v_t among $v_1, ..., v_n$ and these are invariant to constant shifts.
y_{ij} = Dependent variable for the individual i in a cluster j
x'_{ki} = Corresponding px1 vector of covariates
For any given score function, Kloke et al. developed the asymptotic theory of rank-based method along with consistent estimators of SE and test statistics of the general linear hypothesis [6]. This asymptotic theory makes an additional assumption that the marginal distribution of error terms is the same. The simple mixed model with a compound symmetry covariance structure usually satisfies this assumption [14]. The score function is a non-decreasing square-integrable function bounded in the interval $(0, 1)$,standardizing the score function such that $\int \phi(u)du = 0$ and $\int \phi^2(u)du = 0$. Scores are calculated as $a[t] = \phi[\frac{t}{N+1}]$ and these scores sum equals zero. Some score functions satisfy assumptions about the distribution of the raw scores and few of them do not satisfy any assumptions such as the Wilcoxon score function. If the underlying distribution of the error term is known, Hajek and Sidak showed that optimal scores can be found as [17]:

$$\phi_f(u) = \phi(u) = \frac{-f'(F^{-1}(u))}{f(F^{-1}(u))}$$

where $f(u)$ and $F(u)$ are pdf and cumulative density function (cdf) of the error distribution respectively.

2.3. **The R Packages.** The package Rfit is used for computing rank-based procedures for simple linear models [12]. The use of different score functions, inferences, and diagnostic measures is provided by several functions available in the package. The authors also included some commonly used score functions in an object of class 'scores' in Rfit. Kloke $\&$ McKean, later on, developed another R package jrfit for analyzing linear models with cluster-correlated error terms [8]. The package includes the estimation of fixed effect and random effect parameters including the covariance structure of linear models. In this study, the rank-based fit is undertaken by jrfit and score functions from package Rfit are applied. Table-1 describes all the score functions provided in package Rfit by Kloke $\&$ McKean [12].

Table-1 shows the score functions available in package Rfit along with their keywords, the shape of the data distribution for which they are optimal. The probability distributions from which error terms have been generated in the simulation study are mentioned in col. 4. The following are expressions for score functions mentioned in Table-1 [18]:

Score Function	R Key-Word	Recommended Data Shape	Data Generating Distribution
Wilcoxon Score	wscores	Symmetric Moderate tailed	Logistic (2,1)
Normal Score	nscores	Symmetric Light-moderate tailed	Normal (0,1)
Bentscores1	Bentscores1	Highly right-skewed	Exponential (1.5)
Bentscores2	Bentscores2	Symmetric Light tailed	Uniform (0,1)
Bentscores3	Bentscores3	Highly left-skewed	Skewed con normal (mu=0,sd=1,alpha=-20)
Bentscores4	Bentscores4	Symmetric Moderately heavy-tailed	Slash (0,1)

TABLE 1. Available score functions in package Rfit with recommended usage, and respective data generating distributions

Wilcoxon score: $\phi(u) = \sqrt{12}[u - \frac{1}{2}]$. It is a linear function of ranks. The function $\phi(u)$ is called a Wilcoxon score function, and $a(i)$ are called the corresponding Wilcoxon scores.Normal scores: $\phi_{ns}(u) = \Phi^{-1}(u)$; where $\Phi^{-1}(u)$ is the inverse cdf of normal distribution.
Bentscores1-Bentscores4 are expressed as below:

$$\phi_1(u) = \begin{bmatrix} s_3 & u > s_1 \\ s_3 + \frac{s_3 - s_2}{s_1}(u - s_1) & otherwise \end{bmatrix}$$

$$\phi_2(u) = \begin{bmatrix} \frac{-s_3}{s_1} & u < s_1 \\ \frac{-s_4}{s_2 - 1}(u - 1) + s_4 & u > s_2 \\ 0 & otherwise \end{bmatrix}$$

$$\phi_3(u) = \begin{bmatrix} s_2 & u < s_1 \\ s_3 + \frac{s_2 - s_3}{s_1 - 1}(u - 1) & otherwise \end{bmatrix}$$

$$\phi_4(u) = \begin{bmatrix} s_2 & u < s_1 \\ s_4 & u > s_2 \\ s_3 + \frac{s_4 - s_3}{s_2 - s_1}(u - s_1) & otherwise \end{bmatrix}$$

where s_1, s_2, s_3, s_4 are the parameters.

3. RESULTS AND DISCUSSION

The selection of optimal score function is discussed via two simulation studies pursuing two different approaches developed for linear models but applied on multilevel models to examine the efficacy of these methods for cluster-correlated data also compared with traditional method REML. In the first study, the rank-based fit is obtained by all the available score functions in Rfit applied on different error distributions. The suitable score function would lead to efficient analysis. The second simulation is conducted to find out the optimal score function from a broad class of scores by using an R function adaptor based on the shape of the error distribution.

3.1. **Simulation Study 1.** The first goal is to assess the appropriateness of score functions developed for linear models (Eq 1) mentioned in Table-1 according to their recommended usage when applied on multilevel models. Data is generated for the multilevel model with cluster correlated errors (Eq 2) using a block design including a covariate and a treatment effect [8]. The sample size at level-1 (n_1) contains levels 5, 10, 30 and 50 within each level-2 sample size (n_2) 5, 10, 30, and 50. Thus it comprises 16 combinations, ranging from 25 to 2500; a small size to quite a large sample size. All the coefficients of fixed effects are chosen as zero.

Steps of Simulation Study 1

- Simulate independent variable X by a standard normal distribution, overall treatment effect $\Delta = 0.5$ and a baseline covariate with a normal distribution.
- The blocking effect and random error are simulated from the desired distribution (mentioned in Table 1).
- Estimate model (Eq 2) by R function jrfit with score function from Table 1 (repeat with each of the six scores). The same model is also estimated with REML using R function lmer.
- Obtain the fixed effect estimate of and its SE from REML and by each rank-based fit.
- Repeat step 1-4, n =1000 runs with the help of nested "for" loops one for each level-1 sample size within each level-2 sample size.

3.2. **Results of Simulation Study-1.** Estimates and SE of the fixed effect (the slope coefficient) are obtained by using appropriate score function and compared with estimates found from other score functions and REML for each grouping of level-2 within level-1 sample size.Since data is generated according to some particular probability distribution, SE against score function suitable for that specific probability distribution is expected to be minimum among others. Wilcoxon score (wscores) is appropriate for moderate tailed data. Random errors are generated with logistic distribution with parameters (location=0, scale=1) to have moderate (little heavier than normal) tails. The rank-based estimation is applied through all score functions on this data; SE for wscores is expected to be minimal because the shape of the error distribution supports its appropriate use.

Figure-1 shows a graph plotting the SE of measured, for four group units each containing four level-1 units. The SE for wscores appeared minimum for group size 30 and 50. For group size 10, SE for bentscores1 appeared least but SE for wscores is very close to that. If the score function chosen is close to the optimal score function then it would lead to efficient analysis [8]. For group size 5, SE for bentscores1 and bentscores3 appeared minimum. This might be due to the small sample size, the shape of simulated data might not match with its recommended usage.For normal scores, error terms are generated from a standard normal distribution (mean=0, sd=1).

Figure 2 depicts this pattern, all the SE values are very close to each other in larger sample sizes in block 4. The SE obtained from nscores does not come out minimum but very close to a minimal value. Wilcoxon scores appeared minimum in block 3 and block 4 because Wilcoxon and normal score functions are appropriate for almost the same shape of the distribution. Wscores are suitable for moderate tailed and nscores are suitable for light-moderate tailed distribution.Regarding bentscores1, random error from level-1 and level-2

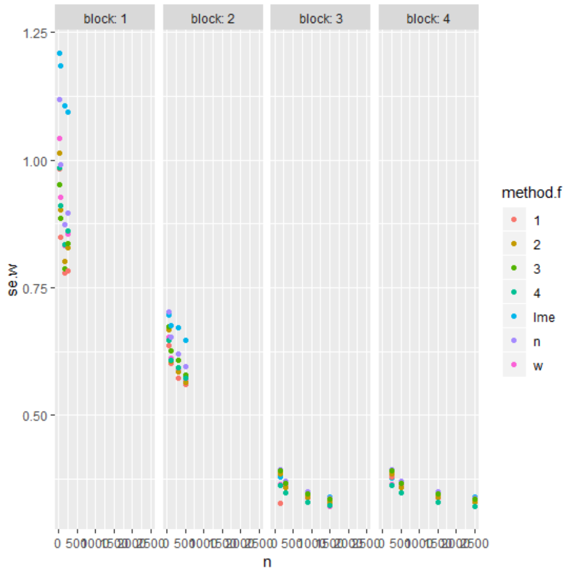

FIGURE 1. Error terms are generated by logistic distribution with parameters (2, 1)

are generated from an exponential distribution with parameter lambda=1.5, showing the highly right-skewed form of data. As expected, SE for bentscores1 must be minimum in all sample sizes ranging from smallest to largest.

Figure 3 showed a big difference in the SE obtained through REML fit as it produces high value in case of skewed data. Over a wide range of all sample sizes, SE for bentscores1 appeared minimum as expected. Another remarkable point here is the effect of sample size. SE being minimum among all other scores, still is reduced a lot from 0.259 (smallest sample size) to 0.059 (largest sample size). For bentscores2, error terms are generated from continuous Uniform distribution with parameters (0, 1) to form light-tailed symmetric data.

Figure 4 depicts minimum SE for bentscores2 for all sample sizes in group 2, group 3 and group 4. In group 1, only for sample sizes 25 and 50, SE for bentscores3 appeared minimum, this upset was may be due to the effect of a very small sample size. Though for remaining all sample sizes 150-2500, SE for bentscores2 appeared minimum. Also, SE reduces in magnitude with increasing sample size. Bentscores3 is appropriate for highly left-skewed data. Random errors are generated with contaminated normal error distribution with parameters (location=0, omega=1, alpha=-20) showing left skewness.

It is clear in Figure 5, SE for bentscores3 is minimum as compared to other score functions for each sample size. It also reduces significantly in magnitude from 0.294 (smallest) to 0.0794 (largest) sample size. Afterward, the random errors are generated for the symmetric moderate heavy-tailed form of data by using slash distribution with parameters (location=0, scale=1). Bentscores4 is appropriate for slash distribution with parameters (0, 1) to form a moderately heavy-tailed shape of the data. SE for bentscores4 is expected to get the minimum value.

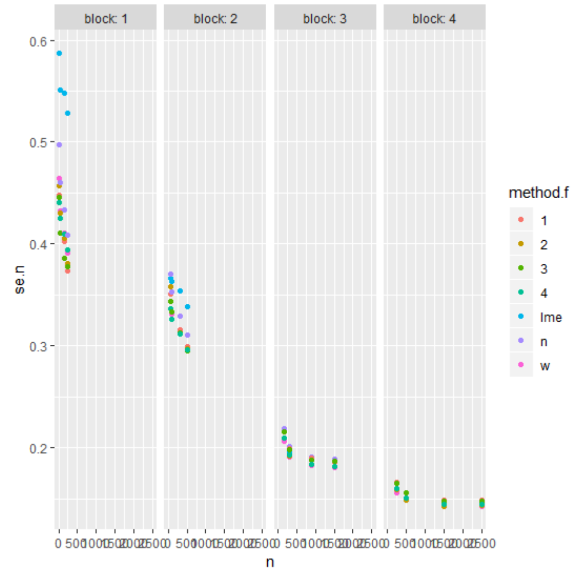

FIGURE 2. Error terms are generated by a normal distribution with parameters (0, 1)

Figure 6 clearly shows a large value of SE produced by REML for all sample sizes. The score function Bentscores4 generated minimum SE for group size 10, 30 and 50. It also reduces significantly from 4.131 (smallest sample size) to 0.400 (largest sample size). Only for group size 5, bentscores3 is smallest, might be due to the very small sample size. The rank-based method overall performed well with all score functions, as they are efficient in terms of the scale parameter comparing with traditional method REML. All the score functions produced minimum SE in large sample sizes, i.e., group size 10, 30 and 50. Within-group size 5, the result was not following literature in some instances as such a small sample size at both level-1 and level-2 is unable to depict the correct shape of the probability distribution.

3.3. **A Hogg Type Adaptive Procedure.** Different schemes in literature are developed for selecting optimal score function, for which the rank based estimates are asymptotically efficient; see, for example, [19]-[22]. This procedure is called a data-driven adaptive scheme [8]. It is similar to Hogg‘s adaptive scheme for two-sample location problems, a family of optimal score functions is selected for a class of distribution and then it selects a suitable score with the help of a data-driven ‘selector‘ [13]. Hogg et al. proposed a distribution-free two-sample adaptive test [23]. It uses a classification scheme that selects the tail weight and amount of skewness of the underlying distribution of error term, then it selects the suitable rank test based on simple linear rank statistics with corresponding scores. This scheme, in general, is developed to include heavy, moderate and light-tailed distribution for both symmetric and asymmetric cases (right-skewed and left-skewed). Initial residuals are obtained by initial rank-based fit for each cluster by using the Wilcoxon score. An adaptation

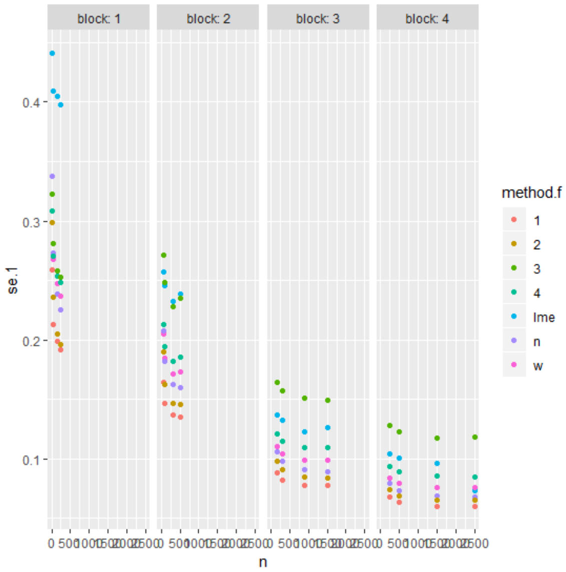

FIGURE 3. Error terms are generated by an exponential distribution with parameters (1.5)

has opted within clusters. This scheme is also modified for the family of skewed normal distributions by [24].

3.4. **Data-Driven Selector.** Hogg developed two selector statistics Q_1 and Q_2 that measure skewness and tail weight respectively [23]. These selector functions must be functions of ordered statistics of residuals. Selector statistics for skewness and tail weight respectively are presented as below [18]:

$$Q_1 = \frac{\overline{U}_{0.05} - \overline{M}_{0.5}}{\overline{M}_{0.5} - \overline{L}_{0.05}}$$

$$Q_2 = \frac{\overline{U}_{0.05} - \overline{L}_{0.05}}{\overline{U}_{0.5} - \overline{L}_{0.5}}$$

where $\overline{U}_{0.05}$, $\overline{M}_{0.5}$ and $\overline{L}_{0.05}$ are the averages of the largest 5%, middle 50% and the smallest 5% of the ordered Wilcoxon residuals respectively, and denote the skewness and tail heaviness respectively [8]. Al-Shomorani proposed benchmark values that are similar and asymptotically approach to Hogg based on sample size [10]. Although, a simple family of scores is generated by a non-decreasing piecewise continuous function defined on (0, 1) called Winsorized Wilcoxon score (bent scores) functions. Initially, regions are defined based on the values of selector statistics. For nine winsorized score functions there are nine regions defined as D_k, for $k = 1, 2, ...9$. Bent score functions numbered from 1-9 through this scheme are selected[8].

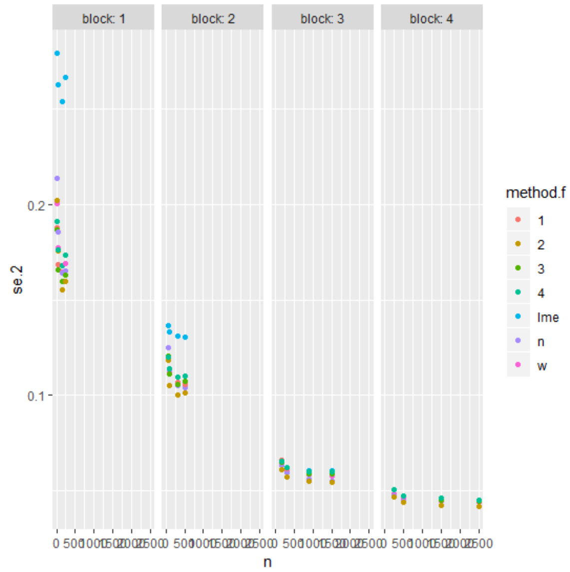

FIGURE 4. Error terms are generated by Uniform distribution with parameters (0, 1)

$Q_1 \leq c_{lq1}, Q_2 \leq c_{lq2}$ select score#1 for left skewed light tailed distribution
$Q_1 \leq c_{lq1}, c_{lq2} < Q_2 \leq c_{uq2}$ select score#2 for left skewed moderate tailed distribution
$Q_1 \leq c_{lq1}, Q_2 \leq c_{uq2}$ select score#3 for left skewed heavy tailed distribution
$c_{lq1} < Q_1 \leq c_{uq1}, Q_2 \leq c_{lq2}$ select score#4 for symmetric light tailed distribution
$c_{lq1} < Q_1 \leq c_{uq1}, c_{lq1} < Q_2 \leq c_{uq2}$ select score#5 for symmetric moderate tailed distribution
$c_{lq1} < Q_1 \leq c_{uq1}, Q_2 \geq c_{uq2}$ select score#6 for symmetric heavy tailed distribution
$Q_1 > c_{uq1}, Q_2 \leq c_{lq2}$ select score#7 for right skewed light tailed distribution
$Q_1 > c_{uq1}, c_{lq2} < Q_2 \leq c_{uq2}$ select score#8 for right skewed moderate tailed distribution
$Q_1 > c_{uq1}, Q_2 \geq c_{uq2}$ select score#8 for right skewed heavy tailed distribution
where $Q_{1l}, Q_{1u}, Q_{2l}, Q_{2u}$ are benchmarks from ordered residuals obtained from initial-fit.

3.5. **The R function Adaptor.** Kloke et al. developed an R function 'adaptor' in R package npsmReg2 to calculate this adaptive scheme using data of response variable Y and design matrix X from the linear model as input [8]. This function gives output about the selected score function, rank-based fit through this function and initial (Wilcoxon) fit. A small simulation study is conducted to investigate the performance of the adaptive scheme on the multilevel model with cluster correlated errors with each combination of level-1 and level-2 sample sizes.

3.6. **Simulation Study 2.** Another simulation study is carried out to investigate the performance of the adaptor function when it is applied to the multilevel model. The efficiency

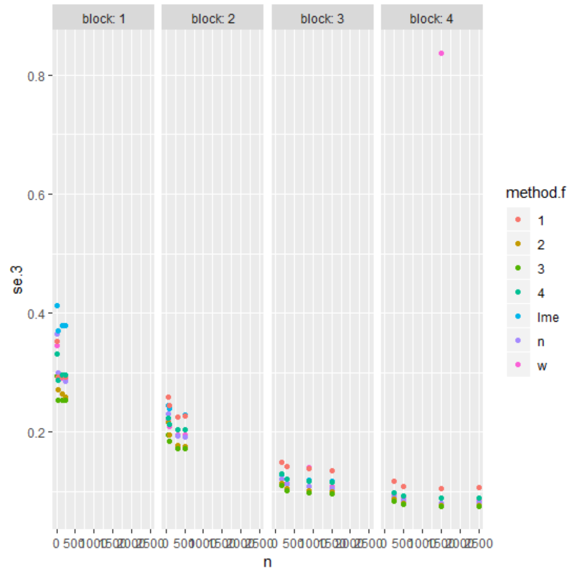

FIGURE 5. Error terms are generated by Skewed contaminated normal distribution with parameters (mu=0,sd=1,alpha=-20)

of the analysis based on the selected score is compared with a Wilcoxon fit. The multi-level model is generated from block design with the same settings from simulation study-1.
Steps of Simulation Study 2

- Simulate independent variable X by a standard normal distribution, overall treatment effect $\Delta = 0.5$ and a baseline covariate with a normal distribution.
- The block effect and error are simulated from the desired distribution (mentioned in Table 1).
- Estimate model (Eq 2) by R function adaptor (repeat with each of the six probability distributions).
- Obtain the fixed effect estimate of and its SE by each rank-based fit from the selected score and Wilcoxon fit (default).
- Precision is calculated by taking a ratio of variances of both methods for comparing the efficiency.
- Repeat step 1-4, n =10,000 runs with the help of nested "for" loops one for each level-1 sample size within each level-2 sample size.

3.7. **Results of Simulation Study 2.** The precision of the fit is calculated by taking the ratio of the squared's (variance of the selected score function to that of the Wilcoxon score function). The value of precision smaller than '1' indicates that the selected score is more precise than Wilcoxon fit in terms of SE. The value of precision for different error distributions for all level-1 sample sizes within each level-2 unit is calculated and provided in Table 2. First, both level-1 and level-2 error terms are distributed by logistic distribution with parameters (location=3, scale=6) to form symmetric moderate tailed data. The adaptor

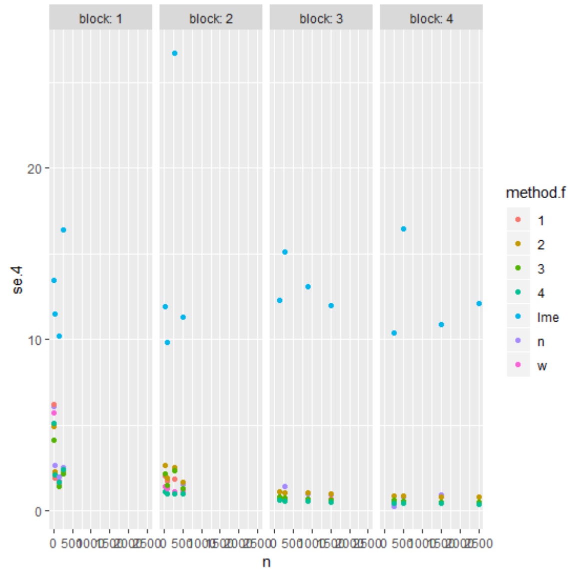

FIGURE 6. Data is generated by bentscores3 distribution with parameters (2, 1)

function selects score#5 as appropriate from the class of optimal scores because it is suitable for symmetric moderate tailed distribution. For sample size greater than 300 precision is 1, which indicates the selected score and Wilcoxon score produces the same analyses. Secondly, level-1 and level-2 error terms are generated by a standard normal distribution. Precision is less than 1 when the sample size is less than 900, implies that the selected score provides efficient analyses. In a sample size larger than 900, both fits provide similar efficiency. Random errors for level-1 and level-2 errors are generated with exponential distribution ($\lambda = 1.5$) to produce highly right-skewed data. Score #6 is selected by the adaptive scheme. For all group sizes, precision is less than 1 which shows rank-based fit from the selected score is more efficient than Wilcoxon fit. Afterward, random errors are distributed by Uniform distribution (0, 1) for light-tailed symmetric distribution. Score# 5 is selected, then the value of precision 0.90 appeared in the smallest sample size and highest 0.99 in the largest sample size. This shows score#5 is more suitable than Wilcoxon fit in small sample size. To produce a highly left-skewed form, errors are generated by a contaminated normal distribution with negative skewness. The adaptive scheme selected score# 3. The smallest value of precision is 0.88 in the smallest sample size. For the largest sample size greater than 900 precision becomes almost '1'. Next, Random errors for level-1 and level-2 are generated by slash (0, 1) distribution to form symmetric moderately heavy-tailed form. Adaptive scheme selected score function# 6 which is suitable for symmetric heavy-tailed distribution. In all combinations of level-1 and level-2 sample sizes, precision is greater than 0.80 and less than 0.96, i.e., the selected score makes the fit more precise than the Wilcoxon fit in all group sizes.

Furthermore, the appropriate selection of score functions through the adaptor function is also described when performing the rank-based analysis in the presence of outliers, through the following data example.

Level-2\Level-1	5	10	30	50
Error terms are generated by logistic distribution (moderate tailed)				
5	0.903	0.970	0.998	1.000
10	0.903	0.970	0.998	1.000
30	0.988	1.000	1.001	1.000
50	1.000	1.001	1.001	1.000
Error terms are generated by a normal distribution (symmetric light-moderate tailed)				
5	0.911	0.973	0.998	0.999
10	0.975	0.993	0.999	0.999
30	0.999	0.999	1.000	1.000
50	0.999	1.000	1.000	1.000
Error terms are generated by exponential distribution (highly right-skewed)				
5	0.802	0.833	0.857	0.863
10	0.865	0.896	0.922	0.926
30	0.911	0.919	0.928	0.929
50	0.915	0.919	0.927	0.927
Error terms are generated by uniform distribution (symmetric light-tailed)				
5	0.900	0.919	0.910	0.908
10	0.958	0.972	0.987	0.992
30	0.985	0.994	0.999	0.999
50	0.992	0.998	0.999	0.999
Error terms are generated by a skewed contaminated normal distribution (highly left-skewed)				
5	0.886	0.950	0.988	0.994
10	0.959	0.981	0.996	0.997
30	0.993	0.999	1.000	1.000
50	0.999	0.999	1.000	1.000
Error terms are generated by slash distribution (symmetric moderately heavy-tailed)				
5	0.837	0.917	0.944	0.948
10	0.854	0.885	0.910	0.906
30	0.922	0.922	0.958	0.910
50	0.929	0.919	0.909	0.904

TABLE 2. Precision calculated through an adaptive scheme for all sample sizes for several shapes of error distributions

4. DATA APPLICATION

An example of a randomized complete block design taken from [25] presents data collected on the production of vascular grafts (artificial veins). The presence of any defect called ‘flick‘ causes the rejection of graft, and these flicks might be produced by extrusion pressure. The experiment is designed to examine the influence of four different levels (treatment) of extrusion pressure on flicks in six groups of resin (blocks). The percentage of grafts in the production run without any flicks is measured as the response variable. Six complete blocks were run with four treatment levels so there are $n = 24$ observations. Block has a random effect and treatments are considered as fixed. A two-level random intercept multilevel model is considered, blocks work as level-2 units. The comparison of the rank-based estimates of fixed effects and their SEs with traditional REML analysis of the model is concerned. An initial robust analysis is shown by Wilcoxon (default) score, afterward, analysis with other score functions is undertaken as well. Table-3 shows estimates of fixed effects, their SE, t-value for REML and p-value for rank-based analysis. The first row in Table-3 contains the results for the intercept parameter. There are significant differences in both REML and Wilcoxon analysis. Four different levels (8500, 8700, 8900 and 9100) of treatment extrusion pressure were tested. For Wilcoxon analysis, treatment levels ‘8500‘ and ‘8700‘ are significant and only the last level ‘8900‘ appeared insignificant. Though in REML treatment ‘8500‘ is not significant and the rest of the levels are significant at a 5% level of significance. The SE for each main effect is magically reduced. The value of the scale parameter is $n = 4.666$. Variance component estimates through REML are $\tau_{00} = 7.3752, \sigma^2 = 7.308, \rho = 0.5022$. There are two types of robust estimators of variance components based on residuals of rank-based fit called MAD-median (mm) [12]. Mean Absolute Deviation (MAD) is used as a scale estimator which is the median absolute deviation from the median. Another estimator Hodges-Lehmann (thl) is chosen for comparison; it substitutes median with Hodges-Lehmann estimator and MAD with an estimator of the scale parameter . The detail of the estimator selected is described by [1]. The rank-based analysis includes the calculation of these two robust estimates of variance components. For MAD-median (mm) $\tau_{00} = 0.319, \sigma^2 = 0.0274, \rho = 0.9209$, and for Hodges-Lehmann (thl)$\tau_{00} = 0.322, \sigma^2 = 0.1532, \rho = 0.6775$. The MAD-median (mm) estimator shows a stronger correlation.

	REML(with outlier)			Rank Based Method (with outlier)		
Method	Estimates	SE	t-value	Estimates	SE	P-value
Constant	85.767	73.782	1.162	70.520	6.961	4.263e-09
Trt 8500	153.567	103.696	1.481	7.025	2.155	0.004
Trt 8700	5.917	103.696	0.057	5.6156	2.155	0.017
Trt 8900	3.150	103.696	0.030	2.960	2.155	0.185

TABLE 3. Summary of regression coefficients for the rank-based and REML analyses for example of vascular grafts including outlier

An outlier is responsible for much of these significant differences. In the response variable, the 21st observation is 97 and mistakenly typed as 977 that appeared as an outlier. On replacing the outlier with its actual observation, the rank-based and REML analysis become quite similar. The last level ‘9100‘ works as a reference category. Table-4 contains fixed effect estimates after correcting outlier value by REML and rank-based estimation method.

	REML(without outlier)			Rank Based Method (without outlier)		
Method	Estimates	SE	t-value	Estimates	SE	P-value
Constant	85.767	1.564	54.826	86.115	5.794	0.000
Trt 8500	6.900	1.564	4.421	6.686	1.712	0.004
Trt 8700	5.917	1.564	3.791	5.549	1.712	0.099
Trt 8900	3.150	1.564	2.018	2.970	1.745	0.116

TABLE 4. Summary of regression coefficients for the rank-based and REML analyses for example of vascular grafts data after replacing the outlier with its correct value

For rank-based (Wilcoxon) analysis, all the effects are highly significant at a 5% level of significance except 'trt8900'. Whereas all treatment levels are significant in REML. SE has also been reduced in a rank-based analysis. After the outlier removal, the REML method shows lower SE, but even for the case of the normal distribution, the rank-based method has more power than REML. The Wilcoxon analysis is radically efficient in the presence of outliers. But rank-based analysis appeared still more efficient. The scale parameter is = 5.532. Variance component estimates through REML are $\tau_{00} = 404.3527, \sigma^2 = 32258.33, \rho = 0.0224$. The 'mm' estimates are $\tau_{00} = 0.0482, \sigma^2 = 0.0074, \rho = 0.8672$, and 'thl' estimates are $\tau_{00} = 0.0509, \sigma 2 = 21.0913, \rho = 0.0024$. The MAD-median (mm) estimator performs like robust and shows the strongest correlation, while Hodges-Lehmann (thl) remains efficient.

	est.w	Se.w	Est.sc	Se.sc
Constant	71.3532	8.5869	68.7459	6.3610
Trt 8500	6.5680	2.7480	7.5000	1.9849
Trt 8700	5.3854	2.7480	5.6510	1.9849
Trt 8900	2.8623	2.7480	3.8021	1.9849

TABLE 5. Comparison of estimate and SE of default Wilcoxon score and scores selected by adaptor function

Figure-7 contains a scatter plot of studentized residuals obtained from the rank-based fit with the Wilcoxon score function and normal Q-Q plot of the response variable. After removing the outlier, a clear picture of the remaining data is obtained. The response variable seems quite good with normality. Wilcoxon studentized residuals are scattered randomly. The precision of rank-based analysis can be improved by using any other appropriate score function. A suitable score function can be obtained according to residuals obtained through the adaptor function. Residuals have light-tailed distribution shape with skewness=0.12 and kurtosis=1.95. Adaptor function selects bentscores#4, which gives the impression of proper choice here. Thus the analysis with bentscores#4 is described and compared with Wilcoxon (default) fit in Table-5. Rank-based fit with score selected from the adaptor function produces lower SE than Wilcoxon fit. Hence bentscores#4 appears as the right choice and leads to more efficient results.

5. Summary and Conclusion

The rank-based method is a robust estimation method in the presence of outliers and performs as a substitute to OLS and REML developed for linear models. This estimation method is based on a pseudo norm established on score functions. The rank-based fit could be sufficiently improved by selecting the accurate score function according to the underlying distribution of the error term. For the case when the distribution of the error term is not known, this paper investigates the performance of two selection criteria developed for linear models for a class of error distribution, i.e., symmetric, asymmetric, light-tailed to heavy-tailed distributions. We evaluated two selection schemes for random intercept multilevel models with cluster-correlated error terms. The selection of appropriate score functions is made from a class of suitable score functions. A newly developed R package Rfit for rank-based estimation includes a library of several built-in score functions. Estimates of fixed effects and their SE from each rank-based fit for 16 combinations of level-1 and level-2 sample sizes are observed through a simulation study. The efficiency of each score function is compared with other score functions by following the recommended shapes of error distributions. All the score functions performed well when group size is 30 or more and the individual sample size is 5, 10, 30 and 50. Bentscores1 and Bentscores3 show minimum SE among all other score functions even for the smallest sample size and its magnitude reduces as sample size increases. Some score functions like wscores, nscores, Bentscores1, and Bentscores4 are minimum only when the level-2 sample size is 10 or more. This implies that when the sample size is quite insufficient like group size less than 10 or say 5, the shape of error distribution would be misleading and appropriate selection of score function would be affected. Another criterion for choosing an appropriate score function is Hogg type adaptive scheme. This adaptive scheme is applied by using a recently developed R function named adaptor. A simulation study is conducted to verify the performance of the adaptor function for several shapes of distributions on the multilevel model. The rank-based fit through the selected score and Wilcoxon fit are shown and compared. For this comparison, precision is calculated by taking the ratio of estimated variances of rank-based fit from the selected score with Wilcoxon fit. For the case of highly right-skewed, moderately heavy-tailed and light-tailed distributions, selected fit by adaptor function is more precise than Wilcoxon fit in each sample size. For contaminated normal distribution selected fit is more precise even in small sample sizes say less than 30. For the rest of the situations, Wilcoxon and the selected fit produces the same efficiency. In sample size more than 900, almost for every score function, precision tends to reach 1. The results indicate the significance of sample size at each level, particularly at level-2. Generally speaking, when the total sample size is around 1000, rank-based fit through selected score function and by Wilcoxon fit produces quite similar SE of fixed effect estimates. One should be cautious when selecting a score function for small level-1 and level-2 sample sizes using either selection criteria. The application of both the selection schemes is illustrated through an example of block design with cluster-correlated errors. All the score functions provided through both selection procedures give good results in terms of lower SE and unbiased estimates for multilevel models compared with REML.

REFERENCES

[1] A. A. Al-shomrani, *A Comparison of Different Schemes for Selecting and Estimating Score Functions Based on Residuals*,, PhD. Dissertation, Western Michigan University, Michigan, 2003.

[2] Z. Ahmad, E. Mahmoudi, G. G. Hamedani, and O. Kharazmi, *New methods to define heavy-tailed distributions with applications to insurance data*,, J. Taibah Univ. Sci.**14**, No.4(2020) 359382.

[3] R. Beran, *Asymptotically Efficient Adaptive Rank Estimates in Location Model*,Ann. Stat. **2**,No.1 (1974) 6374.

[4] L. Dionne, *Efficient Nonparametric Estimators of Parameters in the General Linear Hypothesis*, Ann. Stat. **9**,No.2 (1981) 457460.

[5] C. M. Douglas, *Design and Analysis of Experiments*, Eighth Edit. John Wiley and Sons, Inc. 2012.

[6] H. Finch, *Multilevel modeling in the presence of outliers: A comparison of robust estimation methods*,, Psicologica.**38**,No.(1) (2017) 5792.

[7] H. Goldstein, *Multilevel Statistical Models*, John Wiley and Sons. **922**, 2011.

[8] J. Hajek and Z. Sidak, *Theory of Rank Tests*, Academic Press, New York. 1967

[9] T. P. Hettmansperger and J. W. McKean, *Robust nonparametric statistical methods*, Second Edi, CRC Press. 2010.

[10] R. V Hogg, *Adaptive robust procedures: A partial review and some suggestions for future applications and theory*, J. Am. Stat. Assoc. **69**,No.348 (1974) 909923.

[11] R. V Hogg, D. M. Fisher, and R. H. Randles, *A two-sample adaptive distribution-free test*, J. Am. Stat. Assoc. **70**,No.351a (1975) 656661.

[12] J. L. Hodges and E. L. Lehmann, *The efficiency of some nonparametric competitors of the t-test*, Ann. Math. Stat.**27**,No.2 (1956) 324335.

[13] L. A. Jaeckel, *Estimating Regression Coefficients by Minimizing the Dispersion of the Residuals*, Ann. Math. Stat. **43**, No.5 (1972) 14491458.

[14] J. Kapenga and J. W. McKean, *Spline estimation of the optimal score function for fitting linear models*, A.S.A. Proc. Stat. Comp. **9**,No.2 (1989) 227232.

[15] J. D. Kloke, J. W. McKean, and M. M. Rashid, *Rank-Based Estimation and Associated Inferences for Linear Models With Cluster Correlated Errors*, J. Am. Stat. Assoc. **104**, No.485(2009) 384390.

[16] J. D. Kloke and J. W. Mckean, *Nonparametric Statistical Methods Using R*, Chapman and Hall/CRC, 2014.

[17] J. D. Kloke and J. W. McKean, *Rfit: Rank-based Estimation for Linear Models*, The R Journal. **4**,No. 2 (2012) 57-64.

[18] J. D. Kloke and J. W. McKean, *Rank-Based Estimation for Arnold-Transformed Data*, Nonparametric Statistics And Mixture Models: A Festschrift in Honor of Thomas P Hettmansperger, World Scientific. (2011) 183203.

[19] J. De Leeuw, E. Meijer, and H. Goldstein, *Handbook of multilevel analysis*, Springer, (2008).

[20] R. Y. Liu and W. J. Mckean, *Robust Rank- Based and Nonparametric Methods*, Switzerland: Springer, 2015.

[21] J. W. McKean and G. L. Sievers, *Rank scores suitable for analyses of linear models under asymmetric error distributions*, Technometrics. **31**,No.2 (1989) 207218.

[22] G. A. Okyere, *Robust Adaptive Scheme for Linear Mixed Models*, Western Michigan University, Michigan. 2011.

[23] H. D. Patterson and R. Thompson, *Recovery of inter-block information when block sizes are unequal*, Biometrika. **58**,No.3 (1971) 545554.

[24] M. M. Rashid, J. W. McKean, and J. D. Kloke, *R estimates and associated inferences for mixed models with covariates in a multicenter clinical trial*, Stat. Biopharm. Res. **4**,No.1 (2012) 3749.

[25] F. Wilcoxon, *Individual Comparisons by Ranking Methods*, Biometrics Bull. **1,**No.6 (1945) 8083.