



Asian Journal of Mathematics & Statistics

ISSN 1994-5418

Efficient Wald Type Estimators for Simple Linear Measurement Error Model

¹A. Al-Radaideh, ²A. Al-Nasser and ³E. Ciavolino

¹Dipartimento di Scienze Pedagogiche, Psicologiche e Didattiche,
Decorate Course in Sciences of the Mind and Human Relations,
Università del Salento, Lecce, Italia

²Department of Statistics, Science Faculty, Yarmouk University, Irbid, Jordan

³Dipartimento di Filosofia e Scienze Sociali, Università del Salento, Lecce, Italia

Abstract: In this study, a new sampling technique called L Ranked Set Sampling (LRSS) was used to estimate the Error in Variable Model (EIV) parameter using Wald-type estimators. The new estimators were made in order to reduce the cost and increase the efficiency of the estimators. The new formulas of the Wald type estimators are shown to be unbiased towards the EIV model and are evaluated by comparing LRSS with Simple Random Sampling (SRS) and Ranked Set Sampling (RSS). Two Monte Carlo experiments were considered to study the performance of Wald-type estimators with the three sampling techniques. It appears that the suggested estimators based on LRSS are more accurate and more efficient. Moreover, a real data set of student achievements is studied.

Key words: Errors in variables, grouping methods, ranked set sampling, L ranked set sampling, simple random sampling

INTRODUCTION

The ability to understand causes and predict outcomes of events is critical in business, medicine, education, government and in fact, nearly every facet of our lives. We intuitively have the ability to detect that variables are related. The concepts of regression, however, provide us with a means to establish concrete evidence of such relationships. Simple linear regression is a powerful tool that can be used in several ways. It can be used to describe populations or to make predictions about other subjects in the population or even to test causal hypotheses. However, it is important to note that in the simple linear regression model it is assumed that the explanatory variable X is measured exactly without errors. However, in most regression problems, measurement errors affect the explanatory and the response variable (Fuller, 1987). To illustrate the effect of measurement errors on explanatory variables suppose that ξ and η are two variables having a relationship as in the following form:

$$\eta = \alpha + \beta \cdot \xi \quad (1)$$

These two variables (ξ and η) are measured with random errors, where the observed values (x, y) are given by:

Corresponding Author: Ahmed Al-Radaideh, Dipartimento di Scienze Pedagogiche, Psicologiche e Didattiche, Decorate Course in Sciences of the Mind and Human Relations, Università del Salento, Lecce, Italia

$$y = \eta + \varepsilon \tag{2}$$

$$x = \xi + \delta \tag{3}$$

The ε and δ are the random errors assumed to be mutually independent with known means and unknown variances (σ^2, τ^2), respectively. The model given in the equations is well known in the literature as Errors in Variables (EIV) model.

There are many methodologies to estimate the unknown parameters α and β for the EIV model suggested in the literatures, the easiest is the Grouping Methods (GMs) which were proposed by Wald (1940) and Bartlett (1949), also known from the name of the researchers, Wald type estimators.

Alternative EIV model estimation methods such as the least square, method of moment, maximum likelihood, instrumental variable and generalized maximum entropy, have been suggested by Madansky (1959), Kendall and Stuart (1979), Fuller (1987), Cheng and Van Ness (1999) and Al-Nasser *et al.* (2005).

In this study, we adopt the Wald-type estimators for the estimation of EIV model parameters, by combining with the new sampling technique called L Ranked Set Sampling (LRSS), proposed by Al-Nasser (2007) that can improve the efficiency of Wald type estimators.

As well as improving the efficiency, there are some practical situations that the combination of the EIV model and LRSS can help to solve.

For instance, Chen and Wang (2004), in their application showed that SRS cannot be used because of the high cost and time involved in SRS, so in these cases, they proposed RSS as a more suitable sampling scheme. Moreover, in this kind of application such as genetics, ecological and environmental studies, all the variables are affected by measurement errors so the use of EIV model will be appropriate for the estimation of the parameters.

WALD-TYPE ESTIMATORS

In order to fit a straight line when both variables are measured with error, the GMs is first proposed by Wald (1940), which is simpler than the other methods and hence or otherwise may easily be applied in many practical problems, especially in economics.

The grouping estimators, as we said, are also well known as Wald-type estimators. Suppose that we have two variables X, Y and n pairs $\{(x_i, y_i), I = 1, 2, \dots, n\}$, then the grouping methods can be described by the following steps:

- Step 1:** Order the n pairs (x_i, y_i) , by the magnitude of x_i ; for all the n observations.
- Step 2:** Select proportions P_1 and P_2 such that $P_1+P_2 \leq 1$, place the first nP_1 pairs in group one (G_1) and the last nP_2 pairs in another group (G_3) and discarding G_2 the middle group of observations; that is to say:

$$\text{Place } (x_i, y_i) \text{ in } \begin{cases} G_1 & \text{if } x_i \leq x_{p_1} \\ G_2 & \text{if } x_{p_1} < x_i \leq x_{1-p_2} \\ G_3 & \text{if } x_i > x_{1-p_2} \end{cases} \tag{4}$$

where, x_{p_i} is the p_i percentile. The slope can be estimated or formulated as follows:

$$\hat{\beta} = \frac{P_2^{-1} \sum_{G_3} Y_{np_2(i)} - P_1^{-1} \sum_{G_1} Y_{np_1(i)}}{P_2^{-1} \sum_{G_3} X_{np_2(i)} - P_1^{-1} \sum_{G_1} X_{np_1(i)}}, i = 1, \dots, n \quad (5)$$

Consequently, the intercept estimator will be:

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} \quad (6)$$

Noting that when $P_1=P_2=1/2$ then the grouping method is called two group (Wald, 1940) and when $P_1 = P_2 = 1/3$ then the method is called three group (Nair and Shrivastava, 1942; Bartlett, 1949).

L RANKED SET SAMPLING FOR BIVARIATE DATA

In the early 1950's, McIntyre (1952) without mathematical proof introduced a sampling method known as Ranked Set Sampling (RSS). This technique was introduced as an efficient alternative method to simple random sampling for estimating expected pasture yields. In McIntyre's case, measuring the plots of pasture yields requires moving and weighting crop yields which is time-consuming. However, a small number of plots can be sufficiently well ranked by eye without actual measurement. He suggested that using RSS to estimate population mean gives an unbiased estimator with smaller variance than the estimator produced by a simple random sample of the same size. Therefore, since the ranking could be done with negligible cost, he developed the RSS technique to implement this advantage.

Al-Nasser (2007) proposed a robust procedure based on L statistics, which is denoted by LRSS for detecting outliers, which a generalization for many types of RSS that introduced in the literature for estimation of the population mean. Based on the LRSS scheme, the estimator of the population mean when $r = 1$ is defined as:

$$\hat{\mu}_{LRSS} = \frac{1}{m} \left(\sum_{i=1}^k X_{i(k+1)} + \sum_{i=k+1}^{m-k} X_{i(i)} + \sum_{i=m-k+1}^m X_{i(m-k)} \right) \quad (7)$$

and its variance is given by:

$$\text{var}(\hat{\mu}_{LRSS}) = \frac{1}{m^2} \left(\sum_{i=1}^k \text{var}(X_{i(k+1)}) + \sum_{i=k+1}^{m-k} \text{var}(X_{i(i)}) + \sum_{i=m-k+1}^m \text{var}(X_{i(m-k)}) \right) \quad (8)$$

where, m is the set size, r is the number of cycles and k is the LRSS coefficient.

Al-Nasser (2007) proved that $\hat{\mu}_{LRSS}$ is an unbiased estimator of the population mean μ and has smaller variance than $\hat{\mu}_{SRS}$ if the underlying distribution is symmetric. Later, Al-Nasser and Al-Radaideh (2008) extend LRSS from an univariate sampling method to a bivariate sampling method in order to estimate the intercept and the slope for the simple linear regression model; they showed that the use of LRSS, especially with the data that have some outliers, gives more efficient estimates for the parameters. In this study, LRSS is combined with the Wald-type estimator in case the regression model considers that the error may have occurred on both dependent and independent variables.

The general idea of LRSS for choosing a bivariate sample from a population, where the response variable is too expensive to measure but the predictor variable can be measured easily, can be briefly shown in the following steps:

- Step 1:** Randomly draw m independent sets each of size m bivariate sampling units
- Step 2:** Rank the units within each sample with respect to the X 's by visual inspection or any other cheap method
- Step 3:** Select LRSS coefficient, $K = [mp]$ such that $0 = p < 0.5$ and $[z]$ is the largest integer value less than or equal to z
- Step 4:** For each of the first $(k+1)$ ranked samples, select the unit with rank $k+1$, $x_{(k+1)i}$ and select the corresponding Y , say $y_{[k+1]i}$
- Step 5:** For $j = k+2, \dots, m-k-1$, the unit with rank j in the j th ranked sample is selected
- Step 6:** With respect to the first characteristic, the procedure continues until $(m-k)$ th unit is selected from the each of the last $(k+1)$ ranked samples and the corresponding y value is measured

The steps can be repeated r times in order to choose the required sample size $n = (r \cdot m)$ and it can be noted that, for any set size when $K = 0$, the steps will lead to the novel RSS, McIntyre's procedure, also for a set size m with $K = [(n-1)/2]$, then the selected sample leads to the traditional MRSS. Moreover, PRSS can be also considered as a special case of LRSS.

WALD-TYPE ESTIMATORS BASED ON LRSS

Here, instead of using the regular SRS method to estimate the EIV model parameter, we introduce the two group and the three group estimators using the samples taken by using the LRSS method, which will be easier to use in the applications in cases where, the researcher has difficulty in measuring the study units and we also confirm that the proposed estimators based on LRSS are unbiased estimators. Now in order to fit the EIV model, let:

$$L_{(i)j}^X = \begin{cases} X_{(k+1)j} & i \leq k+1 \\ X_{(i)j} & k+2 \leq i \leq m-k-1 ; j=1,2,\dots,r \\ X_{(m-k)j} & m-k \leq i \leq m \end{cases} \quad (9)$$

be LRSS of X variable and

$$L_{[i]j}^Y = \begin{cases} Y_{[k+1]j} & i \leq k+1 \\ Y_{[i]j} & k+2 \leq i \leq m-k-1 ; j=1,2,\dots,r \\ Y_{[m-k]j} & m-k \leq i \leq m \end{cases} \quad (10)$$

are the corresponding observed values obtained from the response variable Y . Then, using Eq. 4 and 5; the two group estimates can be identified as:

$$\begin{cases} \hat{\beta}_{LRSS_{2g}} = \frac{\bar{L}_2^Y - \bar{L}_1^Y}{\bar{L}_2^X - \bar{L}_1^X} \\ \hat{\alpha}_{LRSS_{2g}} = \bar{L}_{LRSS}^Y - \hat{\beta}_{LRSS_{2g}} \bar{L}_{LRSS}^X \end{cases} \quad (11)$$

Similarly, the three group estimates can be formulated as:

$$\begin{cases} \hat{\beta}_{LRSS_{3g}} = \frac{\bar{L}_3^y - \bar{L}_1^y}{\bar{L}_3^x - \bar{L}_1^x} \\ \hat{\alpha}_{LRSS_{3g}} = \bar{L}_{LRSS}^y - \hat{\beta}_{LRSS_{3g}} \bar{L}_{LRSS}^x \end{cases} \quad (12)$$

Theorem: Assuming that model in Eq. 1-3 is satisfied, then the grouping estimators based on LRSS given in Eq. 11 are unbiased.

Proof:

$$E\left(\hat{\beta}_{LRSS_{2g}}\right) = E\left(\frac{\bar{L}_2^y - \bar{L}_1^y}{\bar{L}_2^x - \bar{L}_1^x}\right) = \left(\frac{(\alpha + \beta\bar{L}_2^x) - (\alpha + \beta\bar{L}_1^x)}{\bar{L}_2^x - \bar{L}_1^x}\right) = \beta$$

With associated variance given as:

$$\begin{aligned} \text{var}\left(\hat{\beta}_{LRSS_{2g}}\right) &= \frac{1}{\left(\bar{L}_2^x - \bar{L}_1^x\right)^2} \left\{ \text{var}\left(\bar{L}_2^y - \bar{L}_1^y\right) \right\} \\ &= \frac{\text{var}\left(\bar{L}_2^y\right) + \text{var}\left(\bar{L}_1^y\right) - 2\text{cov}\left(\bar{L}_2^y, \bar{L}_1^y\right)}{\left(\bar{L}_2^x - \bar{L}_1^x\right)^2} \end{aligned}$$

Also,

$$E\left(\hat{\alpha}_{LRSS_{2g}}\right) = E\left(\bar{L}_{LRSS}^y - \hat{\beta}_{LRSS_{2g}} \bar{L}_{LRSS}^x\right) = \left(\alpha + \beta\bar{L}_{LRSS}^x - \beta\bar{L}_{LRSS}^x\right) = \alpha$$

With

$$\begin{aligned} \text{var}\left(\hat{\alpha}_{LRSS_{2g}}\right) &= \text{var}\left(\bar{L}_{LRSS}^y - \hat{\beta}_{LRSS_{2g}} \bar{L}_{LRSS}^x\right) \\ &= \text{var}\left(\bar{L}_{LRSS}^y\right) + \left(\bar{L}_{LRSS}^x\right)^2 \text{var}\left(\hat{\beta}_{LRSS_{2g}}\right) - 2\bar{L}_{LRSS}^x \text{cov}\left(\bar{L}_{LRSS}^y, \bar{L}_{LRSS}^x\right) \end{aligned}$$

Theorem: Assuming model in Eq. 1-3 is satisfied, the LRSS based grouping estimators given in Eq. 12 are unbiased.

Proof:

$$E\left(\hat{\beta}_{LRSS_{3g}}\right) = E\left(\frac{\bar{L}_3^y - \bar{L}_1^y}{\bar{L}_3^x - \bar{L}_1^x}\right) = \left(\frac{(\alpha + \beta\bar{L}_3^x) - (\alpha + \beta\bar{L}_1^x)}{\bar{L}_3^x - \bar{L}_1^x}\right) = \beta$$

Where:

$$\begin{aligned} \text{var}(\hat{\beta}_{\text{LRSS}_{3g}}) &= \frac{1}{(\bar{L}_3^x - \bar{L}_1^x)^2} \left\{ \text{var}(\bar{L}_3^y - \bar{L}_1^y) \right\} \\ &= \frac{\text{var}(\bar{L}_3^y) + \text{var}(\bar{L}_1^y) - 2\text{cov}(\bar{L}_3^y, \bar{L}_1^y)}{(\bar{L}_3^x - \bar{L}_1^x)^2} \end{aligned}$$

However,

$$E(\hat{\alpha}_{\text{LRSS}_{3g}}) = E(\bar{L}_{\text{LRSS}}^y - \hat{\beta}_{\text{LRSS}_{3g}} \bar{L}_{\text{LRSS}}^x) = (\alpha + \beta \bar{L}_{\text{LRSS}}^x - \beta \bar{L}_{\text{LRSS}}^x) = \alpha$$

And its variance is given as:

$$\begin{aligned} \text{var}(\hat{\alpha}_{\text{LRSS}_{3g}}) &= \text{var}(\bar{L}_{\text{LRSS}}^y - \hat{\beta}_{\text{LRSS}_{3g}} \bar{L}_{\text{LRSS}}^x) \\ &= \text{var}(\bar{L}_{\text{LRSS}}^y) + (\bar{L}_{\text{LRSS}}^x)^2 \text{var}(\hat{\beta}_{\text{LRSS}_{3g}}) - 2\bar{L}_{\text{LRSS}}^x \text{cov}(\bar{L}_{\text{LRSS}}^y, \bar{L}_{\text{LRSS}}^x) \end{aligned}$$

EMPIRICAL EXAMPLE

The proposed procedure is illustrated by using a real data set from Franklin and Hariharan (1994), in which the data that contains 2600 pairs of numbers (X, Y) where, Y is the score (in percentage) obtained by a student on a standardized calculus test administered at a certain university and X is the number of hours (recorded to the nearest hour) that the students spent studying for this test. It is obvious that these two variables are subject to errors. Therefore, the EIV is applicable to such data. A simple random sample of sizes n = 60, 80, 100 and 120 are chosen from the population. For the ranked data scheme, a set of size m = 3, 4, 5 and 6, that repeated r = 20 times (cycles) were used. The results based on these samples are shown in Table 1 and 2.

Table 1: Subpopulation counts, means and standard deviations

Hours (X)	No. of items	Mean	SD
0	200	45	2.881
1	200	49	2.881
2	200	53	2.881
3	200	57	2.881
4	200	61	2.881
5	200	65	2.881
6	200	69	2.881
7	200	73	2.881
8	200	77	2.881
9	200	81	2.881
10	200	85	2.881
11	200	89	2.881
12	200	93	2.881

Table 2: The estimated slope and intercept using wald-type estimator

Methods		SRS			RSS			LRSS		
m	r	$\hat{\alpha}$	$\hat{\beta}$	MSE	$\hat{\alpha}$	$\hat{\beta}$	MSE	$\hat{\alpha}$	$\hat{\beta}$	MSE
Two-group										
3	20	74.59	-0.18	122.82	44.53	4.03	15.6	48.16	3.74	9.42
4	20	67.85	-0.11	161.47	45.18	4.09	3.86	45.73	3.94	5.26
5	20	67.61	0.40	194.12	42.09	4.34	7.93	43.55	3.91	9.72
6	20	60.18	0.33	113.44	42.69	4.24	13.7	44.96	3.88	9.64
Three-group										
3	20	76.03	-0.30	120.39	44.74	3.95	15.1	48.99	3.50	8.21
4	20	68.20	0.23	150.23	45.03	4.11	4.34	47.07	3.74	4.75
5	20	66.33	0.14	201.19	42.42	4.33	8.57	42.56	3.98	11.41
6	20	59.85	0.28	65.01	42.29	4.29	16.5	44.51	3.89	12.94

Where:

$$MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - 2} \tag{11}$$

It could be noted that, for a sample of size 60, SRS gives unreliable results and the slope values are negative, which means if the student studies his score will decrease. However, more reliable results were obtained by the ranked data schemes. This is an advantage in using LRSS and RSS over SRS, in this example. Moreover, the results showed that the estimators based on ranked data have a smaller MSE comparing to the estimators based on SRS. There is no effect of the sample size on reducing the value of MSE. Table 2 shows a good visual comparison among the three sampling methods. The results show that the ranked data are more efficient and more accurate estimators than SRS.

Table 2 comparing MSE for the three proposed method RSS and LRSS with SRS, the results show that the ranked data are more efficient and more accurate estimators than SRS.

SIMULATION STUDY

Two different Monte Carlo experiments were performed to gain insight into the properties of the EIV model parameters estimate using ranked data. The simulation was performed for sample size $n = 15, 20, 25, 30, 40$ and 50 where, the set size was selected to be $r = 5, 10$ and cycle size was $m = 3, 4, 5$ for RSS and LRSS; with coefficient ($k = 1$). Then 100,000 random samples were generated, assuming that $\alpha = 1, \beta = 2$. The error terms and ξ were generated from the standard Normal distribution. Then the bias, the mean square errors for the model parameters and for the overall model and the relative efficiency were computed for each method. The formulas used in our simulation are as follows:

$$\begin{aligned} \text{Bias}(\hat{\alpha}) &= \frac{1}{100000} \sum_{i=1}^{100000} (\alpha - \hat{\alpha}_i) & \text{Bias}(\hat{\beta}) &= \frac{1}{100000} \sum_{i=1}^{100000} (\beta - \hat{\beta}_i) \\ \text{MSE}(\hat{\alpha}) &= \frac{1}{100000} \sum_{i=1}^{100000} (\alpha - \hat{\alpha}_i)^2 & \text{MSE}(\hat{\beta}) &= \frac{1}{100000} \sum_{i=1}^{100000} (\beta - \hat{\beta}_i)^2 \\ \text{RE}(\cdot) &= \frac{\text{MSE}_{\text{SRS}}(\cdot)}{\text{MSE}_{(\cdot)}(\cdot)} \\ \text{MSE}(\text{model}) &= \frac{1}{100000} \left[\sum_{i=1}^{100000} \left(\frac{\sum_{j=1}^n \hat{\epsilon}_j^2}{(n-2)} \right) \right]; \text{where} \\ \hat{\epsilon}_i &= y_i - \hat{y}_i \text{ and } \hat{y}_i = \hat{\alpha} + \hat{\beta}x_i \end{aligned}$$

Experiment 1

To study the performance of the EIV model parameters, the bias and the mean square error for the model under the simulation assumptions, we simulate ξ_1 , δ_1 and ϵ_1 from normal (0, 1) as symmetric distribution. The simulation results, bias, MSE and relative efficiencies are shown in Table 3-5, respectively. From the result Table 3-5, it appears that the use of

Table 3: Bias for estimated EIV parameters

Methods		$\hat{\alpha}$			$\hat{\beta}$		
m	r	SRS	RSS	LRSS	SRS	RSS	LRSS
Two-groups							
3	5	-8E-06	9E-06	-5E-06	1E-05	5E-06	1E-05
3	10	-9E-06	-6E-06	-1E-05	1E-05	8E-06	2E-05
4	5	-5E-07	-3E-05	-2E-05	8E-06	9E-06	1E-06
4	10	-4E-06	-1E-06	3E-06	1E-05	9E-06	4E-06
5	5	1E-05	-2E-05	-2E-05	7E-06	2E-06	4E-07
5	10	2E-06	6E-06	9E-06	8E-06	9E-06	8E-06
Three-group							
3	5	-6E-06	4E-06	6E-06	9E-06	5E-06	8E-06
3	10	-9E-06	5E-06	-1E-05	9E-06	1E-05	1E-05
4	5	-7E-06	-4E-05	7E-06	1E-05	1E-05	-4E-07
4	10	5E-06	1E-07	-5E-06	1E-05	1E-05	1E-06
5	5	3E-06	-2E-05	-3E-05	9E-06	4E-06	9E-07
5	10	9E-06	1E-05	1E-05	1E-05	7E-06	9E-06

Table 4: MSE for estimated EIV parameters

Methods		$\hat{\alpha}$			$\hat{\beta}$		
m	r	SRS	RSS	LRSS	SRS	RSS	LRSS
Two-group							
3	5	4E-05	4E-05	4E-05	2E-06	2E-06	5E-06
3	10	2E-05	2E-05	2E-05	1E-06	1E-06	3E-06
4	5	3E-05	3E-05	3E-05	2E-06	2E-06	4E-06
4	10	1E-05	1E-05	1E-05	9E-07	8E-07	2E-06
5	5	2E-05	2E-05	2E-05	1E-06	1E-06	3E-06
5	10	1E-05	1E-05	1E-05	7E-07	7E-07	1E-06
Three-group							
3	5	5E-05	5E-05	5E-05	2E-06	2E-06	5E-06
3	10	3E-05	3E-05	3E-05	9E-07	9E-07	2E-06
4	5	4E-05	4E-05	4E-05	2E-06	1E-06	3E-06
4	10	2E-05	2E-05	2E-05	7E-07	7E-07	2E-06
5	5	3E-05	3E-05	3E-05	1E-06	1E-06	2E-06
5	10	2E-05	1E-05	1E-05	5E-07	5E-07	1E-06

Table 5: RE for EIV parameters

Methods		$\hat{\alpha}$		$\hat{\beta}$	
r	m	RSS	LRSS	RSS	LRSS
Two-group					
3	5	1.073	1.009	1.059	0.454
3	10	1.035	1.000	1.032	0.451
4	5	1.048	1.010	1.055	0.455
4	10	1.018	1.007	1.043	0.455
5	5	1.047	1.029	1.056	0.523
5	10	1.019	1.012	1.018	0.510
Three-group					
3	5	1.063	1.022	1.049	0.448
3	10	1.038	1.008	1.031	0.451
4	5	1.038	1.020	1.060	0.454
4	10	1.012	1.009	1.033	0.453
5	5	1.035	1.026	1.057	0.521
5	10	1.002	1.003	1.019	0.508

grouping methods gives unbiased estimated parameter values for all-sampling methods. The bias using all-methods are around $(.) \cdot 10^{-5}$ and $(.) \cdot 10^{-7}$. The values for the estimated parameter decrease when the sample size increases. Moreover, in the case of the estimated intercept, the LRSS scheme is more accurate than the SRS scheme. LRSS is more efficient than SRS, while in all situations RSS is more efficient than other schemes. However, in estimating the slope RSS is more efficient for the slope t than SRS and LRSS.

Experiment 2

Here, the performance of the proposed grouping estimators in the presence of the outliers is investigated under the same simulation assumptions. Then one random outlier in each case is generated according to the sample size.

The simulated bias and MSE are summarized in Table 6-8, the results indicate that the proposed estimators based on RSS and SRS when the data contain some outliers give inaccurate estimators for the EIV parameters. However, in case of LRSS the proposed estimators are more accurate and more efficient than the other schemes. MSE is also much better for LRSS. Moreover, the estimated parameters decrease as the sample size increases. The result indicates that LRSS is more efficient than RSS and SRS, also with two and three-group methods to estimate the EIV parameters. And it indicates that the RE for all methods increases when the number of cycles increases.

Table 6: Bias for the EIV parameters: outliers case

Method		$\hat{\alpha}$			$\hat{\beta}$		
m	r	SRS	RSS	LRSS	SRS	RSS	LRSS
Two-group							
3	5	-0.004	-0.001	0.001	0.507	0.491	0.210
3	10	0.001	0.003	0.000	0.503	0.496	0.209
4	5	0.000	-0.001	-0.001	0.404	0.392	0.164
4	10	0.002	0.001	0.000	0.402	0.396	0.164
5	5	-0.005	-0.001	0.000	0.336	0.327	0.140
5	10	-0.002	0.001	0.000	0.334	0.330	0.139
Three-group							
3	5	-0.004	0.000	0.002	0.530	0.513	0.210
3	10	0.001	0.006	0.000	0.528	0.520	0.210
4	5	0.000	-0.002	-0.001	0.421	0.409	0.165
4	10	0.003	0.001	-0.001	0.426	0.420	0.166
5	5	-0.006	0.000	0.000	0.358	0.349	0.142
5	10	-0.003	0.001	0.000	0.352	0.348	0.142

Table 7: MSE for the EIV parameters: outliers case

Method		$\hat{\alpha}$			$\hat{\beta}$		
m	r	SRS	RSS	LRSS	SRS	RSS	LRSS
Two-group							
3	5	0.845	0.815	0.180	0.271	0.271	0.063
3	10	0.393	0.386	0.085	0.259	0.260	0.053
4	5	0.543	0.527	0.098	0.173	0.177	0.039
4	10	0.281	0.278	0.052	0.166	0.169	0.033
5	5	0.436	0.421	0.084	0.120	0.125	0.028
5	10	0.212	0.205	0.040	0.115	0.118	0.024
Three-group							
3	5	1.676	1.503	0.339	0.298	0.297	0.066
3	10	0.903	0.816	0.186	0.286	0.287	0.055
4	5	1.118	1.024	0.194	0.189	0.193	0.041
4	10	0.678	0.623	0.119	0.187	0.190	0.034
5	5	0.995	0.900	0.182	0.138	0.143	0.030
5	10	0.466	0.426	0.085	0.128	0.131	0.025

Table 8: RE for EIV parameters: outliers case

Method		$\hat{\alpha}$		$\hat{\beta}$	
m	r	RSS	LRSS	RSS	LRSS
Two-group					
3	5	1.216	1.811	1.035	3.774
3	10	1.095	3.047	1.009	5.580
4	5	1.155	2.159	1.012	3.755
4	10	1.078	3.668	1.005	5.304
5	5	1.100	2.517	0.987	3.667
5	10	1.042	3.878	0.987	4.715
Three-group					
3	5	1.142	0.603	1.030	4.228
3	10	0.969	0.485	1.011	6.406
4	5	1.066	1.199	1.015	4.259
4	10	0.851	0.945	1.006	5.941
5	5	0.883	0.693	0.988	4.052
5	10	0.839	0.584	0.989	5.216

CONCLUSIONS

Wald-type estimators for the Error in Variable Model (EIV) were made based on the new sampling technique called L Ranked Set Sampling (LRSS). The proposed procedure was exemplified by a real data set from Franklin and Hariharan (1994), in order to reduce the cost and increase the efficiency of the estimators. The data contained bivariate data (X, Y). Where, Y was the score (in percentage) obtained by a student on a standardized calculus test administered at a certain university and X was the number of hours (recorded to the nearest hour) that the students spent studying for this test. The results of the statistical analysis recognized the robustness of the LRSS scheme over SRS by giving more reliable estimates. Moreover, it has been seen that the proposed Wald type estimators are unbiased toward the EIV model. Two Monte Carlo experiments were considered in order to study the performance of the estimators with the three sampling techniques, LRSS, RSS and SRS. It appears that the suggested estimators based on LRSS are more accurate and more efficient. The simulation confirmed that estimating both the slope and the intercept of the EIV model using LRSS is in general more efficient than using SRS. Note that the experiments show that the relative precision of the proposed estimators decreases as the set size or the cycle size increase. It seems that, for moderate or large sample size, the performance of parameter estimation of the EIV model was affected by the sampling scheme used. However, using LRSS is still more efficient than using SRS or RSS for the analysis.

REFERENCES

- Al-Nasser, A.D., M.E. Al-Haj and A. Al-Masri, 2005. On the use of l-statistics in wald-type estimators for fitting a straight line when both variables are subject to error. Pak. J. Stat., 21: 179-194.
- Al-Nasser, A., 2007. L ranked set sampling: A generalization procedure for robust visual sampling. Commun. Stat. Simul. Comput., 36: 33-43.
- Al-Nasser, A.D. and A. Al-Radaideh, 2008. Estimation of simple linear regression model using L ranked set sampling. Int. J. Open Problems Comput. Sci. Math., 1: 18-33.
- Bartlett, M.S., 1949. Fitting straight line when both variables are subject to error. Biometrics, 5: 207-212.

- Chen, Z. and Y.G. Wang, 2004. Efficient regression analysis with ranked-set sampling. *Biometrics*, 60: 997-1004.
- Cheng, C.L. and J.W. Van Ness, 1999. *Statistical Regression with Measurement Error*. Arnold, New York, USA.,
- Franklin, A.G. and K.L. Hariharan, 1994. *Regression Analysis, Concepts and Application*. Wands Worth, California, USA.
- Fuller, W.A., 1987. *Measurement Error Models*. Wiley, New York.
- Kendall, M.G. and A. Stuart, 1979. *The Advanced Theory of Statistics*. 2nd Edn., Griffin, London.
- Madansky, A., 1959. The fitting of straight lines when both variables are subject to error. *J. Am. Stat. Soc.*, 55: 173-205.
- McIntyre, G.A., 1952. A method for unbiased selective sampling, using ranked sets. *Aust. J. Agric. Res.*, 3: 385-390.
- Nair, R.K. and P.M. Shrivastava, 1942. On a simple method of curve fitting. *Sanyukhya*, 6: 121-132.
- Wald, A., 1940. The fitting of straight lines if both variables are subject to error. *Ann. Math. Stat.*, 11: 284-300.