Hacettepe Journal of Mathematics and Statistics
Volume 36 (1) (2007), 65 – 73

BEST LINEAR UNBIASED ESTIMATORS FOR THE MULTIPLE LINEAR REGRESSION MODEL USING RANKED SET SAMPLING WITH A CONCOMITANT VARIABLE

Yaprak Arzu Özdemir* and A. Alptekin Esin*

Received 27 : 01 : 2006 : Accepted 20 : 11 : 2006

Abstract

Ranked set sampling is a more efficient sampling technique than simple random sampling for estimating the population mean when the measurement of the sampling units according to the variable of interest is expensive or difficult, but ranking them is relatively cheap and easy. In this study, the best linear unbiased estimators in the class of linear combinations of the ranked set sample values are obtained for multiple linear regression models with replicated observations. During the sample selection procedure of ranked set sampling, it is assumed that the ranking is done according to a concomitant variable. The estimators obtained from the ranked set sampling and simple random sampling with the same sample size are compared with respect to the relative efficiency.

Keywords: Ranked set sampling, Best linear unbiased estimators, Optimal L-estimators, Order statistics, Multiple linear regression model, Concomitant ranking, Ranking error.

2000 AMS Classification: 62 D 05.

1. Introduction

In many environmental, ecological, agricultural and medical studies, the measurement of the variable of interest is generally expensive or difficult. Thus, in such fields, it is natural to prefer a sampling technique that represents the population to the best level possible while using the smallest sample size. A sampling technique suiting this purpose was initially suggested by McIntyre [12] under the name Ranked Set Sampling (RSS). McIntyre estimated the mean of pasture yields and indicated that RSS is a more efficient

*Department of Statistics, Faculty of Science and Arts, Gazi University, 06500 Beşevler, Ankara, Turkey.

sampling technique than Simple Random Sampling (SRS) for estimating the population mean.

In the RSS technique, the sample selection procedure is composed of two stages. In the first stage of sample selection, m random samples of size m are drawn from the population with replacement by SRS. Each of these samples is called a "set", and each set is roughly ranked with respect to the variable of interest Y from the lowest value to the highest value. Ranking of the units is done with a low-level measurement such as using previous experiences, visual measurement or using a concomitant variable. At the second stage; the first unit from the first set, the second unit from the second set, and generally the m^{th} unit from the m^{th} set are taken and measured with respect to Y at a desired level. The resulting sample $Y_{(1)}, Y_{(2)}, \ldots, Y_{(m)}$ constitute a ranked set sample with size m. Here, $Y_{(i)}$ denotes the i^{th} order statistic in the i^{th} set for $i = 1, 2, \ldots, m$ based on the assumption of no ranking error. So, $Y_{(1)}, Y_{(2)}, \ldots, Y_{(m)}$ are all independent but not identically distributed. This sample selection procedure may be repeated r times until one reaches the desired sample size n. After the r^{th} cycle, the estimator of the population mean, which is generated from a ranked set sample of size $n = mr$, is obtained as in [12],

$$\overline{Y}_{\mathrm{RSS}} = \frac{1}{mr} \sum_{i=1}^{m} \sum_{j=1}^{r} Y_{(i)j}. \tag{1}$$

Takahasi and Wakimoto [18] proved that $\overline{Y}_{\mathrm{RSS}}$ is an unbiased estimator of the population mean which is more efficient than the simple random sample mean $\overline{Y}_{\mathrm{SRS}}$. Furthermore, Dell and Clutter [7] showed that $\overline{Y}_{\mathrm{RSS}}$ is an unbiased estimator of the population mean regardless of ranking errors, and that is more efficient than $\overline{Y}_{\mathrm{SRS}}$ unless the ranking is so poor as to yield a random sample.

As is seen from (1), $\overline{Y}_{\mathrm{RSS}}$ gives equal weight to all ranked set sample units. In recent years, Stokes [17], Sinha et al.[15], Barnett and Moore [5], Barnett and Barreto [3], Barnett and Barreto [4], Gang and Al-Saleh [8] and Balakrishnan and Li [2] have considered more general linear combinations of ranked set sample units. They considered best linear unbiased estimators (BLUEs) for the class of linear combinations of ranked set sample values, namely optimal L-estimators. In these studies, generally location and scale parameters of the selected distributions were estimated. However, Barnett and Barreto [3] were the first to estimate the regression parameters by using optimal L-estimators. They considered a simple linear regression model with replicated observations when the dependent variable is normally distributed and ranking is assumed to be perfect. This model is compared with the simple linear regression model with respect to Relative Efficiency (RE). It is remarked that the optimal L-estimators of slope and intercept parameters based on RSS are more efficient than the estimators based on SRS. Besides, the residual variance is also more efficient than the estimators based on SRS except in the cases where the number of replicated observations is smaller than 6.

In this study, we extend the work of Barnett and Barreto [3] by introducing the optimal L-estimators of multiple linear regression model parameters based on the assumption of ranking by a concomitant variable. We compare this model with the traditional multiple linear regression model by means of RE. Also, we calculate the RE values relating to residual variance by using the uniformly minimum variance unbiased estimator of σ for the normal distribution, while Barnett and Barreto [3] gives an approximate variance formulation for the $\widehat{\sigma}$ under SRS.

2. BLUE's in the case of concomitant ranking

Let the independent variables $X_1, X_2, \dots, X_p$ constitute the regression model for c different predetermined values ($X_l = x_{lj},\ l = 1, 2, \dots, p;\ j = 1, 2, \dots, c$), thus the conditional mean and variance of the dependent variable Y are defined respectively by;

$$\text{(2)}\qquad \begin{aligned} \mathrm{E}(Y/X_l = x_{lj}) &= \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \cdots + \beta_p x_{pj} \\ \mathrm{Var}(Y/X_l = x_{lj}) &= \sigma^2 \mathbf{I}_{cxc}. \end{aligned}$$

To obtain optimal L-estimators for the model parameters, Y is observed more than once at each set of c distinct values of the independent variables. Suppose that Y is observed from a ranked set sample with size m for the j^{th} value set of the independent variables $(x_{1j}, x_{2j}, \dots, x_{pj})$ for $j = 1, 2, \dots, c$. To obtain the ranked set sample, the sample selection procedure mentioned in Section 1 can be applied. Using this sample selection procedure, Özdemir [13] obtained optimal L-estimators of the multiple linear regression model parameters based on the assumption of no ranking error. In applications, units are generally ranked in RSS by visual techniques, or by using a concomitant variable, which is cheap and easy. Since the units are not ranked with exact measurements of the variable of interest Y, it is possible to have a ranking error. The most important studies of ranking error were made by Dell and Clutter [7], David and Levine [6], Ridout and Cobby [14] and Stokes [16]. In these studies, Stokes [16] used the simple linear regression model to estimate the population mean in the case of ranking according to a concomitant variable. We use this approach of Stokes [16] to take into account the effect of ranking according to a concomitant variable to the multiple linear regression model.

Let the concomitant variable, which will be used for the ranking of Y, be defined as a variable W with mean μ_W and variance σ_W^2. Then, in the sample selection procedure in RSS, units are ranked according to the concomitant variable W and the selected units are measured with respect to variable Y. The order statistic $Y_{[i]}$ for the dependent variable Y which corresponds to the order statistic $W_{(i)}$ for $i = 1, 2, \dots, m$i s called the *induced order statistic*, and the conditional mean of $Y_{[i]}$ is given as follows [16];

$$\text{(3)}\qquad \mathrm{E}(Y_{[i]}/W_{(i)}) = \mu_y + \rho \frac{\sigma_y}{\sigma_W}(W_{(i)} - \mu_W),$$

where ρ is the correlation coefficient between Y and W. For the general case of a continuous random variable W with a distribution function $F[\frac{(w-\mu_W)}{\sigma_W}]$, the standardized variable $U_i = \frac{W_i - \mu_W}{\sigma_W}$, has a parameter-free (and hence completely known) distribution with $F(u; 0, 1)$ and the i^{th} order statistic for U_i is $U_{(i)} = \frac{W_{(i)} - \mu_W}{\sigma_W}$. Then, the mean and the variance of $U_{(i)}$ from such a distribution is given by $\mathrm{E}(U_{(i)}) = \eta_i$, $\mathrm{Var}(U_{(i)}) = \frac{\mathrm{Var}(W_{(i)})}{\sigma_W^2} = \tau_i$, respectively [1]. So, the mean and the variance of $Y_{[i]}$ are obtained as;

$$\text{(4)}\qquad \mathrm{E}(Y_{[i]}) = \mathrm{E}(\mathrm{E}(Y_{[i]}/W_{(i)})) = \mu_y + \rho\sigma_y\eta_i,$$

$$\text{(5)}\qquad \mathrm{Var}(Y_{[i]}) = \sigma_y^2[(1-\rho^2) + \rho^2\tau_i].$$

Let $U_{[i]} = \frac{Y_{[i]} - \mu_y}{\sigma_y}$ be the i^{th} induced order statistic of a ranked set sample from a standardized distribution $F(u; 0, 1)$. Then, the mean and the variance of $U_{[i]}$ are written using (4) and (5) as follows,

$$\text{(6)}\qquad \mathrm{E}(U_{[i]}) = \frac{\mathrm{E}(Y_{[i]}) - \mu_y}{\sigma_y} = \rho\eta_i$$

$$\text{(7)}\qquad \mathrm{Var}(U_{[i]}) = \frac{\mathrm{Var}(Y_{[i]})}{\sigma_y^2} = (1-\rho^2) + \rho^2\tau_i.$$

For estimating the model parameters in the multiple regression model given in (2), m observations from Y_{ij}, $(i = 1, 2, \ldots, m, \ j = 1, 2, \ldots, c)$ are taken using RSS at each value set of $(x_{1j}, x_{2j}, \ldots, x_{pj})$ based on the assumption of ranking Y with respect to the concomitant variable W. Then, the i^{th} induced order statistic at the j^{th} value set $(x_{1j}, x_{2j}, \ldots, x_{pj})$ obtained by ranked set sampling from the standardized distribution $F(u; 0, 1)$ takes the form;

$$U_{[i]j} = \frac{Y_{[i]j} - (\beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \cdots + \beta_p x_{pj})}{\sigma}, \tag{8}$$

where $Y_{[i]j}$ is the i^{th} induced order statistic of the variable Y for the j^{th} value set of $(x_{1j}, x_{2j}, \ldots, x_{pj})$. So, the mean and the variance of $U_{[i]j}$ are obtained from (6) and (7) as,

$$\mathrm{E}(U_{[i]j}) = \mathrm{E}(U_{[i]}) = \rho\eta_i, \tag{9}$$

$$\mathrm{Var}(U_{[i]j}) = \mathrm{Var}(U_{[i]}) = (1-\rho^2) + \rho^2\tau_i, \tag{10}$$

where the subscript j disappear since the m samples are of equal size in the replications. Using (8), (9) and (10), the mean and the variance of $Y_{[i]j}$ can be written as,

$$\mathrm{E}(Y_{[i]j}) = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \cdots + \beta_p x_{pj} + \sigma\rho\eta_i \tag{11}$$

$$\mathrm{Var}(Y_{[i]j}) = \sigma^2 \mathrm{Var}(U_{[i]j}) = \sigma^2[(1-\rho^2) + \rho^2\tau_i]. \tag{12}$$

Thus, the matrix form for this model is;

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \boldsymbol{\epsilon} \tag{13}$$

where,

$$\mathbf{Y} = \begin{bmatrix} Y_{[1]1} \\ Y_{[2]1} \\ \vdots \\ Y_{[m]1} \\ Y_{[1]2} \\ Y_{[2]2} \\ \vdots \\ Y_{[m]2} \\ \vdots \\ Y_{[1]c} \\ Y_{[2]c} \\ \vdots \\ Y_{[m]c} \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{p1} & \eta_1 \\ 1 & x_{11} & x_{21} & \cdots & x_{p1} & \eta_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & x_{11} & x_{21} & \cdots & x_{p1} & \eta_m \\ 1 & x_{12} & x_{22} & \cdots & x_{p2} & \eta_1 \\ 1 & x_{12} & x_{22} & \cdots & x_{p2} & \eta_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & x_{12} & x_{22} & \cdots & x_{p2} & \eta_m \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1c} & x_{2c} & \cdots & x_{pc} & \eta_1 \\ 1 & x_{1c} & x_{2c} & \cdots & x_{pc} & \eta_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & x_{1c} & x_{2c} & \cdots & x_{pc} & \eta_m \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \\ \sigma \end{bmatrix} \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_{[1]1} \\ \epsilon_{[2]1} \\ \vdots \\ \epsilon_{[m]1} \\ \epsilon_{[1]2} \\ \epsilon_{[2]2} \\ \vdots \\ \epsilon_{[m]2} \\ \vdots \\ \epsilon_{[1]c} \\ \epsilon_{[2]c} \\ \vdots \\ \epsilon_{[m]c} \end{bmatrix} \tag{14}$$

with ϵ a random error vector having $\mathrm{E}(\boldsymbol{\epsilon}) = 0$ and

$$\mathrm{Var}(\boldsymbol{\epsilon}) = \mathrm{Var}(\mathbf{Y}) = \sigma^2\mathbf{V}, \tag{15}$$

where $\mathbf{V} = (1-\rho^2)\mathbf{I}_{nxn} + \rho^2\mathbf{D}$ and $\mathbf{D} = diag(\tau_1, \tau_2, \ldots \tau_m, \tau_1, \tau_2, \ldots \tau_m, \ldots, \tau_1, \tau_2, \ldots \tau_m)$. So, in the case of ranking error, using the generalized least square technique based on order statistics suggested by Llyod [11], the BLUE of the parameter vector $\mathbf{B}$ is given by,

$$\widehat{\mathbf{B}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}, \tag{16}$$

with variance covariance matrix

$$\mathrm{Var}(\widehat{\mathbf{B}}) = \sigma^2(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}. \tag{17}$$

3. The relative efficiencies of the BLUEs

In this section, we compare the efficiencies of BLUEs obtained by RSS and SRS using the RE measure. To calculate the RE values of the BLUEs, firstly the variance covariance matrix of $\widehat{\mathbf{B}}$ in RSS is established. We assume that the dependent variable Y has a normal distribution. Then, $\sum_{i=1}^{m} \rho\eta_i = \rho\sum_{i=1}^{m} \eta_i = 0$ and $\sum_{i=1}^{m} \frac{\rho\eta_i}{(1-\rho^2)+\rho^2\tau_i} = 0$ are satisfied. Hence, the matrix of $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$ in (17) can be partitioned as follows,

$$(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} = \begin{bmatrix} C_{11}^{-1} & C_{12} \\ C_{21} & C_{22}^{-1} \end{bmatrix} \tag{18}$$

where,

$$C_{11}^{-1} = \left(\sum_{i=1}^{m} \frac{1}{(1-\rho^2)+\rho^2\tau_i}\right)^{-1} \begin{bmatrix} c & \sum_{j=1}^{c} x_{1j} & \sum_{j=1}^{c} x_{2j} & \cdots & \sum_{j=1}^{c} x_{pj} \\ & \sum_{j=1}^{c} x_{1j}^2 & \sum_{j=1}^{c} x_{1j}x_{2j} & \cdots & \sum_{j=1}^{c} x_{1j}x_{pj} \\ & & \sum_{j=1}^{c} x_{2j}^2 & \cdots & \sum_{j=1}^{c} x_{2j}x_{pj} \\ & & & \ddots & \cdots \\ & \text{sym} & & & \sum_{j=1}^{c} x_{pj}^2 \end{bmatrix}^{-1} \tag{19}$$

$$C_{22}^{-1} = \frac{1}{c\sum_{i=1}^{m} \dfrac{\rho^2\eta_i^2}{(1-\rho^2)+\rho^2\tau_i}} \tag{20}$$

$$C_{21} = C_{12}' = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \end{bmatrix}. \tag{21}$$

For the same sample size n, and assuming that Y is obtained using SRS with the same number of replications for each value set of the independent variables with RSS, the multiple linear regression model is defined in matrix form as follows,

$$\mathbf{Y}^* = \mathbf{X}^*\mathbf{B}^* + \boldsymbol{\epsilon}^* \tag{22}$$

where,

$$\mathbf{Y}^* = \begin{bmatrix} Y_{11} \\ Y_{21} \\ \vdots \\ Y_{m1} \\ Y_{12} \\ Y_{22} \\ \vdots \\ Y_{m2} \\ \vdots \\ Y_{1c} \\ Y_{2c} \\ \vdots \\ Y_{mc} \end{bmatrix} \quad \mathbf{X}^* = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{p1} \\ 1 & x_{11} & x_{21} & \cdots & x_{p1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{11} & x_{21} & \cdots & x_{p1} \\ 1 & x_{12} & x_{22} & \cdots & x_{p2} \\ 1 & x_{12} & x_{22} & \cdots & x_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{12} & x_{22} & \cdots & x_{p2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1c} & x_{2c} & \cdots & x_{pc} \\ 1 & x_{1c} & x_{2c} & \cdots & x_{pc} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1c} & x_{2c} & \cdots & x_{pc} \end{bmatrix} \quad \mathbf{B}^* = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \quad \epsilon^* = \begin{bmatrix} \epsilon_{11} \\ \epsilon_{21} \\ \vdots \\ \epsilon_{m1} \\ \epsilon_{12} \\ \epsilon_{22} \\ \vdots \\ \epsilon_{m2} \\ \vdots \\ \epsilon_{1c} \\ \epsilon_{2c} \\ \vdots \\ \epsilon_{mc} \end{bmatrix}. \tag{23}$$

In this model, $\boldsymbol{\epsilon}^*$ is the random error vector with $\mathrm{E}(\boldsymbol{\epsilon}^*) = 0$ and $\mathrm{Var}(\boldsymbol{\epsilon}^*) = \sigma^2\mathbf{I}$. Using the least square estimation technique, $\widehat{\mathbf{B}}^*$ and its variance are given respectively by;

$$\widehat{\mathbf{B}}^* = (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{Y}, \tag{24}$$

$$\mathrm{Var}(\widehat{\mathbf{B}}^*) = \sigma^2(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}. \tag{25}$$

Here, $\widehat{\mathbf{B}}^*$ is the BLUE estimator of the parameter vector $\mathbf{B}$ based on SRS. Then, the matrix of $(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}$ can be written as follows:

$$(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1} = \frac{\sum_{i=1}^{m} \frac{1}{(1-\rho^2)+\rho^2\tau_i}}{} mC_{11}^{-1}. \tag{26}$$

So, the RE of $\widehat{\mathbf{B}}$ relative to $\widehat{\mathbf{B}}^*$ can be obtained by dividing the diagonal elements of the matrix $\sigma^2(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}$ by the first $(p+1)$ diagonal elements of the matrix $\sigma^2(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$. As seen from (18), the first $(p+1)x1$ part of $\sigma^2(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$ is $\sigma^2 C_{11}^{-1}$. So, using (19) and (26), the RE of $\widehat{\mathbf{B}}$ relative to $\widehat{\mathbf{B}}^*$ is given by;

$$\mathrm{RE}(\widehat{\beta}_l^*, \widehat{\beta}_l) = \frac{\mathrm{Var}(\widehat{\beta}_l^*)}{\mathrm{Var}(\widehat{\beta}_l)} = \frac{\sum_{i=1}^{m} \frac{1}{(1-\rho^2)+\rho^2\tau_i}}{m}, \; l = 0, 1, 2, \ldots, p. \tag{27}$$

Let $\widehat{\sigma}$ be the estimator of σ for RSS and $\widehat{\sigma}^*$ the estimator of σ for SRS. To find the RE of $\widehat{\sigma}$ relative to $\widehat{\sigma}^*$, it is necessary to derive variance formula for $\widehat{\sigma}^*$. Barnett and Barreto [3] considered the variance of $\widehat{\sigma}^*$ using the Rao-Cramer lower-bound variance from Kendall and Stuart [10]. The least square estimation of $\widehat{\sigma}^*$ can be defined as;

$$\widehat{\sigma}^* = \sqrt{\frac{1}{mc-(p+1)} \sum_{j=1}^{c}\sum_{i=1}^{m} (Y_{ij} - \widehat{\beta}_0 - \widehat{\beta}_1 x_{1j} - \widehat{\beta}_2 x_{2j} - \cdots - \widehat{\beta}_p x_{pj}}, \tag{28}$$

which is asymptotically unbiased for σ. In this study, based on the assumption of normal distribution, the estimator $\widehat{\sigma}_{\mathrm{UMVU}}$, which is a uniformly minimum variance unbiased estimator of σ under SRS, will be used [9] and it is written as,

$$\widehat{\sigma}_{\mathrm{UMVU}} = \frac{\sqrt{mc-(p+1)}\,\Gamma\left(\frac{mc-(p+1)}{2}\right)}{\sqrt{2}\,\Gamma\left(\frac{mc-(p+1)+1}{2}\right)} \widehat{\sigma}^*. \tag{29}$$

The variance of $\widehat{\sigma}_{\mathrm{UMVU}}$ is defined as;

$$\mathrm{Var}(\widehat{\sigma}_{\mathrm{UMVU}}) = \sigma^2 \left[\frac{mc-(p+1)}{2} \left(\frac{\Gamma\left(\frac{mc-(p+1)}{2}\right)}{\Gamma\left(\frac{mc-(p+1)+1}{2}\right)} \right)^2 - 1 \right]. \tag{30}$$

Also, the RE of $\widehat{\sigma}$ relative to $\widehat{\sigma}_{\mathrm{UMVU}}$ is obtained by dividing (30) by the $(p+2)^{nd}$ diagonal element of $\sigma^2(X'V^{-1}X)^{-1}$, which is $\sigma^2 C_{22}^{-1}$, and so $\mathrm{RE}(\widehat{\sigma}, \widehat{\sigma}_{\mathrm{UMVU}})$ is given by,

$$\begin{aligned} \mathrm{RE}(\widehat{\sigma}, \widehat{\sigma}_{\mathrm{UMVU}}) &= \frac{\mathrm{Var}(\widehat{\sigma}_{\mathrm{UMVU}})}{\mathrm{Var}(\widehat{\sigma})} \\ &= c \sum_{i=1}^{m} \frac{\rho^2\eta_i^2}{(1-\rho^2)+\rho^2\tau_i} \\ &\qquad \times \left[\frac{mc-(p+1)}{2} \left(\frac{\Gamma\left(\frac{mc-(p+1)}{2}\right)}{\Gamma\left(\frac{mc-(p+1)+1}{2}\right)} \right)^2 - 1 \right]. \end{aligned} \tag{31}$$

As can be seen from (27) and (31), the value of $\mathrm{RE}(\widehat{\beta}_l^*, \widehat{\beta}_l)$ depends only on $|\rho|$, τ_i and m, but the value of $\mathrm{RE}(\widehat{\sigma}, \widehat{\sigma}_{\mathrm{UMVU}})$ depend on $|\rho|$, τ_i, η_i, m, c and p. The values calculated for $\text{RE}(\widehat{\beta}_l^*, \widehat{\beta}_l)$ with $|\rho| = 1.00,\ 0.90,\ 0.70,\ 0.50,\ 0.30,\ 0.10$, and $m = 2(1)10$ are given in Table 1.

Table 1. The RE values of $\widehat{\beta}_l^*$ relative to $\widehat{\beta}_l$ for $l = 0, 1, 2, \ldots, p$.

$\text{RE}(\widehat{\beta}_l^*,\ \widehat{\beta}_l)$						
$\lvert\rho\rvert$						
m	**1,00**	**0,90**	**0,70**	**0,50**	**0,30**	**0,10**
2	1,4669	1,3474	1,1848	1,0865	1,0295	1,0032
3	1,9345	1,6388	1,3069	1,1358	1,0449	1,0048
4	2,4040	1,8873	1,3940	1,1679	1,0545	1,0058
5	2,8751	2,1018	1,4596	1,1907	1,0611	1,0064
6	3,3475	2,2890	1,5107	1,2076	1,0659	1,0069
7	3,8211	2,4538	1,5519	1,2208	1,0695	1,0073
8	4,2955	2,6000	1,5858	1,2314	1,0724	1,0076
9	4,7706	2,7308	1,6143	1,2400	1,0748	1,0078
10	5,2463	2,8485	1,6385	1,2473	1,0767	1,0080

Table 2. The RE values of $\widehat{\sigma}$ relative to $\widehat{\sigma}_{\text{UMVU}}$.

$\text{RE}(\widehat{\sigma},\ \widehat{\sigma}_{\text{UMVU}})$												
$p = 1$												
$c = 2$							$c = 10$					
$\lvert\rho\rvert$							$\lvert\rho\rvert$					
m	**1,00**	**0,90**	**0,70**	**0,50**	**0,30**	**0,10**	**1,00**	**0,90**	**0,70**	**0,50**	**0,30**	**0,10**
2	0,5103	0,3797	0,2020	0,0945	0,0322	0,0035	0.2629	0,1956	0,1040	0,0487	0,0166	0,0018
3	0,6747	0,4754	0,2359	0,1061	0,0354	0,0038	0,4612	0,3250	0,1612	0,0725	0,0242	0,0026
4	0,8303	0,5561	0,2610	0,1141	0,0375	0,0040	0,6356	0,4257	0,1998	0,0873	0,0287	0,0031
5	0,9800	0,6256	0,2804	0,1200	0,0390	0,0041	0,7976	0,5092	0,2282	0,0976	0,0317	0,0034
6	1,1258	0,6867	0,2960	0,1245	0,0401	0,0042	0,9521	0,5808	0,2503	0,1053	0,0340	0,0036
7	1,2691	0,7411	0,3089	0,1281	0,0410	0,0043	1,1019	0,6435	0,2682	0,1113	0,0356	0,0038
8	1,4105	0,7901	0,3197	0,1311	0,0418	0,0044	1,2484	0,6993	0,2829	0,1160	0,0370	0,0039
9	1,5505	0,8346	0,3289	0,1336	0,0424	0,0044	1,3925	0,7495	0,2954	0,1200	0,0380	0,0040
10	1,6895	0,8753	0,3370	0,1357	0,0429	0,0045	1,5348	0,7951	0,3061	0,1233	0,0389	0,0041
$p = 6$												
$c = 2$							$c = 10$					
$\lvert\rho\rvert$							$\lvert\rho\rvert$					
m	**1,00**	**0,90**	**0,70**	**0,50**	**0,30**	**0,10**	**1,00**	**0,90**	**0,70**	**0,50**	**0,30**	**0,10**
2	-	-	-	-	-	-	0,3658	0,2722	0,1448	0,0677	0,0231	0,0025
3	-	-	-	-	-	-	0,5625	0,3963	0,1967	0,0884	0,0295	0,0032
4	5,4789	3,6696	1,7224	0,7528	0,2474	0,0263	0,7326	0,4907	0,2303	0,1007	0,0331	0,0035
5	2,7133	1,7322	0,7764	0,3322	0,1080	0,0114	0,8908	0,5687	0,2549	0,1091	0,0355	0,0038
6	2,2978	1,4016	0,6042	0,2541	0,0819	0,0087	1,0424	0,6358	0,2741	0,1153	0,0372	0,0039
7	2,2035	1,2868	0,5363	0,2225	0,0713	0,0075	1,1897	0,6948	0,2895	0,1201	0,0385	0,0041
8	2,2134	1,2399	0,5016	0,2057	0,0655	0,0069	1,3342	0,7474	0,3024	0,1240	0,0395	0,0042
9	2,2698	1,2217	0,4815	0,1956	0,0620	0,0065	1,4766	0,7948	0,3133	0,1272	0,0403	0,0042
10	2,3508	1,2178	0,4689	0,1888	0,0596	0,0062	1,6176	0,8380	0,3226	0,1299	0,0410	0,0043

In Table 2, the values of $RE(\widehat{\sigma}, \widehat{\sigma}_{UMVU})$ are given for $|\rho| = 1.00,\ 0.90,\ 0.70,\ 0.50,\ 0.30,\ 0.10$, $m = 2(1)10$ at $c = 2$ and 10; $p = 1$ and 6.

Based on the assumption of a normal distribution, τ_i and η_i denote respectively the mean and the variance of the i^{th} order statistic from the standard normal distribution. These values are taken from Arnold et al. [1] for $m = 2(1)10$.

4. Conclusion

As presented in Table 1, the values of $\mathrm{RE}(\widehat{\beta}_l^*, \widehat{\beta}_l)$ are equal for the same values of m, and they increase as the set size m increases for all values of $|\rho|$. Because there is no ranking error, the RE values have a maximum at $|\rho| = 1,00$. In addition, when $|\rho|$ is decreasing, the RE values converge to 1 and consequently the efficiency of RSS relative to SRS decreases.

In Table 2, the following comments can be listed;

- For $p = 6$ and $c = 2$, the values of $RE(\widehat{\sigma}, \widehat{\sigma}_{\mathrm{UMVU}})$ cannot be calculated at $m = 2$ and $m = 3$, since $n \geq (p + 1)$ is not valid.
- For $p = 1$, which is the same as the simple regression model, the values of $\mathrm{RE}(\widehat{\sigma}, \widehat{\sigma}_{\mathrm{UMVU}})$ increase with m for fixed values of c and $|\rho|$.
- For $p = 1$ and $|\rho| = 1,00$, the value of $\mathrm{RE}(\widehat{\sigma}, \widehat{\sigma}_{\mathrm{UMVU}})$ exceeds 1 when $m > 5$ and $c = 2$, and when $m > 6$ and $c = 10$.
- For fixed values of c and m, when p increases the value of $\mathrm{RE}(\widehat{\sigma}, \widehat{\sigma}_{\mathrm{UMVU}})$ also increases.
- The values of RE reach a maximum for $|\rho| > 0.7$, $p = 6$, $c = 2$ and $m = 4$. However, on increasing the value of m, $RE(\widehat{\sigma}, \widehat{\sigma}_{\mathrm{UMVU}})$ decrease rapidly, for example from $5,4789$ to $2,7133$.

Finally, $\mathrm{RE}(\widehat{\beta}_l^*,\ \widehat{\beta}_l)$ and $\mathrm{RE}(\widehat{\sigma},\ \widehat{\sigma}_{\mathrm{UMVU}})$ take their highest values at $|\rho| = 1,00$, because of the lack of ranking error. When $|\rho|$ approaches 0.1, $\mathrm{RE}(\widehat{\beta}_l^*,\ \widehat{\beta}_l)$ converge to 1 but $\mathrm{RE}(\widehat{\sigma},\ \widehat{\sigma}_{\mathrm{UMVU}})$ converges to 0. In addition, estimation of the parameter σ is negatively affected by ranking error much more than the estimation of the parameter vector $\mathbf{B}$.

Although, RSS is not such an effective technique for estimating the parameter σ, it is a more efficient technique than SRS for estimating the parameter vector $\mathbf{B}$ for all values of $|\rho|$ considered.

References

[1] Arnold,B., Balakrishnan, N., and Nagaraja, H. N. *A First Course in Order Statistics* (John Wiley, New York, 1993).

[2] Balakrishnan, N. and Li, T. 2006. *BLUEs of parameters of generalized geometric distribution using ordered ranked set sampling*, Communications in Statistics: Simulation and Computation **34**, 253–266, 2006.

[3] Barnett, V. and Barreto, M. C. M. *Best linear unbiased estimators for the simple linear regression model using ranked set sampling*, Environmental and Ecological Statistics **6**, 119–133, 1999.

[4] Barnet, V. and Barreto, M. C. M. *Estimators for a Poisson parameter using ranked set sampling*, Journal of Applied Statistics **8**, 929–941, 2001.

[5] Barnett, V. and Moore, K. *Best Linear unbiased estimates in ranked set sampling with particular referance to imperfect ordering*, Journal of Applied Statistics **24**, 697–710, 1997.

[6] David, H. A. and Levine, D. N. *Ranked set sampling in the presence of judgment error*, Biometrics **28**, 553–555, 1972.

[7] Dell, D. R. and Clutter, J. L., *Ranked set sampling theory with order statistics background*, Biometrics **28**, 545–555, 1972.

[8] Gang, Z. and Al-Saleh, M. *Improving the best linear unbiased estimator for the scale parameter of symmetric distributions by using the absolute value of ranked set samples*, Journal of Applied Statistics **30**, 253–265, 2003.

[9] Graybill, F. A. *An Introduction to Linear Statistical Models, Vol. 1* (McGraw-Hill Book Company, NewYork, 1961).

[10] Kendall, S., Kendall, M. and Stuart, A. *The Advanced Theory of Statistics, Vol. 2* (Charles Griffin and Company Limited, London, 1979).

[11] Lloyd, E. H. In *Contributions to Order Statistics* (John Wiley, New York, 1952).

[12] McIntyre, G. A. *A method of unbiased selective sampling using ranked sets*, Australian Journal of Agricaltural Research **3**, 385–390, 1952.

[13] Özdemir, Y. A. *Investigation of Parameter Estimation In Linear Regression Models Using Ranked Set Sampling* (Ph.D. Thesis, Ankara, 2005).

[14] Ridout, M. S. and Coby, J. M. *Ranked set sampling with non-random selection of sets and errors in ranking*, Applied Statistics **36**, 145–152, 1987.

[15] Sinha, Bimal K., Sinha, Bikas K. and Purkayastha, S. *On some aspects of ranked set sampling for estimation of normal and exponential parameters*, Statistical Decisions **14**, 223–240, 1996.

[16] Stokes, L. S. *Ranked set sampling with concomitant variables*, Commun. Statist. Theor. Meth. **A6** (12), 1207–1211, 1977.

[17] Stokes, L. S. *Parametric ranked set sampling*, Annals of the Institude of Statistical Mathematics **47**, 465–482, 1995.

[18] Takahasi, K. and Wakimoto, K. *On Unbiased estimates of the population mean based on the sample stratified by means of ordering*, Annals of The Institude of Statistical Mathematics **21**, 249–255, 1968.