Mathematics

Research article

A modified regression-cum-ratio estimator for finite population mean in presence of nonresponse using ranked set sampling

Mehreen Fatima[1],*, Saman Hanif Shahbaz[2], Muhammad Hanif[1] and Muhammad Qaiser Shahbaz[2]

[1] Department of Statistics, National College of Business Administration and Economics, Lahore 54000, Pakistan

[2] Department of Statistics, King Abdulaziz University, Jeddah, Saudi Arabia

* Correspondence: Email: Fatima.stat0786@gmail.com; Tel: +923444273700.

Abstract: Several situations arise where decision-making is required for some characteristics of an asymmetrical population for example estimation of the weekly number of server breakdowns at a company. The estimation methods based upon classical sampling designs are not suitable in such situations and some specialized methods and/or estimators are required. The ranked set sampling is a procedure that is suitable in such situations. In this paper, a new estimator is proposed that can be used to estimate population characteristics in case of asymmetrical populations. The proposed estimator is useful for estimation of population mean in the presence of non-response in study variable by using ranked set sampling procedure. The estimator is based upon two auxiliary variables to reduce the effect of asymmetry. The use of two auxiliary variables is also helpful in minimizing the variation in the estimation of the population mean of the study variable. The ranked set sampling procedure is used to get better accuracy as the actual measurements may be time-consuming, expensive, or difficult to obtain in a small sample size. The use of ranked set sampling also reduces the effect of asymmetry in the characteristics under study. The expressions for the mean square error and bias for the proposed estimators have been derived. The performance of the proposed estimator is evaluated by using real-life data and a simulation study is carried out to get an overview of efficiency. The relative efficiency of the proposed estimator is compared with some existing estimators. It has been found that the proposed estimator is highly efficient as compared with Mohanty's regression cum ratio estimator in simple random sampling and is more reliable in the case of non-response with a small sample size.

1. Introduction

There are several situations where some characteristics are to be estimated for a highly skewed or asymmetrical population. The sample drawn from such a population by using simple random sampling may lead to some unreliable results. The *ranked set sampling* (RSS) is a much appropriate method to draw the sample in such situations as it is based upon ordering a large number of sampling units on the basis of their relative sizes and then selecting a smaller number of units in all ranks for actual measurements. The *ranked set sampling* (RSS) thus improves the estimates in such a way that the amount of sampling error is decreased. The improvement of estimation is also subjected to the availability of some suitable auxiliary variables that are correlated with the variable under study.

This technique was first introduced by [1], where the units of study variable are ranked by using simple judgment. An unbiased estimator of the population mean under ranked set sampling has been proposed by [2]. The use of an auxiliary variables has also played an important role in ranking the study variable as proposed by [3,4]. The utilization of an auxiliary variable is also suitable to reduce the sampling variance and improve efficiency. Several authors have argued the use of auxiliary variables in estimation for more efficient results. It is pertinent that the increase in the number of auxiliary variables will increase the efficiency of the estimates. The use of two auxiliary variables has been studied by several authors in the case of simple random and ranked set sampling. Some notable references in this regard are [5–9].

Sometimes it happens that information on some of the units of study variable is not available, for example, we may not be able to record information about some sensitive characteristics from some of the units. In such cases, the non–response is observed that asks for specialized treatment. Usually, the problem of non-response is reduced by using the subsampling technique and various authors have proposed estimators to overcome this situation. A subsampling approach to adjust the nonresponse in the survey has been proposed by [10]. The problem of non-response can also be reduced by using the information of some auxiliary variable with the full response as discussed by [11]. Some other notable references on dealing with the non-response are [12–19].

The estimation in ranked set sampling is done to reduce the sampling variance and different new estimators have been proposed from time to time by various authors to reduce the sampling variance. A class of ratio-in-exponential type estimators in ranked set sampling has been proposed by [20] by using the information of two auxiliary variables. A single auxiliary variable ratio-cum-product estimators of finite population mean in ranked set sampling has been proposed by [21]. The median and neoteric ranked set sampling provides more efficient estimates as compared with the classical ranked set sampling and [22] have proposed some regression estimators in median and neoteric ranked set sampling. These estimators are based upon single and two auxiliary variables. The transformation of auxiliary variable is sometimes helpful in improving the efficiency of the estimates as discussed by [23] where the authors have proposed a ranked set sampling estimator by using transformation of the auxiliary variable. The new methods of some log type class of estimators using *RSS* have been proposed that have been considered as novel class of estimators in current literature by [24].

In some situations, some sensitive auxiliary information is to be used and non-response is evident for such an auxiliary variable. Such situations usually arise in medical research where the study variable is related to post-operative conditions or transplantation of organs and the collection of observation is difficult or missed. Such situations also have an effect on the sample size and in this scenario the ranked set sampling technique is suitable. In this paper, we will propose a new regression-cum-ratio estimator for estimation of population mean when the study variable is sensitive or it is difficult to obtain information about that variable in a normal sample size. The estimator has been proposed by using the information of two auxiliary variables with the view that one of the variables has a high correlation with the study variable.

The plan of the paper follows. A brief description about the ranked set sampling with two auxiliary variables in presence of nonresponse is given in Section 2. Some existing ranked set sampling estimators are given in Section 3. The new estimator is proposed in Section 4 alongside the expressions for mean square error and bias of the proposed estimator. The results of the simulation study are given in Section 5 alongside the relative efficiency of the proposed estimator as compared with Mohanty's estimator. A numerical study is presented in Section 6 followed by conclusions in Section 7.

2. Ranked set sampling with two auxiliary variables in presence of non-response

The ranked set sampling is based upon ranking the variable of interest either visually or by any cost independent ranking method. This type of sampling differs from conventional random sampling and hence requires some description. In this section, description of the ranked set sampling technique is given. The technique is described below.

Suppose we have a finite population W, of size N, from where the sample is to be drawn. Suppose further that two auxiliary variables, X and Z, are also known that are highly correlated with the study variable Y. Let μ_x and μ_z are population mean of X and Z respectively. It is assumed that the information about the study variable Y, with the population mean μ_y, is not easy to obtain and potentially nonresponse exist and hence auxiliary information is to be utilized.

Suppose that the population W is divided into two subgroups W_1 and W_2 such that $W = W_1 + W_2$ and let information on $\left(X_i, Z_i, Y_i \right)$ is obtained at the first attempt from W_1 whereas information about Y_i is not available from W_2 at the first attempt. Also, let $s_i = s_1 + s_2$ be the sample at the ith draw with $s_i \subset W_i$. The ranked set sampling starts by drawing a simple random sample of size n from the population and recording information about auxiliary variables X and Z. The other steps in selecting a ranked set sample are given below.

1) Select m^2 values on X, Z and Y from the sample of size n.
2) Rank m^2 values in m sets based on the variable X.
3) Select first observation from set 1 on three variables and let the collection of observations on three variables be $\left\{ x_{1(1)}, z_{1(1)}, y_{1(1)} \right\}$. Continue this procedure on all m sets.
4) Repeat steps 1–3 r times where r is the pre-decided number of cycles. In this way, $m \times r$ observations are selected using r cycles of m^2 observations.

It is assumed that some non-response has occurred such that complete response is available on a sample $s_1 \subset W_1$ of size $n_1 = n_1' + n_2^*$, where n_1' is the number of observations recorded at the first attempt and n_2^* are the number of observations of study variable that are not recorded from the first

sample. Also, assume that incomplete response is available on $s_2 \subset W_2$ observations. The additional steps of ranked set sampling are based upon the procedures given by [10] and [11] and are given below.

5) Following [10], consider $s'_{2(j)} = s_{2(j)}/k_{(j)}$.

6) Draw a second phase sample of size s_2 and record information on *X, Y* and *Z*.

Some notations to be used in ranked set sampling are:

The means and variances of auxiliary variables are:

$$\bar{X}_j = m^{-1}\sum_{j=1}^{m} X_j \; ; \bar{Z}_j = m^{-1}\sum_{j=1}^{m} Z_j \; ; Var\left(X_{(j)}\right) = m^{-1}\sum_{j=1}^{m}\delta_{x(j)}^2 \; ; Var\left(Z_{(j)}\right) = m^{-1}\sum_{j=1}^{m}\delta_{z(j)}^2 \; ,$$

where $\delta_{x(j)}^2 = (n-1)^{-1}\sum_{j=1}^{m}\left(x_{(j)} - \bar{x}_{(j)}\right)^2$ and $\delta_{z(j)}^2 = (n-1)^{-1}\sum_{j=1}^{m}\left(z_{(j)} - \bar{z}_{(j)}\right)^2$.

The variances for ranked set sampling are

$$Var\left(X_{(RSS)}\right) = \left(m^2 r\right)^{-1}\sum_{j=1}^{m}\delta_{x(j)}^2 \quad \text{and} \quad Var\left(Z_{(RSS)}\right) = \left(m^2 r\right)^{-1}\sum_{j=1}^{m}\delta_{z(j)}^2$$

Let $\bar{y}'_{1(j)}$ denotes sample mean of study variable at the first phase s_1, $\bar{y}_{2(j)}$ denotes mean of subsample s_2 and $\bar{y}'_{2(j)}$ denotes the sample mean of study variable at second phase s_2 then the weighted estimator of μ_y is $\bar{\bar{y}}_{(j)} = q_{1(j)}\bar{y}'_{1(j)} + q_{2(j)}\bar{y}'_{2(j)}$ where $q_i = m_i/m$.

The estimator of μ, proposed by [24], in case of ranked set sampling is

$$\bar{\bar{y}}_{RSS} = m^{-1}\sum_{j=1}^{m}\bar{\bar{y}}_{(j)} \tag{1}$$

with $\quad V\left(\bar{\bar{y}}_{RSS}\right) = \left(m^2 r\right)^{-1}\left[\sum_{j=1}^{m}\delta_{(j)}^2 + \sum_{j=1}^{m}q_{2(j)}\left\{k_{(j)} - 1\right\}\delta_{2(j)}^*\right]. \tag{2}$

The sample mean of study variable, in RSS, for each rank *j* is written

$$\bar{\bar{y}}_{j(RSS)} = q_{1(j)}\bar{y}'_{1(j)} + q_{2(j)}\bar{y}'_{2(j)} + q_{2(j)}\left[\bar{y}'_{2(j)} - \bar{y}_{2(j)}\right]. \tag{3}$$

The sample means of two auxiliary variables X and Z at a specific rank *j* are $\bar{x}_{j(RSS)} = m_i^{-1}\sum_{i=1}^{m}x_{ij}$ and $\bar{z}_{j(RSS)} = m_i^{-1}\sum_{i=1}^{m}z_{ij}$. We also have $\bar{\bar{e}}^*_{yRSS} = \mu_y - \bar{\bar{y}}_{RSS}$ as error in estimation for μ_y such that

$$E\left(\bar{\bar{e}}^{*2}_{yRSS}\right) = \left(m^2 r\right)^{-1}\left[\sum_{j=1}^{m}\delta_{y(j)}^2 + \sum_{j=1}^{m}q_{2(j)}\left\{k_{(j)} - 1\right\}\delta_{y2(j)}^{*2}\right] \tag{4}$$

$$E\left(\bar{\bar{e}}^2_{xRSS}\right) = \left(m^2 r\right)^{-1}\sum_{j=1}^{m}\delta_{x(j)}^2 \tag{5}$$

$$E\left(\bar{\bar{e}}^2_{zRSS}\right) = \left(m^2 r\right)^{-1}\sum_{j=1}^{m}\delta_{z(j)}^2 \tag{6}$$

$$E\left(\bar{\bar{e}}^*_{yRSS}\bar{\bar{e}}_{xRSS}\right) = \left(m^2 r\right)^{-1}\left[\sum_{j=1}^{m}\left\{q_{1(j)}\delta'_{x(j)y(j)} + q_{2(j)}\delta^*_{x(j)y(j)}\right\} - \sum_{j=1}^{m}q_{2(j)}\left\{k_{(j)} - 1\right\}\delta^*_{x(j)y(j)}\right] \tag{7}$$

$$E\left(\bar{\bar{e}}^*_{yRSS}\bar{\bar{e}}_{zRSS}\right) = \left(m^2 r\right)^{-1}\left[\sum_{j=1}^{m}\left\{q_{1(j)}\delta'_{z(j)y(j)} + q_{2(j)}\delta^*_{z(j)y(j)}\right\} - \sum_{j=1}^{m}q_{2(j)}\left\{k_{(j)} - 1\right\}\delta^*_{z(j)y(j)}\right] \tag{8}$$

$$E\left(\bar{\bar{e}}_{xRSS}\bar{\bar{e}}_{zRSS}\right) = \left(m^2 r\right)^{-1}\sum\nolimits_{j=1}^{m}\delta_{x(j)z(j)}$$ (9)

Above joint expectations are equal to specific covariance and are useful in deriving the mean square error of estimator in ranked set sampling.

3. Some existing estimators

In this section, we will discuss some existing estimators in ranked set sampling. These estimators are discussed below.

1) A ratio-cum-product estimator by [21]. The estimator is

$$\bar{y}_1 = \bar{y}_{[n]}\left[\alpha\left\{\frac{\bar{X}C_x + \beta_2(x)}{\bar{x}_{(n)}C_x + \beta_2(x)}\right\} + (1-\alpha)\left\{\frac{\bar{x}_{(n)}C_x + \beta_2(x)}{\bar{X}C_x + \beta_2(x)}\right\}\right]$$ (10)

where α is a suitably chosen constant. The MSE and Bias are:

$$MSE\left(\bar{y}_1\right) = \bar{Y}^2\left[\theta S_y^2\left(1-\rho_{yx}^2\right) - \left\{W_{y[i]} - \frac{1}{2}(t_3 - K)W_{x(i)}\right\}^2\right].$$ (11)

and

$$B\left(\bar{y}_1\right) = \bar{Y}\left[\theta t_3 C_x^2\left\{K + \alpha t_3\left(t_t - 2K\right)\right\} - \left\{\alpha t_3^2 W_{x(i)}^2 + (1-2\alpha)t_3 W_{yx(i)}\right\}\right],$$

where $t_3 = \bar{X}C_x\left[\bar{X}C_x + \beta_2(x)\right]^{-1}$ and $K = \rho C_y/C_x$.

2) A class of ratio-in-exponential-type estimators proposed by [20]. The estimator is

$$\bar{y}_2 = \bar{y}_{rss}\left(\frac{\bar{X}}{\bar{x}_{rss}}\right)^{\alpha_1}\left(\frac{\bar{Z}}{\bar{z}_{rss}}\right)^{\alpha_2}\left[k\exp\left(\frac{\bar{X} - \bar{x}_{rss}}{\bar{X} + \bar{x}_{rss}}\right) + (1-k)\exp\left(\frac{\bar{Z} - \bar{z}_{rss}}{\bar{Z} + \bar{z}_{rss}}\right)\right],$$ (12)

where $\bar{y}_{rss} = \left(mr\right)^{-1}\sum\nolimits_{j=1}^{r}\sum\nolimits_{i=1}^{m}y_{[i:m]j}$, $\bar{x}_{rss} = \left(mr\right)^{-1}\sum\nolimits_{j=1}^{r}\sum\nolimits_{i=1}^{m}x_{[i:m]j}$ and $\bar{z}_{rss} = \left(mr\right)^{-1}\sum\nolimits_{j=1}^{r}\sum\nolimits_{i=1}^{m}z_{[i:m]j}$. Also α_1 and α_2 are unknown and their values are to be determined. The mean square error of this estimator is

$$MSE\left(\bar{y}_2\right) \cong \bar{Y}^2\left[V_{200} + \frac{1}{4}\left(k + 2\alpha_1\right)^2 V_{020} + \frac{1}{4}\left(1 - k + 2\alpha_2\right)^2 V_{002} - \left(k + 2\alpha_1\right)V_{110}\right.$$
$$\left. - \left(1 - k + 2\alpha_2\right)V_{101} + \frac{1}{2}\left(k + 2\alpha_1\right)\left(1 - k + 2\alpha_2\right)V_{011}\right].$$ (13)

3) A class of regression cum ratio estimators of population mean in ranked set sampling by [23]

$$\bar{y}_3 = \left[\bar{y}_{rss} + b\left(\bar{X} - \bar{x}_{rss}\right)\right]\left(\frac{\bar{A}\lambda + \delta}{\bar{a}_k\lambda + \delta}\right)$$ (14)

where λ and δ are known constants and can be population coefficient of kurtosis, skewness, variation and correlation, etc. Also $b_{yx} = \bar{Y}\rho_{yx}C_y / \bar{X}C_x$, $\bar{a}_k = \bar{x}_{rss} + \bar{X}(t_k - 1)$, $\bar{A} = \bar{X}t_k$ for $k = 1, 2, 3, 4, 5, 6$, with $t_1 = C_x$, $t_2 = \beta_{2(x)}$, $t_4 = \rho_{yx}$, $t_5 = Q_{1(x)}$ and $t_6 = Q_{3(x)}$. Also C_x is the coefficient of variation, $\beta_{1(x)}$ is the coefficient of skewness, $\beta_{2(x)}$ is the coefficient of kurtosis, ρ_{yx} is the coefficient of correlation, $Q_{1(x)}$ is the lower quartile and $Q_{3(x)}$ is the upper quartile.

The MSE and Bias presented of this class of estimators is

$$MSE(\bar{y}_3) = \bar{Y}\left[\gamma\left\{C_y^2 + (\varphi + \rho C_y / C_x)C_x^2 - 2(\varphi + \rho C_y / C_x)C_{yx}\right\} \right. \\ \left. - \left\{W_y - (\varphi + \rho C_y / C_x)W_x\right\}\right]. \tag{15}$$

Different estimators have been proposed by [23] by using different values of γ and δ in (14)

In the following section, we will propose a new estimator for the population mean in the case of ranked set sampling. The estimator is proposed by using information on two auxiliary variables. The estimator is motivated by the fact that the additional auxiliary variable, having a higher correlation with the study variable, will improve the efficiency of the estimates. These types of estimators are useful for the estimation of population characteristics in different types of studies, for example, in information technology we may be interested to estimate the average time taken by a server to process a task by using the information of the processors and the installed RAM. The proposed estimator will also be suitable when there is non-response on some auxiliary variables.

4. A new ranked set sampling estimator with two auxiliary variables

In the following we will propose a new estimator in ranked set sampling with two auxiliary variables. The proposed estimator is

$$t^*_{RSS} = \bar{y}_2^* + b_{yx}(\bar{X} - \bar{x}_1)\frac{\bar{Z}}{z_1} \tag{16}$$

Using $\bar{y}_2^* = \bar{Y} + \bar{\bar{e}}_{yRSS}^*$ etc, the above estimator can be written as

$$t^*_{RSS} = \bar{Y} + \bar{\bar{e}}_{yRSS}^* - b_{yx}\bar{\bar{e}}_{xRSS}\left(1 - \frac{\bar{\bar{e}}_{zRSS}}{\bar{Z}}\right)^{-1}.$$

Using Taylor expansion of $\left(1 - \bar{\bar{e}}_{zRSS}/\bar{Z}\right)^{-1}$ and retaining only the linear term, the above estimator can be written as

$$t^*_{RSS} - \bar{Y} = \bar{\bar{e}}_{yRSS}^* - b_{yx}\bar{\bar{e}}_{xRSS}\left(1 - \bar{\bar{e}}_{zRSS}/\bar{Z}\right) \tag{17}$$

The bias of the proposed estimator (16) can be easily obtained from (17) as

$$E\left(t^*_{RSS} - \bar{Y}\right) = \bar{Y}E\left(\bar{\bar{e}}_{yRSS}^* - b_{yx}\frac{\bar{X}}{\bar{Y}}\bar{\bar{e}}_{xRSS} + \frac{1}{\bar{Z}}\bar{\bar{e}}_{zRSS} - \frac{\bar{\bar{e}}_{yRSS}\bar{\bar{e}}_{zRSS}}{\bar{Z}} + \frac{b_{yx}\bar{X}\bar{\bar{e}}_{xRSS}\bar{\bar{e}}_{zRSS}}{\bar{Z}\bar{Y}}\right).$$

Applying the expectation, the bias of the proposed estimator is

$$E\left(t^*_{RSS} - \overline{Y}\right) = \frac{b_{yx}}{\overline{Z}} \frac{1}{m^2 r} \sum_{j=1}^{m} \delta_{x_{(j)}z_{(j)}} - \frac{1}{\overline{Z}(m^2 r)} \left[\sum_{j=1}^{m} \left(q_{1(j)} \delta'_{y_{(j)}z_{(j)}} + q_{2(j)} \delta^*_{y(j)z(j)} \right) - \sum_{j=1}^{m} q_{2(j)} \left\{ k_{(j)} - 1 \right\} \delta^*_{y_{(j)}z_{(j)}} \right].$$

Again, squaring (17) and applying expectation the mean square error of the proposed estimator is

$$MSE\left(t^*_{RSS}\right) = E\left(t^*_{RSS}\right)^2 = E\left(\overline{\overline{e}}_{yRSS} + \frac{\overline{Y}}{\overline{Z}} \overline{\overline{e}}_{zRSS} - b_{yx} \overline{\overline{e}}_{xRSS} \right)^2.$$

Expanding the square and applying expectations we have

$$MSE\left(t^*_{RSS}\right) = E\left(t^*_{RSS} - \overline{Y}\right)^2 = E\left(\overline{\overline{e}}^*_{yRSS} \right)^2 + \frac{\overline{Y}}{\overline{Z}^2} E\left(\overline{\overline{e}}_{zRSS} \right)^2 - b^2_{yx} E\left(\overline{\overline{e}}_{xRSS} \right)^2$$

$$-2\frac{\overline{Y}}{\overline{Z}} E\left(\overline{\overline{e}}^*_{yRSS} \overline{\overline{e}}_{zRSS} \right) - 2b_{yx} E\left(\overline{\overline{e}}^*_{yRSS} \overline{\overline{e}}_{xRSS} \right) + 2\frac{\overline{Y}}{\overline{Z}} b_{yx} E\left(\overline{\overline{e}}_{xRSS} \overline{\overline{e}}_{zRSS} \right).$$

Using expectations from (4) to (9) etc., we have

$$MSE\left(t^*_{RSS}\right) = \frac{1}{m^2 r} \left[\left\{ \sum_{j=1}^{m} \delta^2_{y(j)} + \sum_{j=1}^{m} q_{2(j)} \left(k_{(j)} - 1 \right) \delta^{*2}_{y_2(j)} \right\} + \frac{\overline{Y}^2}{\overline{Z}_2} \sum_{j=1}^{m} \delta^2_{z(j)} - b^2_{yx} \sum_{j=1}^{m} \delta^2_{x_{(j)}} \right.$$

$$-2\frac{\overline{Y}}{\overline{Z}} \rho_{yz} \left\{ \sum_{j=1}^{m} \left(q_{1(j)} \delta'_{y_{(j)}z_{(j)}} + q_{2(j)} \delta^*_{y(j)z(j)} \right) - \sum_{j=1}^{m} q_{2(j)} \left(k_{(j)} - 1 \right) \delta^{*2}_{y(j)z(j)} \right\}$$

$$-2b_{yx}\rho_{yx} \left\{ \sum_{j=1}^{m} \left(q_{1(j)} \delta'_{y_{(j)}x_{(j)}} + q_{2(j)} \delta^*_{y(j)x(j)} \right) - \sum_{j=1}^{m} q_{2(j)} \left(k_{(j)} - 1 \right) \delta^{*2}_{y(j)x(j)} \right\}$$

$$\left. +2\frac{\overline{Y}}{\overline{Z}} b_{yx}\rho_{xz} \sum_{j=1}^{m} \delta_{x(j)z(j)} \right]$$

Using $C^2_{y(2)} = \delta^2_{y(2)} / \overline{Y}^2$ etc., the mean square error can be written as

$$MSE\left(t^*_{RSS}\right) = \frac{\overline{Y}^2}{m^2 r} \left[\left\{ \sum_{j=1}^{m} C^2_{y(j)} + \sum_{j=1}^{m} q_{2(j)} \left(k_{(j)} - 1 \right) C^{*2}_{y_2(j)} \right\} + \sum_{j=1}^{m} C^2_{z(j)} - b^2_{yx} \sum_{j=1}^{m} C^2_{x_{(j)}} \right.$$

$$-2\rho_{yz} \left\{ \sum_{j=1}^{m} \left(q_{1(j)} C_{y_{(j)}z_{(j)}} + q_{2(j)} C^*_{y(j)z(j)} \right) - \sum_{j=1}^{m} q_{2(j)} \left(k_{(j)} - 1 \right) C^{*2}_{y(j)z(j)} \right\}$$

$$-2b_{yx}\rho_{yx} \left\{ \sum_{j=1}^{m} \left(q_{1(j)} C_{y_{(j)}x_{(j)}} + q_{2(j)} C^*_{y(j)x(j)} \right) - \sum_{j=1}^{m} q_{2(j)} \left(k_{(j)} - 1 \right) C^{*2}_{y(j)x(j)} \right\} \tag{18}$$

$$\left. +2b_{yx}\rho_{xz} \sum_{j=1}^{m} C_{x(j)z(j)} \right]$$

where $b_{yx} = \overline{Y} C_y \rho_{yx} / \overline{X} C_x$. It is to be noted that the bias of the estimator is not zero.

In the following, we have conducted a simulation study to see the performance of the proposed estimator.

5. Simulation study

A simulation study has been conducted to see the performance of the proposed estimator as compared with the existing estimator. The simulation study has been conducted by generating random observations from the normal distribution. We have generated artificial populations of size $N = 5000$ on auxiliary variables as $x = rnorm(5000, 0, 1)$ and $z = rnorm(5000, 0, 1)$. The artificial population values of the study variable Y are generated by using the linear equation

$$Y_i = 0.02 + 0.004X + 0.02Z + \varepsilon_i$$

where ε_i is $N(0,1)$. After generating the random population we have computed the proposed estimator for different combinations of the parameters. The values of parameters that we have used are $m = 4, 5, 6$ and 7, $r = 20$, 30 and 40 and $k = 2$ and 4. A response of 60% is considered giving a non-response of 40%. We have drawn subsamples from the non-responded group. This procedure is repeated 20,000 times giving 20,000 values of the estimators. Mohanty's estimator is also computed by using the above-mentioned procedure. The mean square errors of the proposed estimator and Mohanty's estimator are computed by using the 20,000 values of each of the estimators at different combinations of the parameters. The mean square errors of the proposed estimator and its relative efficiency relative to Mohanty's estimator are given in Table 1 below.

Table 1. Mean square errors t^*_{RSS} and its relative efficiency.

	$MSE\left(t^*_{RSS}\right)$		Relative Efficiency	
	k = 2	**k = 4**	**k = 2**	**k = 4**
m = 2, r = 20	0.0000820	0.0000009	5.63135	5.63441
m = 2, r = 30	0.0000042	0.0000004	15.00015	12.13133
m = 2, r = 40	0.0000002	0.0000183	55.83486	35.81538
m = 3, r = 20	0.0000296	0.0004195	14.39062	14.59429
m = 3, r = 30	0.0000734	0.0000760	62.66464	62.92887
m = 3, r = 40	0.0000387	0.0000214	71.34428	51.71975
m = 4, r = 20	0.0000017	0.0000227	18.3212	10.1797
m = 4, r = 30	0.0000080	0.0000008	90.3362	82.6455
m = 4, r = 40	0.0000608	0.0000006	95.7313	95.6457
m = 5, r = 20	0.0002140	0.0000378	28.5117	16.0599
m = 5, r = 30	0.0000184	0.0000551	16.5051	10.2424
m = 5, r = 40	0.0000088	0.0000010	18.7858	17.5502
m = 6, r = 20	0.0000062	0.0000175	68.5501	55.7342
m = 6, r = 30	0.0000008	0.0000124	19.8887	13.5740
m = 6, r = 40	0.0000154	0.0000348	40.2506	35.0131
m = 7, r = 20	0.0000301	0.0000058	73.5355	85.9495
m = 7, r = 30	0.0000059	0.0000046	23.6716	18.3407
m = 7, r = 40	0.0000017	0.0000099	51.5398	46.0115

From the above table, we can see that the efficiency of the proposed estimator is far better as compared with Mohanty's estimator and hence the proposed estimator is much better than Mohanty's estimator.

We have also obtained the mean square error of the proposed estimator for different correlation combinations between study variable and auxiliary variables. These mean square errors are given in Table 2 below.

Table 2. Mean square error of proposed estimator for different correlation combinations.

m, r, k	Correlations between Y,X and Y, Z		
	0.72, 0.70	**0.90, 0.45**	**0.97, 0.26**
m = 4, r = 20, k = 4	0.000051063	0.000004507	0.000002055
m = 5, r = 20, k = 4	0.000885543	0.000235453	0.000004507
m= 6, r = 20, k = 4	0.000004314	0.000004948	0.000022795

From above table we can see that for small value of m, the mean square error decreases with increase in correlation coefficient between X and Y.

6. Numerical study

In this section, we have given a real data example for the proposed estimator. For this, we have used an artificial population of 60 patients of type–II diabetes II [25]. The study has been conducted by using the resistive index of the renal artery as the study variable, Y; alongside two auxiliary variables, blood urea, X; and serum creatinine, Z. The study variable had obvious non-response and hence the proposed estimator is suitable to estimate its mean. The calculated values are $\mu_X = 54.55$, $\mu_Z = 3.98$, $\mu_z = 3.98$, $\sigma_X^2 = 24.25$, $\sigma_Z^2 = 0.50$, $\sigma_Y^2 = 0.000249$, $\rho_{XY} = 0.575$, $\rho_{XZ} = 0.544$ and $\rho_{YZ} = 0.438$. The mean square errors of the proposed estimator under RSS turned out to be $MSE(t_{RSS}^*) = 0.000234$ and the mean square error of Mohanty's estimator under SRS turned out to be $MSE(t_M) = 0.001422$. We can see that the proposed estimator again has a much smaller mean square error as compared with Mohanty's estimator.

7. Conclusions

In this paper, we have proposed a new regression cum ratio estimator of mean by using the ranked set sampling approach. The estimator has been proposed by using the information of two auxiliary variables. The proposed estimator is useful for the estimation of population characteristics when the study variable has non-response. The expression for the mean square error of the proposed estimator has been obtained. The simulation study has been conducted to see the performance of the proposed estimator as compared with Mohanty's estimator. The simulation study has been conducted by using different proportions of non-response. Specifically, the proportion of non-response used is 50% ($k = 2$) and 25% ($k = 4$). The mean square error of the proposed estimator is compared with the Mohanty's estimator by using three different settings. We have seen that for all the combinations of m and r, the proposed estimator is better that the Mohanty's estimator.

Conflict of interest

The author declares no conflict of interest in this paper.

References

1. G. A. Mclntyre, A method for unbiased selective sampling using ranked sets, *Crop Pasture. Sci.,* **3** (1952), 385–390. https://doi.org/10.1071/AR9520385

2. K. Takahasi, K. Wakimoto, On unbiased estimates of the population mean based on the sample stratifed by means of ordering, *Ann. Inst. Stat. Math.,* **20** (1968), 1–31. https://doi.org/10.1007/BF02911622

3. S. L. Stokes, Ranked set sampling with concomitant variables, *Commun. Stat. Theory Methods,* **6** (1977), 1207–1211. https://doi.org/10.1080/03610927708827563

4. H. P. Singh, R. Tailor, S. Singh, General procedure for estimating the populations mean using ranked set sampling, *J. Stat. Comput. Simul.,* **84** (2014), 931–945. https://doi.org/10.1080/00949655.2012.733395

5. W. A. Abu-Dayyeh, M. S. Ahmed, R. A. Ahmed, H. A. Muttlak, Some estimators of a finite population mean using Mathematical Problems in Engineering auxiliary information, *Appl. Math. Comput.*, **139** (2003), 287–298. https://doi.org/10.1016/S0096-3003(02)00180-7

6. C. Kadilar, H. Cingi, A new estimator using two auxiliary variables, *Appl. Math. Comput.*, **162** (2005), 901–908. https://doi.org/10.1016/j.amc.2003.12.130

7. S. Malik, R. Singh, An improved estimator using two auxiliary attributes, *Appl. Math. Comput.*, **219** (2013), 10983–10986. https://doi.org/10.1016/j.amc.2013.05.014

8. P. Sharma, R. Singh, A class of exponential ratio estimators of finite population mean using two auxiliary variables, *Pak. J. Stat. Oper. Res.*, **1** (2015), 221–229. https://doi.org/10.18187/pjsor.v11i2.759

9. S. Muneer, J. Shabbir, A. Khalil, Estimation of finite population mean in simple random sampling and stratified random sampling using two auxiliary variables, *Commun. Stat. Theory Methods,* **46** (2017), 2181–2192. https://doi.org/10.1080/03610926.2015.1035394

10. M. N. Hansen, W. N. Hurwitz, The problem of non-response in sample surveys, *J. Am. Stat. Assoc.,* **41** (1946), 517–529. https://doi.org/10.1080/01621459.1946.10501894

11. A. A. Sodipo, K. O. Obisesan, Estimation of the population mean using difference cum ratio estimator with full response on the auxiliary character, *Res. J. Appl. Sci.,* **6** (2007), 769–772.

12. W. Cochran, *Sampling techniques*, 2 Eds., New York: Wiley and Sons, 1977.

13. P. S. R. S. Rao, Ratio estimation with sub sampling the non-respondents, *Surv. Methodol.*, **12** (1986), 217–230.

14. B. B. Khare, S. Srivastava, Estimation of population mean using auxiliary character in presence of non-response, *Natl. Acad. Sci. Lett. India*, **16** (1993), 111–114.

15. F. C. Okafor, H. Lee, Double sampling for ratio and regression estimation with sub-sampling the non-respondents, *Surv. Methodol.*, **26** (2000), 183–188.

16. H. P. Singh, S. Kumar, Estimation of mean in presence of non-response using two phase sampling schemes, *Stat. Pap.*, **5** (2010), 559–582. https://doi.org/10.1007/s00362-008-0140-5

17. G. K. Agarwal, S. M. Allende, C. N. Bouza, Double sampling with ranked set selection in the second phase with non-response: analytical results and Monte Carlo experiences, *J. Probab. Stat.,* **10** (2012), 12–20. https://doi.org/10.1155/2012/214959

18. S. Riaz, G. Diana, J. Shabbir, A general class of estimators for the population mean using multiphase sampling with the non-respondents, *Hacettepe J. Math. Stat.*, **43** (2014), 511–527.

19. B. B. Khare, S. Srivastava, Study of conventional and alternative two-phase sampling ratio, product and regression estimators in presence of non-response, *Proc. Natl. Acad. Sci. India Sect. A*, **65** (1995), 195–204.

20. L. Khan, J. Shabbir, Improved ratio-type estimators of population mean in ranked set sampling using two concomitant variables, *Pak. J. Stat. Oper. Res.*, **12** (2016), 507–518. https://doi.org/10.18187/pjsor.v12i3.1271

21. N. Mehta, V. L. Mandowara, A modified ratio-cum-product estimator of finite population mean using ranked set sampling, *Commun. Stat. Theory Methods,* **45** (2016), 267–276. https://doi.org/10.1080/03610926.2013.830748

22. N. Koyuncu, Regression estimators in ranked set, median ranked set and neoteric ranked set sampling, *Pak. J. Stat. Oper. Res.,* **14** (2018), 89–94, https://doi.org/10.18187/pjsor.v14i1.1825.

23. L. Khan, J. Shabbir, A. Khalil, A new class of regression cum ratio estimators of Population mean in ranked set sampling, *Life Cycle Reliab. Saf. Eng.,* **8** (2019), 201–204. https://doi.org/10.1007/s41872-019-00079-y

24. S. Bhushan, A. Kumar, S. A. Lone, On some novel classes of estimators using ranked set sampling, *Alexandria Eng. J.*, **61** (2021), 5465–5474. https://doi.org/10.1016/j.aej.2021.11.001

25. M. Nasir, *Correlation between renal artery resistive index and renal function tests in type II diabetic patients*, MPhil dissertation, The University of Lahore, Lahore Pakistan, 2021.