Int. J. Nonlinear Anal. Appl. 13 (2022) 1, 3537-3549
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2022.6134

Some ratio estimators of finite population variance using auxiliary information in ranked set sampling

Rikan A. Ahmed[a,*], Saja Mohammad Hussein[b]

[a]Department of Statistics and Informatics, Mosul University, Iraq
[b]Department of Statistics, Baghdad University, Iraq

(Communicated by Madjid Eshaghi Gordji)

Abstract

The method of selecting or designing the sample may be expensive or take a long time in some studies. And with the existence of the relationship between the main and auxiliary variables, which can employ in the process of selecting sampling units through the possibility of ranking for the auxiliary variable at the lowest possible cost. Ranked set sampling (RSS) is a method to achieve this objective, and in sample surveys, it is usual to use auxiliary information to increase the precision of estimators. This article addresses the problem of estimating the finite population variance in ranked set sampling using auxiliary information, and that is through some suggested estimators. The bias and mean squared error of the proposed estimators are obtained up to the first degree of approximation. An asymptotic optimum estimator is identified with its approximate mean squared error (MSE) formula. An estimator based on "estimated optimum values" is also investigated. Some special cases of these estimators are considered and compared using computer simulation. Finally, we showed how to extend the proposed estimator if more than one auxiliary variable is available.

Keywords: Bias, Mean square error, Auxiliary variables, Relative Efficiency, Ranked set sampling

1. Introduction

In his outstanding efforts to create an estimator that would be more effective for estimating the produce of Australia's vast grazing areas, McIntyre [9] was the first to propose the idea of Ranked Set Sampling (RSS). The RSS concept appears to have gained traction only after Halls and Dell [7] used it to estimate the production of animal fodder in pine woods, and they were the first to

*Corresponding author
Email addresses: rikan.al_rahman1101@coadec.uobaghdad.edu.iq (Rikan A. Ahmed), saja@coadec.uobaghdad.edu.iq (Saja Mohammad Hussein)

Received: March 2021 *Accepted:* June 2021

coin the name Ranked Set Sampling. The two scientists who first provided mathematical proofs for this type of sampling, Takahase and Wakimoto [15], proved that the arithmetic mean of this type of sampling is an unbiased estimator of the population's arithmetic mean, and that the variance is less than the variance of the arithmetic mean of a Simple Random Sample (SRS), assuming perfect ranking of the elements. Dell & Clutter [6] got to the same conclusion as above, but without the requirement that the elements are in perfect order, implying whether or not there are ranking errors. Stokes [13] suggested using the auxiliary variable to estimate the ranks of the variable we want to study (the main variable), which is difficult to rank units with the naked eye. Stokes [14] proposed a variance estimator based on RSS data that is an asymptotic unbiased estimate of the population variance. Yu et al. [16] investigated a set of unbiased variance estimates for a normal population in the parameter case. MacEachern et al. [8] provide a new unbiased estimator for Population variance of ranked set sample data, showing that it is more efficient than Stokes' estimator, as well as conventional sample variance from a simple random sample. Perron and Sinha [11] demonstrated that in the nonparametric context, it is possible to create a class from an unbiased quadratic for variance estimates in both the balanced and unbalanced cases. AL-Saleh and Samawi [1] the suggested estimators are compared to other existing estimators based on bivariate simple random sample and application to the bivariate normal distribution, using the bivariate ranked set sampling process to estimate the correlation coefficient between two variables. Sengupta and Mukhuti [12] proposed some unbiased estimators of the variance of an exponential distribution using a ranked set sample, and all the proposed estimators are better than the non-parametric minimum variance unbiased quadratic estimator based on a balanced ranked set sample as well as the uniformly minimum variance unbiased estimator based on a simple random sampling. Chen and Lim [5] propose a plug-in estimator for each variance group, which is more efficient than the empirical group for estimating variances of strata in ranked set sampling. Biswas et al. [2] studied the two alternative variance estimation approaches in a ranked set sample under a finite population framework using the Jackknife method. Ozturk and Demirel [10] created a population variance estimator based on a multi-ranker partially rank-ordered set design, and a simulation analysis offers experimental proof for the proposed estimator's efficiency. Zamanzade and Al-Omari [17] used Monte Carlo simulation to compare empirical mean and variance estimators based on new ranked set sampling to their counterparts in ranked set sampling and simple random sampling. When ranked set sampling design is utilized under ranking criteria instead of using the process of a simple random sample, Bouza et al. [4] propose a model for the estimation of the variance of sensitive variables in a randomized response procedure. Biswas et al. [3] offer two unbiased variance estimations of ranked set sampling estimator under finite population framework using the bootstrap method and compare the efficiency of these proposed variance estimation strategies using a simulated study and real data application.

2. Ranked set sampling (RSS)

If we have the two variables Y and X representing the main and auxiliary variable respectively, and to select a Ranked Set Sample(RSS) in the form of rank pairs with size $n = mr$, we first selecting a simple random sample it size m^2 is drawn from the population, and then the selected sample is divided into m of sets, each set with size m represent a random sample. Each sample is arranged based on one of the two variables, we assume that the arrangement depends on the ranks of the auxiliary variable X and the main variable Y follows in the order, then accordingly the auxiliary variable is perfect ranking, and the main variable contains a possible error in the ranking say imperfect ranking, and the (RSS) items are selected according to the following. From the first-ranked sample, we choose the first observation, and from the second-ranked sample, we choose the second observation and continue to the last sample m where we choose the last observation, and in this case, the first cycle is completed, then we repeat this process r times to get the required sample size n from the ranked sample, which is $n = rm$. The RSS procedure can be summarized as follows:

1. First, we select bivariate sample units with size m^2 from the population are be random.

2. Divided the selected sample m^2 as possible into m of sets each with size m.

3. Ranked each sample with respect to one of the main variables Y or the auxiliary variable X. And upon it, we assume that judgment perfect ranking is done based on the variable X, while the imperfect ranking of variable Y.

4. The first sample is taken the measured of a unit from with the smallest rank of variable X, together with variable Y associated with the smallest rank of variable X. From the second sample, the second smallest rank of variable X associated with the variable Y is measured, the process is continued until the m^{th} sample, since the highest rank of variable X associated with the variable, Y is measured.

5. To reach the required sample size n, we repeat steps one to four r times until we get the size $n = mr$.

Note we will use some simple formulas to differentiate between perfect and imperfect ranking. In the case of perfect ranking, we will put the rank index in parentheses, say(i). Otherwise, in imperfect ranking, we will put the rank index in brackets, say$[i]$.
The elements of (RSS) for the auxiliary variable X and the main variable Y from one cycle can be described as follows:

$$\left\{\begin{matrix} X_{11} & X_{21} & \cdots & X_{m1} \\ X_{12} & X_{22} & \cdots & X_{m2} \\ \vdots & \vdots & \cdots & \vdots \\ X_{1m} & X_{2m} & \cdots & X_{mm} \end{matrix}\right\} \Rightarrow \left\{\begin{matrix} X_{(1)1} & X_{(2)1} & \cdots & X_{(m)1} \\ X_{(1)2} & X_{(2)2} & \cdots & X_{(m)2} \\ \vdots & \vdots & \cdots & \vdots \\ X_{(1)m} & X_{(2)m} & \cdots & X_{(m)m} \end{matrix}\right\};$$

$$\left\{\begin{matrix} Y_{11} & Y_{21} & \cdots & Y_{m1} \\ Y_{12} & Y_{22} & \cdots & Y_{m2} \\ \vdots & \vdots & \cdots & \vdots \\ Y_{1m} & Y_{2m} & \cdots & Y_{mm} \end{matrix}\right\} \Rightarrow \left\{\begin{matrix} Y_{[1]1} & Y_{[2]1} & \cdots & Y_{[m]1} \\ Y_{[1]2} & X_{[2]2} & \cdots & Y_{[m]2} \\ \vdots & \vdots & \cdots & \vdots \\ Y_{[1]m} & Y_{[2]m} & \cdots & Y_{[m]m} \end{matrix}\right\},$$

And upon it, the (RSS) units from one cycle for the auxiliary variable X and the main variable Y are respectively. $\left(X_{(1)1}, X_{(2)2}, \cdots X_{(m)m}\right)$ & $\left(Y_{[1]1}, Y_{[2]2}, \cdots Y_{[m]m}\right)$. And the binary pairs of the (RSS) units for both the auxiliary and main variables in the case r of cycles are defined as follows when the ranking is based on the auxiliary variable$X$.

$$\left\{\left(Y_{[1]j}, X_{(1)j}\right) \ , \ \left(Y_{[2]j}, X_{(2)j}\right), \ \ldots, \left(Y_{[m]j}, X_{(m)j}\right)\right\} \qquad j = 1, 2, \ \ldots, \ r$$

3. Some Population Variance estimators based on RSS

Assume that ($Y_{[i]j}$; $i = 1, 2, \cdots, m$ & $j = 1, 2, \cdots, r$) denote the ranked value of a unite of the jth cycle having the ith rank from the complete balanced Ranked Set Sample (RSS) with set size m and r cycles, and let the mean and variance of the ith judgment order statistic from a set of size m are respectively $\mu_{y(i)}$ & $\sigma^2_{y(i)}$. And assume that also Y_{ij} denote a Simple Random Sample (SRS)

of size $n = mr$ which this sample is *iid* random variables selected from the same population with mean and variance, respectively μ_y & σ_y^2 .Oftentimes, the overall variation between population units is extremely important. It is obvious that the parameter referred to is the population variance σ_y^2, from natural that the unbiased estimate of σ_y^2 in case of (SRS) of size n it is s_y^2 where

$$s_y^2=\frac{1}{\mathrm{n}-1}\sum_{\mathrm{j}=1}^{\mathrm{r}}\sum_{\mathrm{i}=1}^{\mathrm{m}}(\mathrm{Y_{ij}}-\overline{\mathrm{Y}}_{\mathrm{SRS}})^2 \quad ; \qquad \overline{\mathrm{Y}}_{\mathrm{SRS}} = \frac{1}{\mathrm{n}-1}\sum_{j=1}^{r}\sum_{i=1}^{m}Y_{ij}$$

In the case of (RSS), Stokes [14] proposed estimating the population variance σ_y^2, for one cycle, it is defined as follows.

$$\sigma_s^2=\frac{1}{\mathrm{n}-1}\sum_{\mathrm{j}=1}^{\mathrm{r}}\sum_{\mathrm{i}=1}^{\mathrm{m}}(\mathrm{Y}_{(i)\mathrm{j}}-\overline{\mathrm{Y}}_{\mathrm{RSS}})^2 \quad ; \quad \overline{\mathrm{Y}}_{\mathrm{RSS}}=\frac{1}{\mathrm{n}}\sum_{\mathrm{j}=1}^{\mathrm{r}}\sum_{\mathrm{i}=1}^{\mathrm{m}}\mathrm{Y}_{(i)\mathrm{j}}$$

She showed that her estimator is asymptotically unbiased regardless of the presence of ranking error, but this estimator owns a substantial amount of bias for finite sets and sample sizes. It is noticeable on the method of calculating the Stokes estimator that it relied on the well-known formula in calculating the sample variance, regardless of the information available from the design of the ranked set sample. MacEachern et al. [8] suggested developing to stokes estimator. This estimator for any set and cycle sizes is unbiased. Also, it is noted that the MacEachern estimator works better than the Stokes variance estimator for small sample sizes and the MacEachern estimator as follows.

$$\sigma_{Rss(y)}^2 = \frac{1}{2r^2m^2}\sum_{i\neq l}^{m}\sum_{j=1}^{r}\sum_{k=1}^{r}(Y_{(i)j}-Y_{(l)k})^2+\frac{1}{2r\,(r-1)\,m^2}\sum_{i=1}^{m}\sum_{j=1}^{r}\sum_{k=1}^{r}(Y_{(i)j}-Y_{(i)k})^2 \tag{3.1}$$

MacEachern studied the properties of the above estimator and proved that it is unbiased with respect to the population variance σ_y^2 of any number of the cycles, and estimated its variance according to the following formula.

$$\begin{aligned} V\left(\sigma_{Rss(y)}^2\right) =& \frac{1}{rm^2}\sum_{i=1}^{m}\mu_{y(i)4}+\frac{4}{rm^2}\sum_{i=1}^{m}\mu_{y(i)3}\tau_{y(i)}+\frac{4}{rm^2}\sum_{i=1}^{m}\sigma_{y(i)}^2\tau_{y(i)}^2 \\ &+\frac{4}{r^2m^2}\sum_{i<l}^{m}\sigma_{y(i)}^2\sigma_{y(l)}^2-\frac{m^2\,(r-1)-2}{r(r-1)\,m^4}\sum_{i=1}^{m}\sigma_{y(i)}^4 \end{aligned} \tag{3.2}$$

where

$$\mu_{y(i)a}=\sum_{i=1}^{m}(Y_{(i)}-\mu_{y(i)})^a \quad , \quad \tau_{y(i)}=\mu_{y(i)}-\mu_y \quad \& \quad \sigma_{y(i)}^2=Var(y_{(i)})$$

4. Estimators Definition and Properties

Infinite populations, the classical methods of estimation the population parameters are based on direct estimators, that means those which only use the main (study) variable Y. In another case, indirect estimation of the unknown population parameters can be used by employing auxiliary information to any auxiliary variable X (or more) associated with the main variable Y for the purpose of estimating it; the indirect estimates are usually better than direct estimates because of the ancillary information, which gives an increase inaccuracy. As an example of indirect methods, the ratio method, the regression method, etc. Now assume that the problem is to estimate the population variance σ_y^2 of the study variable Y by using the indirect methods from via the information provided by one or two auxiliary variables based onRSS. And suppose Y denote the study variable and auxiliary variable X whose population mean and variance are μ_y, μ_x and σ_y^2 , σ_x^2 respectively. Assuming that the mean and variance of the population of the auxiliary variables are known, and the jth unit of the population for the main variable and the auxiliary variable they respectively Y_j & X_j , and the ith order statistics of a sample of size m in the jth cycle of the variables $Y\&X$ is respectively represent by $Y_{(i)j}$ & $X_{(i)j}$ Based on a RSS of size $n = rm$ drawn from the population. We will now suggest a ratio estimator for population variance of the variable Y based on RSS, as shown below.

$$\widehat{\sigma}_a^2 = \sigma_{Rss[y]}^2 \ \left(\frac{\sigma_{Rss(x)}^2}{\sigma_x^2} \right)^a \tag{4.1}$$

Where $\sigma_{Rss[y]}^2$ has been defined in (3.1) and represents the usual estimator σ_y^2 , $\sigma_{Rss(x)}^2$ and σ_x^2 denote the sample and the population variance of the auxiliary variable X respectively, also, the estimated formula to $\sigma_{Rss(x)}^2$ is defined as in (3.1) after replacing the symbol X instead of Y , and a it is a real number will be chosen so that the variance of the estimator$\widehat{\sigma}_a^2$ is as small as possible.
It is cannot to ranked two or more variables data at the same time. Therefore, ranking one of the variables, say the auxiliary variable (x), and taking the corresponding values of other variables here be the study variable [y] is an option. So it was written the proposed ratio estimator with the above formula to indicate that the auxiliary variable is perfect ranking while the main variable is imperfect ranking.

4.1. Some Notions

In this research, the following definitions will be used to studying the properties of the proposed estimator. Let $e_0 = \frac{\widehat{\sigma}_{Rss_{[y]}}^2 - \sigma_y^2}{\sigma_y^2}$; $e_1 = \frac{\widehat{\sigma}_{Rss_{(x)}}^2 - \sigma_x^2}{\sigma_x^2}$ Then $E\left(e_0\right) = E\left(e_1\right) = 0$, $E\left(e_0^2\right) = \frac{V_y}{\sigma_y^4}$, $E\left(e_1^2\right) = \frac{V_x}{\sigma_x^4}$ & $E\left(e_0 e_1\right) = \frac{C_{yx}}{\sigma_y^2\ \sigma_x^2}$, where

$$\begin{aligned} V_y =& \frac{1}{rm^2} \sum_{i=1}^{m} \mu_{y[i]4} + \frac{4}{rm^2} \sum_{i=1}^{m} \mu_{y[i]3} \tau_{y[i]} + \frac{4}{rm^2} \sum_{i=1}^{m} \sigma_{y[i]}^2 \tau_{y[i]}^2 + \\ & \frac{4}{r^2 m^2} \sum_{i<l}^{m} \sigma_{y[i]}^2 \sigma_{y[l]}^2 - \frac{m^2\left(r-1\right)-2}{r(r-1)\, m^4} \sum_{i=1}^{m} \sigma_{y[i]}^4 \end{aligned} \tag{4.2}$$

$$\begin{aligned} V_x =& \frac{1}{rm^2} \sum_{i=1}^{m} \mu_{x(i)4} + \frac{4}{rm^2} \sum_{i=1}^{m} \mu_{x(i)3} \tau_{x(i)} + \frac{4}{rm^2} \sum_{i=1}^{m} \sigma_{x(i)}^2 \tau_{x(i)}^2 + \\ & \frac{4}{r^2 m^2} \sum_{i<l}^{m} \sigma_{x(i)}^2 \sigma_{x(l)}^2 - \frac{m^2\left(r-1\right)-2}{r(r-1)\, m^4} \sum_{i=1}^{m} \sigma_{x(i)}^4 \end{aligned} \tag{4.3}$$

$$\begin{aligned} C_{yx} =& \frac{1}{rm^2} \sum_{i=1}^{m} \mu_{yx[i]} + \frac{4}{rm^2} \sum_{i=1}^{m} \mu_{yx[i]3} \tau_{yx[i]} + \frac{4}{rm^2} \sum_{i=1}^{m} \sigma_{yx[i]} \tau_{yx[i]} + \\ & \frac{4}{r^2 m^2} \sum_{i<l}^{m} \sigma_{yx[i]} \sigma_{yx[l]} - \frac{m^2\left(r-1\right)-2}{r(r-1)\, m^4} \sum_{i=1}^{m} \sigma_{yx[i]} \end{aligned} \tag{4.4}$$

$$\mu_{y[i]a} = \sum_{i=1}^{m} \left(Y_{[i]} - \mu_{y[i]}\right)^{a}, \tau_{y[i]} = \mu_{y[i]} - \mu_{y} \quad , \quad \mu_{x(i)a} = \sum_{i=1}^{m} \left(X_{(i)} - \mu_{x(i)}\right)^{a},$$

$$\tau_{x(i)} = \mu_{x(i)} - \mu_{x}, \tau_{yx[i]} = (\mu_{y[i]} - \mu_{y})(\mu_{x(i)} - \mu_{x}) \ , \sigma_{yx[i]} = cov(Y_{[i]}, X_{(i)})$$

and $\mu_{yx[i]a} = \sum_{i=1}^{m} \left((Y_{[i]} - \mu_{y[i]}) \left(X_{(i)} - \mu_{x(i)}\right)\right)^{a}$

The estimator $\widehat{\sigma}_a^2$ will be rewritten as follows based on the given definitions.

$$\begin{aligned} \widehat{\sigma}_a^2 &= \sigma_y^2 \ (\ 1 + e_0\ )\ (\ 1 + e_1\ )^{a} \\ &= \sigma_y^2 (1 + e_0 + ae_1 + ae_0 e_1 + \frac{a\,(a-1)}{2}\ e_1^2 + \cdots .\ ) \end{aligned}$$

We will discuss the properties of the proposed estimator represented by the derivation of the bias, and mean square error expressions are considered up to the terms of order n^{-1} only, are, therefore.

$$E\left(\widehat{\sigma}_a^2\right) = \sigma_y^2 [1 + a\ \frac{C_{yx}}{\sigma_y^2\ \sigma_x^2} + \frac{a\,(a-1)}{2}\ \frac{V_x}{\sigma_x^4}].$$

The bias to $\widehat{\sigma}_a^2$ is given by the following formula.

$$Bias\ \left(\widehat{\sigma}_a^2\right) = \ a\ \frac{C_{yx}}{\sigma_x^2} + \frac{a\,(a-1)}{2}\ \frac{V_x\ \sigma_y^2}{\sigma_x^4}. \tag{4.5}$$

The expression for the mean square error of $\widehat{\sigma}_1^2$ define as follows.

$$MSE\left(\widehat{\sigma}_a^2\right) = E\left(\widehat{\sigma}_a^2 - \sigma_y^2\right)^2 = \ V_y + a^2 \frac{V_x\ \sigma_y^4}{\sigma_x^4} + 2a \frac{C_{yx}\sigma_y^2}{\sigma_x^2}. \tag{4.6}$$

The best value of a that minimizes the mean square error of $\widehat{\sigma}_a^2$ up to the order n^{-1} of can be easily proved as

$$\hat{a} \quad = -\frac{C_{yx}\sigma_x^2}{V_x \sigma_y^2}. \tag{4.7}$$

Therefore the minimum MSE for $\widehat{\sigma}_a^2$ is

$$MSE_{opt}\left(\widehat{\sigma}_{\hat{a}}^2\right) = V_y - \frac{C_{yx}^2}{V_x}. \tag{4.8}$$

While the smallest bias to $\widehat{\sigma}_a^2$ is

$$Bias_{opt}\ \left(\widehat{\sigma}_{\hat{a}}^2\right) = \ \frac{C_{yx}}{2\ \sigma_x^2} - \frac{C_{yx}^2}{2V_x\ \sigma_y^2} \tag{4.9}$$

4.2. Some unique characteristics of a

The suggested estimator in equation (4.1) contains a set of population variance estimators that can be derived by changing the numerical value of a and as follows: when $a = -1$, it is referred to as the ratio estimator of σ_y^2, and its mathematical formula is

$$\widehat{\sigma}_r^2 = \sigma_{Rss[y]}^2 \frac{\sigma_x^2}{\sigma_{Rss(x)}^2} = \hat{R}_{RSS} \sigma_x^2 \tag{4.10}$$

The bias and MSE of this estimator up to the order of n^{-1} are then determined using well-known formulas, respectively.

$$Bias\ \left(\widehat{\sigma}_r^2\right) = Bias\left(\hat{R}_{RSS}\right)\ \sigma_x^2 = \left(\frac{V_x}{\sigma_x^2} - \frac{C_{yx}}{\sigma_y^2}\right) R \tag{4.11}$$

$$MSE\left(\widehat{\sigma}_r^2\right) = MSE\left(\hat{R}_{RSS}\right)\sigma_x^4 = V_y + V_x R^2 - 2C_{yx}\ R, \tag{4.12}$$

where $R = \frac{\sigma_y^2}{\sigma_x^2}$ and $\hat{R}_{RSS} = \frac{\sigma^2_{Rss[y]}}{\sigma^2_{Rss(x)}}$.
If $a = 1$, the product estimator of σ_y^2 is called, and its mathematical formula is

$$\widehat{\sigma}_p^2 = \sigma^2_{Rss[y]}\quad \frac{\sigma^2_{Rss(x)}}{\sigma_x^2} = \frac{\hat{P}_{RSS}}{\sigma_x^2} \tag{4.13}$$

And the estimator's bias, as well as the MSE up to the order of n^{-1} , are, respectively.

$$Bias\ \left(\widehat{\sigma}_p^2\right) = \quad Bias\left(\hat{P}_{RSS}\right) /\ \sigma_x^2 = \frac{C_{yx}}{\sigma_x^2} \tag{4.14}$$

$$MSE\left(\widehat{\sigma}_p^2\right) = \ MSE\left(\hat{P}_{RSS}\right)\ /\sigma_x^4 = \ V_y + V_x\ R^2 + 2C_{yx}\ R, \tag{4.15}$$

where $P = \sigma_y^2\ \sigma_x^2$ and $\hat{P}_{RSS} = \sigma^2_{Rss[y]}\sigma^2_{Rss(x)}$.

If $a = 0$, it's the standard per-unit variance estimator σ_y^2 based on RSS, which is defined in equation (3.1). If the value of a equals $1 - 2\frac{C_{yx}}{V_x R}$, however, the bias amount of the estimator $\widehat{\sigma}_a^2$ becomes zero, indicating that it is unbiased and its variance matching the same the mean square error of the ratio estimator. In which $(\widehat{\sigma}_a^2) = MSE\left(\widehat{\sigma}_r^2\right) = V_y + V_x\ R^2 - 2C_{yx}\ R$. Finally, when the value of a equals $-1-2\frac{C_{yx}}{V_x R}$, the estimator $\widehat{\sigma}_a^2$ is still biased, but it's mean squared error is the same as the product estimator's mean squared error. Which $MSE\left(\widehat{\sigma}_a^2\right) = MSE\left(\widehat{\sigma}_p^2\right) = V_y + V_x\ R^2 + 2C_{yx}\ R$.

4.3. Comparing Estimators

The proposed estimator is compared to some well-known estimators. The bias and mean square error will be compared up to the order of n^{-1}. For bias we compare equation (4.9) to equations (4.11) and (4.14), we can see that the estimator $\widehat{\sigma}_{\hat{a}}^2$ has a lesser bias than the ratio and product estimators, as illustrated below, respectively,

$$\left|Bias_{opt}\ \left(\widehat{\sigma}_{\hat{a}}^2\right)\right| \ \le\ \left|Bias\ \left(\widehat{\sigma}_r^2\right)\right| \quad \text{if and only if} \quad \left|\frac{C_{yx}\sigma_x^2}{V_x\sigma_y^2}\right| \ \le 2$$

and

$$\left|Bias_{opt}\ \left(\widehat{\sigma}_{\hat{a}}^2\right)\right| \ \le\ \left|Bias\ \left(\widehat{\sigma}_p^2\right)\right| \quad \text{if and only if} \quad \left|1 - \frac{C_{yx}\sigma_x^2}{V_x\sigma_y^2}\right| \ \le 2.$$

In terms of MSE, we'll first compare equation (4.6) to equation (4.12), and we'll see that $MSE\left(\widehat{\sigma}_a^2\right) \le MSE\left(\widehat{\sigma}_r^2\right)\ iff\ (a-1)(a-1+2\frac{C_{yx}}{V_x R}) \le 0$, this signifies that a lies between the two values -1 and $1 - 2\frac{C_{yx}}{V_x R}$. After comparing equation (4.6) to equation (4.15), we note $MSE\left(\widehat{\sigma}_a^2\right) \le MSE\left(\widehat{\sigma}_p^2\right)\ iff\ (a-1)(a+1+2\frac{C_{yx}}{V_x R}) \le 0$, this also demonstrates that a is located between

1 and $-\left(1+2\frac{C_{yx}}{V_xR}\right)$. As a result, the MSE of the estimator $\widehat{\sigma}_a^2$ remains lower than the variance estimator σ_y^2 based on RSS as the value of $|a-\hat{a}\ |$ decreases and it is minimum when $|a-\widehat{a}\ |=0$. In addition, whenever $|a-\hat{a}\ |$ is less than $\left|1-\frac{C_{yx}}{V_xR}\right|$, the MSE of the estimator $\widehat{\sigma}_a^2$ is less than the MSE of the ratio estimator and less than the MSE of the product estimator whenever $|a-\hat{a}\ |$ is less than $\left|1+\frac{C_{yx}}{V_xR}\right|$. To gain a better idea of how the estimator $\widehat{\sigma}_{\hat{a}}^2$ works, consider the following basic example. If we suppose $V_y=V_x$ And $R=0.7$, and the correlation coefficient between the main and auxiliary variables is positive; we should use the ratio estimator to compare. And so be $a=-1.4$, The efficiency factor $MSE\left(\widehat{\sigma}_r^2\right)/MSE\left(\widehat{\sigma}_1^2\right)$ Is greater than one for a with a value between $(-1$ and $-1.85714)$, equal to one for a with a value in -1 and -1.85714, and equal to 225 for a with a value of -1.4. In terms of the amount of bias in the estimators, we can see that the ratio of the biases $\left(Bias\left(\widehat{\sigma}_r^2\right)/Bias\ \left(\widehat{\sigma}_1^2\right)\right)$ will range from (1 to ∞) when a is between $[-1,\ -1.85714]$ and equal to 15.555 when $a=-1.4$, and we can also see the ratio of the relative biases when $a=-1.4$ is 1.037.

5. Generalization in the case of several auxiliary variables

If we have data on more than one auxiliary variable, say P, we may utilize these variables to estimate the population variance σ_y^2 by using a linear function of P estimators of the form (4.1), computed independently for each auxiliary variable. If the P auxiliary variables are $X_1,\ X_2,\ \cdots,\ X_P$ then an RSS-based estimator of the population variance of Y is defined as,

$$\widehat{\sigma}_g^2=\sigma_{Rss[y]}^2\ \left[\ \left(\ \frac{\sigma_{Rss[x_1]}^2}{\sigma_{x_1}^2}\ \right)^{a_1}\ \left(\ \frac{\sigma_{Rss[x_2]}^2}{\sigma_{x_2}^2}\ \right)^{a_2}\ \cdots\ \left(\ \frac{\sigma_{Rss(x_i)}^2}{\sigma_{x_i}^2}\ \right)^{a_i}\ \cdots\ \left(\ \frac{\sigma_{Rss[x_P]}^2}{\sigma_{x_P}^2}\ \right)^{a_P}\right] \tag{5.1}$$

$\widehat{\sigma}_g^2$ are known as multiple ratio estimators for population variance of the variable Y based on RSS with ranking dependant, on the variable $(X_i \qquad i=1,\cdots,P)$. where $\sigma_{Rss[y]}^2$ denotes the usual estimator σ_y^2 of the main variable Y based on RSS, and $\sigma_{Rss[x_i]}^2$ $i=1,\cdots,P$ are usual estimators of $\sigma_{x_i}^2$ $i=1,\cdots,P$ respectively based on RSS , and in a first degree of approximating, it may be demonstrated that.

$$Bias\ \left(\widehat{\sigma}_g^2\right)=\sigma_y^2\left\{\quad \sum_{i=1}^{P} a_i\frac{C_{yx_i}}{\sigma_{x_i}^2\sigma_y^2}+\sum_{i\neq j}^{P} a_ia_j\frac{C_{x_ix_j}}{\sigma_{x_i}^2\sigma_{x_j}^2}+\sum_{i=1}^{P}\frac{a_i\left(a_i-1\right)}{2}\ \frac{V_{x_i}}{\sigma_{x_i}^4}\right\} \tag{5.2}$$

in the same manner

$$MSE\left(\widehat{\sigma}_g^2\right)=\ \sigma_y^4\ \left\{\ \frac{V_y}{\sigma_y^4}+\sum_{i=1}^{P}\sum_{j=1}^{P} a_ia_j\ \frac{C_{x_ix_j}}{\sigma_{x_i}^2\sigma_{x_j}^2}+2\ \sum_{i=1}^{P} a_i\frac{C_{yx_i}}{\sigma_{x_i}^2\sigma_y^2}\ \right\} \tag{5.3}$$

where C_{yx_i} & $C_{x_ix_j}$ previously defined, and the following variables are used to compute the bias and mean squared error.

$$e_0=\frac{\widehat{\sigma}_{Rss_{[y]}}^2-\sigma_y^2}{\sigma_y^2}\ ;\ e_1=\frac{\widehat{\sigma}_{Rss_{[x_1]}}^2-\sigma_{x_1}^2}{\sigma_{x_1}^2}\ ,\ e_2=\frac{\widehat{\sigma}_{Rss_{[x_2]}}^2-\sigma_{x_2}^2}{\sigma_{x_2}^2}\ ,\ \cdots,\ e_i=\frac{\sigma_{Rss(x_i)}^2-\sigma_{x_i}^2}{\sigma_{x_i}^2}$$

$$,\ \cdots,\ e_P=\frac{\sigma_{Rss[x_P]}^2-\sigma_{x_P}^2}{\sigma_{x_P}^2}$$

We must now identify the optimal values of a_i to minimize the MSE for estimator $\widehat{\sigma}_g^2$. Therefore, equation (5.3) will be reformulated in another form.

$$MSE\left(\widehat{\sigma}_g^2\right) = \sigma_y^4 \ \{ \ \frac{V_y}{\sigma_y^4} + \acute{a}Ma + 2\acute{a}b \ \} \tag{5.4}$$

where $= a_{(P\times 1)} = (a_1, a_2, \ \cdots, a_P)^{\acute{}}$, $M = M_{(P\times P)} = (m_{ij})$ with $m_{ij} = \frac{C_{x_i x_j}}{\sigma_{x_i}^2 \sigma_{x_j}^2}$; $i, j = 1, 2, \cdots, P$ and $b = b_{(P\times 1)} = (b_i)$ with $b_i = \frac{C_{yx_i}}{\sigma_y^2 \sigma_{x_i}^2}$; $i = 1, 2, \cdots, P$

Now, the deriving of the equation (5.4) with respect to a yields the best values of a that minimize $MSE\left(\widehat{\sigma}_g^2\right)$ .And it is given as follows.

$$a_{opt} = -M^{-1}b \tag{5.5}$$

Then, given the estimator $\widehat{\sigma}_g^2$, the optimum mean square error and optimum bias are, respectively

$$MSE_{opt}\left(\widehat{\sigma}_g^2\right) = \sigma_y^4 \ \{ \ \frac{V_y}{\sigma_y^4} - \acute{b}Mb \ \} \tag{5.6}$$

$$Bias_{opt}\left(\widehat{\sigma}_g^2\right) = \sigma_y^2 \ \{\acute{a}Ma - 2\acute{a}b \ \} \tag{5.7}$$

For $P = 1$, in other words, if you just utilize one auxiliary variable, then $\widehat{\sigma}_g^2 = \widehat{\sigma}_a^2$, $a_{opt} = \hat{a}$, $MSE_{opt}\left(\widehat{\sigma}_g^2\right) = MSE_{opt}\left(\widehat{\sigma}_a^2\right)$ and $Bias_{opt}\left(\widehat{\sigma}_g^2\right) = Bias_{opt}\left(\widehat{\sigma}_a^2\right)$.

If $= 2$, to put it another way, if we only utilize two auxiliary variables, such as $X_1 and\ X_2$, the first auxiliary variable is likely to have a perfect ranking denoted by symbol(X_1). then the second auxiliary variable and the main variable are likely to have for them imperfect ranking denoted by symbol $[X_2]$ *and* $[Y]$ respectively, we will be reformulated estimator $\widehat{\sigma}_g^2$ as follows.

$$\widehat{\sigma}_{a_1 a_2}^2 = \sigma_{Rss[y]}^2 \ \ [\ (\ \frac{\sigma_{Rss(x_1)}^2}{\sigma_{x_1}^2} \ )^{a_1} \ (\ \frac{\sigma_{Rss[x_2]}^2}{\sigma_{x_2}^2} \ )^{a_2} \ ] \tag{5.8}$$

We'll need to define some matrices to achieve the best values for $a_1$and a_2, to get best MSE**,** for the estimator $\widehat{\sigma}_{a_1 a_2}^2$. By using the equation (5.5) and (5.6).

Let's look at the matrices

$$M = \begin{bmatrix} \frac{C_{x_1 x_1}}{\sigma_{x_1}^2 \sigma_{x_1}^2} & \frac{C_{x_1 x_2}}{\sigma_{x_1}^2 \sigma_{x_2}^2} \\ \frac{C_{x_1 x_2}}{\sigma_{x_1}^2 \sigma_{x_2}^2} & \frac{C_{x_2 x_2}}{\sigma_{x_2}^2 \sigma_{x_2}^2} \end{bmatrix} \quad , \quad b = \begin{bmatrix} \frac{C_{yx_1}}{\sigma_y^2 \sigma_{x_1}^2} & \frac{C_{yx_2}}{\sigma_y^2 \sigma_{x_2}^2} \end{bmatrix} \quad and \quad a = \begin{bmatrix} a_1 & a_2 \end{bmatrix} \text{ then}$$

$$a_{1opt} = \frac{(C_{yx_2} C_{x_1 x_2} - C_{yx_1} V_2)\sigma_{x_1}^2}{(\ V_1 V_2 - \ C_{x_1 x_2}^2)\sigma_y^2} \quad , \quad a_{2opt} = \frac{(C_{yx_1} C_{x_1 x_2} - C_{yx_2} V_1)\sigma_{x_2}^2}{(\ V_1 V_2 - \ C_{x_1 x_2}^2)\sigma_y^2}$$

$$MES_{opt}\left(\widehat{\sigma}_{a_1 a_2}^2\right) = V_0 - \ \frac{(C_{yx_1}^2 V_2 + C_{yx_2}^2 V_1 - 2C_{yx_1} \ C_{yx_2} \ C_{x_1 x_2})}{(\ V_1 V_2 - \ C_{x_1 x_2}^2)}$$

6. Study of the simulation

A hypothetical population was formed for the study's main and auxiliary variable based on the bivariate normal distribution having known parameters ($\mu_y = 2, \ \mu_x = 1, \sigma_y^2 = \sigma_x^2 = 1, \rho = \pm 0.95, \pm 0.75, 0.35$), and ranked set samples were created consisting of sets of varied sizes ($m = 4, 5, 6$)

and a defined number of cycles for each volume equal to$(r = 3)$, We estimated the bias, the mean square error $MSE(.)$of the proposed estimator $\widehat{\sigma}_a^2$, the ratio estimator$\widehat{\sigma}_r^2$, and the product $\widehat{\sigma}_p^2$ for each of the different set sizes, as well as the values of the correlation coefficient for the three cycles, based on the auxiliary variable X. The relative efficiency criterion $eff(.)$ for the estimators above was determined with the estimated variance of the main variable $\sigma^2_{Rss[y]}$ based on RSS to find the best estimators where $eff\left(\widehat{\sigma}_i^2\right) = \frac{Var(\sigma^2_{Rss[y]})}{MSE(\widehat{\sigma}_i^2)} \quad i = a, r, p$ .The experiment was repeated 10,000 times, and the simulation results are shown in the tables (1 and 2). From Tables (1-2) can be seen the following.

For the various situations of $m\ and\ \rho$ in estimating the population variance concerning estimators, a gain in efficiency is obtained by employing $\widehat{\text{o}}_a^2$.

The efficiency of $\widehat{\text{o}}_a^2$ increases as the sample size n increases for fixed $m\ and\ \rho$. The efficiency of $\widehat{\text{o}}_a^2$ increases as the sample size n increases for any ρ. When comparing the bias of the generated estimators for different values of $n\ and\ \rho$, we see that the proposed estimator $\widehat{\text{o}}_a^2$ has the most negligible bias. The findings in tables (1-2) show that the suggested estimator works well irrespectively of whether the value of ρ is positive or negative, unlike the other estimators that are affected by the value of ρ. For example, the ratio estimator only works when $\rho > 0$, while the product estimator only works when $\rho < 0$.

7. Conclusions

The problem of estimating population variance using a known auxiliary variable in the situation of ranked set sampling is discussed in this study, and a novel ratio estimator is proposed to estimate the variance. We have shown that the suggested estimator's properties become more efficient than other estimators for the first degree of approximation, and its bias is negligible. The simulation study showed that the type of relationship between the main and auxiliary variables does not affect the efficiency of the proposed estimator, unlike other estimators that are affected by this relationship, and that the estimator works well with any sample size.

References

[1] M.F. Al-Saleh and H.M. Samawi, *Estimation of the correlation coefficient using bivaria- te ranked set sampling with application to the bivariate normal distribution*, Commun. Statist. Theory Meth. 34(4) (2005) 875–889.

[2] A. Biswas, T. Ahmadand and A. Rai, *Variance estimation using Jackknife method in ranked set samplig under finite population framework*, J. Indian Soc. Agric. Statist. 67(3) (2013) 345–353.

[3] A. Biswas, A. Rai and T. Ahmad, *Rescaling bootstrap technique for variance estimation for ranked set samples in finite population*, Commun. Statist. Simul. Comput. 49(10) (2020) 2704–2718.

[4] C.N.Bouza, J.F. García, G.K. Vishwakarma and S.M. Zeeshan, *Variance estimation of persons infected with AIDS under ranked set sampling*, Clin. Epidemiol. Global Health 7(1) (2017) 17–21.

[5] M.Chen and J.Lim, *Estimating variances of strata in ranked set sampling*, J. Stat.Plann.Inference, 141(8)(2011) 2513-2518.

[6] T.R. Dell and J.L. Clutter, *Ranked set sampling theory with order statistics background*,Biometrics,(1972) 545-555.

[7] L.K. Halls and T.R. Dell, *Trial of ranked-set sampling for forage yields*, Forest Sci. 12(1) (1966) 22–26.

[8] S.N. MacEachern, Ö.Öztürk , D.A. Wolfe and G.V.Stark, *A new ranked set sample estimator of variance*, J. Royal Stat. Soc. Series B (Stat. Method.) 64(2) (2002) 177–188.

[9] G.A. McIntyre, *A method for unbiased selective sampling, using ranked sets*, Aust. J. Agric.Res. 3(4) (1952)385–390.

[10] O. Ozturk and N. Demirel, *Estimation of population variance from multi-ranker ranked set sampling designs*, Commun. Stat. Simul. Comput. 45(10) (2014) 3568–3583.

Table 1: The efficiency of $\widehat{\sigma}_i^2$ $i = a.r.p$ *with respect to* $\sigma^2_{Rss[y]}$ and MSE , as well as the bias values of the three estimators, for $n = 12, 15, 18$ and $\rho > 0$

ρ	r	m	$estimator$	MSE	eff	$Bias$
0.95	3	4	$\sigma^2_{Rss[y]}$	1773.09		
			$\widehat{\sigma}_a^2$	886.541	2.00001	2.85688×10^{-6}
			$\widehat{\sigma}_r^2$	1819.99	0.974229	23.502
			$\widehat{\sigma}_p^2$	1820.22	0.974107	0.0286036
		5	$\sigma^2_{Rss[y]}$	581.478		
			$\widehat{\sigma}_a^2$	290.739	2.	8.92416×10^{-7}
			$\widehat{\sigma}_r^2$	537.139	1.08255	7.6871
			$\widehat{\sigma}_p^2$	596.892	0.974176	0.00893454
		6	$\sigma^2_{Rss[y]}$	142.953		
			$\widehat{\sigma}_a^2$	71.4766	1.999	4.05142×10^{-7}
			$\widehat{\sigma}_r^2$	134.916	1.05958	3.48371
			$\widehat{\sigma}_p^2$	149.939	0.953411	0.00405614
0.75	3	4	$\sigma^2_{Rss[y]}$	1796.76		
			$\widehat{\sigma}_a^2$	1347.57	1.33334	1.9608×10^{-6}
			$\widehat{\sigma}_r^2$	1672.66	1.07419	3.09118×10^{-3}
			$\widehat{\sigma}_p^2$	1858.67	0.966691	0.0196205
		5	$\sigma^2_{Rss[y]}$	1094.1		
			$\widehat{\sigma}_a^2$	957.334	1.14286	6.01129×10^{-7}
			$\widehat{\sigma}_r^2$	1009.66	1.08363	0.00138835
			$\widehat{\sigma}_p^2$	1121.89	0.975226	0.0060139
		6	$\sigma^2_{Rss[y]}$	815.049		
			$\widehat{\sigma}_a^2$	692.791	1.17647	2.76454×10^{-7}
			$\widehat{\sigma}_r^2$	747.591	1.09023	0.000780882
			$\widehat{\sigma}_p^2$	830.679	0.981184	0.00276552
0.35	3	4	$\sigma^2_{Rss[y]}$	2648.7		
			$\widehat{\sigma}_a^2$	1615.7	1.63934	6.1456×10^{-7}
			$\widehat{\sigma}_r^2$	2459.63	1.07687	0.00421257
			$\widehat{\sigma}_p^2$	2732.97	0.969162	0.0061465
		5	$\sigma^2_{Rss[y]}$	1173.87		
			$\widehat{\sigma}_a^2$	689.649	1.70213	2.35627×10^{-7}
			$\widehat{\sigma}_r^2$	1099.69	1.06746	0.00240055
			$\widehat{\sigma}_p^2$	1221.89	0.960699	0.0023565
		6	$\sigma^2_{Rss[y]}$	838.649		
			$\widehat{\sigma}_a^2$	511.576	1.63934	1.10339×10^{-7}
			$\widehat{\sigma}_r^2$	782.231	1.07213	0.001525
			$\widehat{\sigma}_p^2$	869.154	0.964903	0.00110347

Table 2: The efficiency of $\widehat{\sigma}_i^2 \quad i = a.r.p$ *with respect to* $\sigma^2_{Rss[y]}$ *and* MSE*, as well as the bias values of the three estimators, for* $n = 12, 15, 18$ *and* $\rho < 0$

ρ	r	m	*estimator*	MSE	eff	$Bias$
-0.95	3	4	$\sigma^2_{Rss[y]}$	1769.12		
			$\widehat{\sigma}^2_a$	928.787	1.90477	2.35714×10^{-6}
			$\widehat{\sigma}^2_r$	1815.81	0.974287	0.0023387
			$\widehat{\sigma}^2_p$	1725.2	1.02546	0.0235952
		5	$\sigma^2_{Rss[y]}$	584.149		
			$\widehat{\sigma}^2_a$	379.697	1.53846	5.26923×10^{-7}
			$\widehat{\sigma}^2_r$	601.113	0.971779	0.000849134
			$\widehat{\sigma}^2_p$	571.098	1.02285	0.0052725
		6	$\sigma^2_{Rss[y]}$	421.731		
			$\widehat{\sigma}^2_a$	316.298	1.33333	1.27745×10^{-7}
			$\widehat{\sigma}^2_r$	431.613	0.977103	0.000494353
			$\widehat{\sigma}^2_p$	410.042	1.02851	0.00127778
-0.75	3	4	$\sigma^2_{Rss[y]}$	1752.31		
			$\widehat{\sigma}^2_a$	1314.23	1.33333	1.46272×10^{-6}
			$\widehat{\sigma}^2_r$	1812.76	0.966651	0.00302529
			$\widehat{\sigma}^2_p$	1722.24	1.01746	0.0146343
		5	$\sigma^2_{Rss[y]}$	1148.59		
			$\widehat{\sigma}^2_a$	1027.99	1.11732	3.20294×10^{-7}
			$\widehat{\sigma}^2_r$	1177.17	0.975724	0.00142944
			$\widehat{\sigma}^2_p$	1118.33	1.02706	0.00320366
		6	$\sigma^2_{Rss[y]}$	859.461		
			$\widehat{\sigma}^2_a$	773.515	1.11111	6.86997×10^{-8}
			$\widehat{\sigma}^2_r$	876.971	0.980034	0.000875611
			$\widehat{\sigma}^2_p$	833.128	1.03161	0.000687051
-0.35	3	4	$\sigma^2_{Rss[y]}$	2964.25		
			$\widehat{\sigma}^2_a$	1971.23	1.50376	4.20697×10^{-7}
			$\widehat{\sigma}^2_r$	3048.68	0.972305	0.00422235
			$\widehat{\sigma}^2_p$	2896.28	1.02347	0.00420739
		5	$\sigma^2_{Rss[y]}$	1206.75		
			$\widehat{\sigma}^2_a$	802.486	1.50376	9.19648×10^{-8}
			$\widehat{\sigma}^2_r$	1254.12	0.962228	0.00236869
			$\widehat{\sigma}^2_p$	1191.42	1.01287	0.000919683
		6	$\sigma^2_{Rss[y]}$	912.049		
			$\widehat{\sigma}^2_a$	506.187	1.8018	1.087701×10^{-8}
			$\widehat{\sigma}^2_r$	943.398	0.966771	0.00156745
			$\widehat{\sigma}^2_p$	896.229	1.01765	0.000108702

[11] F. Perron and B.K. Sinha, *Estimation of variance based on a ranked set sample*, J. Statist. Plann. Infer. 120(1-2) (2004) 21–28.

[12] S. Sengupta and S. Mukhuti, *Unbiased variance estimation in a simple exponential population using ranked set samples*, J. Statist. Plann. Infer. 136(4) (2006) 1526–1553.

[13] S. Stokes, *Ranked set sampling with concomitant variables*, Commun. Statist. Theory Meth. 6(12) (1977) 1207–1211.

[14] S.L. Stokes, *Estimation of variance using judgment ordered ranked set samples*, Biometrics (1980) 35–42.

[15] K. Takahasi and K. Wakimoto, *On unbiased estimates of the population mean based on the sample stratified by means of ordering*, Ann. Inst. Statist. Math. 20(1) (1968) 1–31.

[16] L.H. Yu, P.K.Lam and B.M.K. Sinha, *Estimation of normal variance based on balanced and unbalanced ranked set samples*, Environ. Ecol. Statist. 6(1) (1999) 23–46.

[17] E. Zamanzade and A.I.Al-Omari, *New ranked set sampling for estimating the populat-ion mean and variance*,Hacettepe J. Math. Statist. 45(6) (2016) 1891–1905.