PROBABILITY PROPORTIONAL TO SIZE (PPS) METHOD TO ENHANCE EFFICIENCY OF ESTIMATOR IN TWO STAGE SAMPLING

Olayiwola O. M., Apantaku F. S, Wale-Orojo O. A., Ajayi A. O., Konku A. M.

Department of Statistics, College of Physical Sciences,
Federal University of Agriculture, Abeokuta, Nigeria

Corresponding Author: Olayiwola O. M., olayiwolaom@funaab.edu.ng

ABSTRACT: This study focused on application of Probability Proportional to Size (PPS) sampling method in two-stage sampling to enhance the efficiency of estimator. Secondary data were used. The first preliminary sample of clusters was selected independently by Simple Random Sampling Without Replacement (SRSWOR). The Lahiri's method was used to select First Stage Units (FSU) of clusters from the preliminary sample using Probability Proportional to Size with Replacement (PPSWR). Within each selected First stage clusters, sub-samples of Second Stage Units (SSU) were selected with SRSWOR. An estimator under PPS was derived and the expressions for the Bias and Mean Square Error (MSE) were obtained. The results showed that the PPS enhances the efficiency of the derived estimator. The derived estimator is therefore preferred in the estimation of a heterogeneous population parameter.
***KEYWORDS*:** Probability Proportional to Size, efficiency, Bias and Mean Square Error, estimator.

1. INTRODUCTION

History of learning about population by using sampling methods could be traced out to even very early stages of primitive life of mankind. In general, people are interested to study totality, called population, to decide about its nature. For such purpose, it is necessary to collect information regarding the population with respect to some characteristics. Such information is collected either by complete enumeration or by sample survey. The sample survey method is the most important tool of collection of such information because of its efficiency, accuracy and speed. Sample surveys are widely used as a cost effective apparatus of data collection and for making valid inference about population parameters. Since in sample surveys the sample is only a part of the whole, extrapolation inevitably leads to errors. The role of survey statisticians is to reduce errors either by devising suitable sampling schemes or by formulating efficient estimators of the parameters. To reduce errors, researchers have adopted the use of auxiliary information which correlates to the information under the study. The use of auxiliary information has been applied for improving the efficiencies of the estimators of population parameter(*s*) irrespective of sampling design. Ratio, product and regression methods of estimation are good examples in this context. Cochran ([Coc40]) used auxiliary information at the estimation stage and proposed a ratio estimator for the population mean. A ratio estimator is preferred when the correlation coefficient between the study variate and the auxiliary variate is positive. However, when such information is not completely known or lacking, and it is relatively cheaper to obtain information on the auxiliary variable(s), one can consider taking a large preliminary sample for estimating population mean(s) of the auxiliary variable(s) to be used at the estimation or selection stage of the ultimate estimation strategies. It is well known that suitable use of auxiliary information in probability sampling results in considerable reduction in the variance of the estimators of population parameter viz population mean (total), median, variance, regression coefficient and population correlation coefficients ([Sin03a]). Several research papers have been written on the use of auxiliary information in sample surveys during the last fifty years Tripathy *et al.* ([TDK94]) and Khare ([Kha03]). Two-stage sampling design offers a variety of possibilities for use of auxiliary information.

Two-stage sampling is a sampling scheme which involves selection of sample in two stages with selection of cluster samples in the first stage and then selecting sample of elements within each sampled clusters in the second stage. Two-stage sampling scheme consists in selecting the First Stage Units (FSU) by any of the sampling schemes e.g. Simple Random Sampling With Replacement (SRSWR), Simple Random Sampling Without Replacement (SRSWOR), Systematic Sampling, Probability Proportional to Size With Replacement (PPSWR), Probability Proportional to Size Without Replacement (PPSWOR) and the size of the FSU may be equal or unequal. From each selected FSU, a sample of Second Stage Units (SSU) is selected independently by any of the above suitable sampling procedure.

In almost all practical situations, the sampling units vary considerably in size and simple random sampling may not be effective in such cases as it does not take into account the possible importance of the larger units in the population. In Simple Random Sampling (SRS) probability of selection of every units in the population is equal but when sampling units are varying in their size if in that case SRS is being used for sample selection then unexpected result may arises as some of the more important (larger) units may not be included in the sample. Under this circumstance, a more precise estimator is used which takes into account the size of every sampling units and probability is assigned based on their sizes which is known as Probability Proportional to Size (PPS) sampling.

The aim of this paper is to develop an alternative approach for estimating population parameters in two-stage sampling by using Probability Proportional to Size (PPS) sampling method.

2. METHODOLOGY

Much literature has been produced on sampling from finite populations for the efficient estimation of the mean of a survey variable when auxiliary variable are available. In many situations information on the auxiliary variable is required either at the designing stage or estimation stage or both stages to increase the precision of the estimators. Ratio, product and Regression estimators are often used when advance knowledge of population mean of the auxiliary variable is readily available. In this paper, the ratio method of estimation was adopted.

When the population mean $\bar{X}$ of the auxiliary variable is unknown, the ratio estimator of $\bar{Y}$ with unequal FSU and its corresponding Mean Square Error (MSE) has been given by Singh *et al.* ([S+13a]) and are as presented below:

Ratio Estimator:

$$\bar{y}_1^{*}=\frac{\bar{y}^{*}}{\bar{x}^{*}}\bar{x}' \qquad (1)$$

$$\text{MSE}(\bar{y}_1^{*})=\lambda' S_{1R}^2+\sum_{i=1}^{n'}\lambda_{2i}S_{2Ri}^2-\lambda_3 S_{1R'}^2 -\sum_{i=1}^{n'}\lambda_{3i}S_{2Ri}^2 \qquad (2)$$

where $\lambda=\frac{1-f}{n}$; f=$\frac{n}{N}$; $\lambda_{2i}=\frac{M_i^2(1-f_{2i})}{nNM_i}$; $f_{2i}=\frac{m_i}{M_i}$; $f_1=\frac{n}{n'}$;

$f'=\frac{n'}{N}$; $\lambda_3=\frac{1-f'}{n'}$; $\lambda_{3i}=\frac{M_i^2(1-f_{2i})}{n'Nm_i}$

$S_{1R}^2=S_{1y}^2+R^2S_{1x}^2-2RS_{1xy}$;

$S_{1R'}^2=R^2S_{1x}^2-2RS_{1xy}$;

$S_{2Ri}^2=R^2S_{2xi}^2-2RS_{2xyi}$;

$S_{1p}^2=\frac{1}{n'-1}\sum_{i=1}^{n'}(\bar{P}_i-\bar{P})^2$;

$S_{2pi}^2=\frac{1}{n'-1}\sum_{i=1}^{n'}(P_{ij}-\bar{P}_i)^2$;

$S_{1PQ}=\frac{1}{n'-1}\sum_{i=1}^{n'}(\bar{P}_i-\bar{P})(\bar{Q}_i-\bar{Q})$;

$S_{2PQi}=\frac{1}{n'-1}\sum_{i=1}^{n'}(P_{ij}-\bar{P}_i)(Q_{ij}-\bar{Q}_i)$

P = x, y: Q= x, y

where S_{1p}^2 and S_{2pi}^2 are the variance among FSU means and variance among subunits for the i^{th} FSU while S_{1PQ} and S_{2PQi} are their corresponding co-variances.

3. SAMPLING PROCEDURE

Let U be a finite population partitioned into n' First Stage Units (FSU) denoted by $U_1, U_2, \ldots, U_{n'}$ such that the number Second Stage Units (SSU) in U_i is M_i. Let y_{ij} and x_{ij} be the values of the study variable (y) and an auxiliary variable (x) respectively for the jth SSU of the i^{th} FSU, U_i (i=1,2,…, n'; j=1,2,…, M_i). To estimate the population mean $\bar{Y}$ of the study variable Y when population mean $\bar{X}$ of the auxiliary variable is unknown, firstly, a preliminary sample of size n' FSU is selected out of N units in the population independently by using Simple Random Sampling Without Replacement (SRSWOR) in order to provide an estimate of $\bar{X}$ and information is collected for X only. Further, out of the preliminary sample of size n', a smaller sample of n FSU's is selected with Probability Proportional to Size With Replacement (PPSWR) by using the Lahiri's method.

Let $P_i=\frac{X_i}{X}$ denotes the probability of selecting the i^{th} unit in the sample on the basis of known auxiliary variable where X=$\sum_{i=1}^{X_i}X_i$. Let M=maxX_i, where M is the maximum of the sizes of the N units in the population. In order to select the first stage units using the Lahiri's method, a pair of random numbers (R_i, R_j) such that 1≤ R_i ≤ N and 1≤ R_j ≤ M are chosen from the table of random numbers. If R_j ≤M, the i^{th} unit is selected, otherwise the pair of (R_i, R_j) of random numbers is rejected and another pair of random numbers is chosen. To get a sample of size n, this procedure is repeated till n units are selected. Then from each of the i^{th} selected n FSU's, a subsample of m_i Second Stage Units (SSU's) is selected out of M_i SSU's by using SRSWOR and both y and x characteristics are measured.

4. PROPOSED ESTIMATOR

Utilizing the information on auxiliary character collected on preliminary sample for estimating the population mean and let $x_1', x_2', \ldots\ldots x_{n'}'$ be the first phase sample drawn by SRSWOR from the N units and let only the auxiliary variable X be measured. Also let $(y_1, y_2, \ldots \ldots y_n)$ and $(x_1, x_2, \ldots \ldots x_n)$ denote respectively the second phase sample drawn by PPSWR from the first phase sample for the study variable Y and auxiliary variable X respectively. Define $\bar{s}_n = \frac{1}{n}\sum_{1=1}^{n} s_i$ and $\bar{w}_n = \frac{1}{n}\sum_{i=1}^{n} w_i$ as the sample means of the y and x under PPS respectively and $\bar{x}' = \frac{1}{n'}\sum_{i=1}^{n'} x_i'$ as sample mean of X based on preliminary sample of size n', where $s_i = \frac{\bar{y}_i}{n'P_i}$; $w_i = \frac{\bar{x}_i}{n'P_i}$ and $x_i' = \sum_{i=1}^{n'} x_{ij}$.

Also let $P_i = \frac{M_i}{M}$; $\bar{y}_i = \frac{1}{m_i}\sum_{i=1}^{m_i} y_{ij}$ and $\bar{x}_i = \frac{1}{m_i}\sum_{i=1}^{m_i} x_{ij}$

Let $\bar{x}'$ be an unbiased estimator of X based on the preliminary sample and let $\bar{s}_n$ *and* $\bar{w}_n$ be unbiased estimators of X and Y based on the second sample and denoting the proposed estimator by $\bar{y}_{PR}$. Thus, the proposed ratio estimator proportional to size sampling is defined as:

$$\bar{y}_{PR} = \left[\frac{\bar{s}_n}{\bar{w}_n}\bar{x}'\right] \quad (3)$$

4.1 The Bias and Mean Square Error of the Proposed Estimator

The approximate expression of the Bias and MSE of the proposed estimator $(\bar{y}_{PR})$ can be obtained by defining the following as:

$$\delta_0 = \frac{\bar{s}_n - \bar{Y}}{\bar{Y}} \Rightarrow \frac{\bar{s}_n}{\bar{Y}} - 1 \Rightarrow \bar{s}_n = \bar{Y}(\delta_0 + 1) \quad (4)$$

$$\delta_1 = \frac{\bar{w}_n - \bar{X}}{\bar{X}} \Rightarrow \frac{\bar{w}_n}{\bar{X}} - 1 \Rightarrow \bar{w}_n = \bar{X}(\delta_1 + 1) \quad (5)$$

$$\delta_2 = \frac{\bar{x}' - \bar{X}}{\bar{X}} \Rightarrow \frac{\bar{x}'}{\bar{X}} - 1 \Rightarrow \bar{x}' = \bar{X}(\delta_2 + 1) \quad (6)$$

It is noted from literature that

$$E(\delta_0) = \left(\frac{\bar{s}_n - \bar{Y}}{\bar{Y}}\right) = 0;$$

$$E(\delta_1) = \left(\frac{\bar{w}_n - \bar{X}}{\bar{X}}\right) = 0;$$

$$E(\delta_2) = \left(\frac{\bar{x}' - \bar{X}}{\bar{X}}\right) = 0$$

$$E(\delta_0^2) = E\left(\frac{\bar{s}_n - \bar{Y}}{\bar{Y}}\right)^2 = \frac{V(\bar{s}_n)}{\bar{Y}^2};$$

$$E(\delta_1^2) = E\left(\frac{\bar{w}_n - \bar{X}}{\bar{X}}\right)^2 = \frac{V(\bar{w}_n)}{\bar{X}^2};$$

$$E(\delta_2^2) = E\left(\frac{\bar{x}' - \bar{X}}{\bar{X}}\right)^2 = \frac{V(\bar{x}')}{\bar{X}^2}$$

$$E(\delta_0, \delta_1) = E\left(\frac{(\bar{s}_n - \bar{Y})(\bar{w}_n - \bar{X})}{\bar{Y}\bar{X}}\right) = \frac{cov(\bar{s}_n, \bar{w}_n)}{\bar{Y}\bar{X}}$$

$$E(\delta_0, \delta_2) = E\left(\frac{(\bar{s}_n - \bar{Y})(\bar{x}' - \bar{X})}{\bar{Y}\bar{X}}\right) = \left(\frac{cov(\bar{s}_n, \bar{x}')}{\bar{Y}\bar{X}}\right)$$

$$E(\delta_1, \delta_2) = E\left(\frac{(\bar{w}_n - \bar{X})(\bar{x}' - \bar{X})}{\bar{X}^2}\right) = \left(\frac{cov(\bar{w}_n, \bar{x}')}{\bar{X}^2}\right)$$

Where

$$V(\bar{s}_n) = \frac{1}{n}\sum_{i=1}^{n} P_i (s_i - \bar{s}_n)^2 + \sum_{i=1}^{n'} \frac{M_i^2(1-f_{2i})}{nNm_i} S_{2yi}^2 \quad (7)$$

$$V(\bar{w}_n) = \frac{1}{n}\sum_{i=1}^{n} P_i (w_i - \bar{w}_n)^2 + \sum_{i=1}^{n'} \frac{M_i^2(1-f_{2i})}{nNm_i} S_{2xi}^2 \quad (8)$$

$$V(\bar{x}') = \lambda_3 S_{1x}^2 + \sum_{i=1}^{n'} \frac{M_i^2(1-f_{2i})}{n'Nm_i} S_{2xi}^2 \quad (9)$$

$$cov(\bar{s}_n, \bar{w}_n) = \frac{1}{n}\sum_{i=1}^{n} P_i (s_i - \bar{s}_n)(w_i - \bar{w}_n) + \sum_{i=1}^{n'} \frac{M_i^2(1-f_{2i})}{nNm_i} S_{2xyi}^2 \quad (10)$$

$$cov(\bar{s}_n, \bar{x}') = \frac{1}{n}\sum_{i=1}^{n} P_i (s_i - \bar{s}_n)(\bar{x}' - \bar{X}) + \sum_{i=1}^{n'} \frac{M_i^2(1-f_{2i})}{n'Nm_i} S_{2xyi}^2) \quad (11)$$

$$cov(\bar{w}_n, \bar{x}') = \frac{1}{n}\sum_{i=1}^{N} P_i (w_i - \bar{w}_n)(\bar{x}' - \bar{X}) + \sum_{i=1}^{n'} \frac{M_i^2(1-f_{2i})}{n'Nm_i} S_{2xi}^2 \quad (12)$$

where equation (7), (8) and (9) are the variances of sample means among FSU means and among subunits for the i^{th} FSU respectively while equation (10), (11) and (12) are their corresponding co-variances.

Where

$$S_{1x}^2 = \frac{1}{n'-1}\sum_{i=1}^{n'}(\bar{x}_i - \bar{x})^2 ;$$

$$S_{2xi}^2 = \frac{1}{n'-1}\sum_{i=1}^{n'}\left(x_{ij} - \bar{x}_i\right)^2;$$

$$S_{2yi}^2 = \frac{1}{n'-1}\sum_{i=1}^{n'}\left(y_{ij} - \bar{y}_i\right)^2;$$

$$S_{2xyi} = \frac{1}{n'-1}\sum_{i=1}^{n'}\left(x_{ij} - \bar{x}_i\right)\left(y_{ij} - \bar{y}_i\right)$$

The proposed estimator $(\bar{y}_{PR})$ can be expressed in terms of $\delta's$ by substituting equation (4), (5) and (6) into equation (3) as follows:

$$\bar{y}_{PR} = \bar{Y}(1 + \delta_0)(1 + \delta_2)(1 + \delta_1)^{-1} \quad (13)$$

$$\bar{y}_{PR} = \bar{Y}(1 + \delta_0)(1 + \delta_2)(1 - \delta_1 + \delta_1^2 - \delta_1^3 + \cdots) \quad (14)$$

Expanding the right hand of equation (14) and neglect terms involving powers of greater than two, we have

$$\bar{y}_{PR} = \bar{Y}(1 + \delta_0 + \delta_2 + \delta_0\delta_2 - \delta_1 - \delta_0\delta_1 - \delta_1\delta_2 + \delta_1^2) \quad (15)$$

Taking expectations of both sides of equation (15), we get

$$E(\bar{y}_{PR}) = \bar{Y}[1 + 0 + 0 + E(\delta_0\delta_2) - 0 - E(\delta_0\delta_1) - E(\delta_1\delta_2) + E(\delta_1^2)] \quad (16)$$

By the definition of Bias, we have

$$Bias(\bar{y}_{PR}) = E(\bar{y}_{PR}) - \bar{Y} \quad (17)$$

Substituting equation (16) into equation (17), we have,

$$\text{Bias}(\bar{y}_{PR})=\bar{Y}[E(\delta_0\delta_2)-E(\delta_0\delta_1)-E(\delta_1\delta_2)+E(\delta_1^2)] \quad (18)$$

Thus, the approximate Bias of the proposed estimator $(\bar{y}_{PR})$ to the first order of approximation is given as:

$$B(\bar{y}_{PR})=\bar{Y}\left[\frac{V(\bar{w}_n)}{\bar{X}^2}+\frac{cov(\bar{s}_n,\bar{x}')}{\bar{X}\bar{Y}}-\frac{cov(\bar{s}_n,\bar{w}_n)}{\bar{X}\bar{Y}}-\frac{cov(\bar{w}_n,\bar{x}')}{\bar{X}^2}\right] \quad (19)$$

By the definition of mean square error, we have

$$\text{MSE}(\bar{y}_{PR})=\text{E}(\bar{y}_{PR}-\bar{Y})^2 \quad (20)$$

Substituting equation (15) into equation (20), we have

$$\text{MSE}(\bar{y}_{PR})=\text{E}[\bar{Y}(\delta_0+\delta_2+\delta_0\delta_2-\delta_1-\delta_0\delta_1-\delta_1\delta_2+\delta_1^2)]^2 \quad (21)$$

By retaining the terms of equation (21) to order two, we have

$$\text{MSE}(\bar{y}_{PR})\cong\text{E}[\bar{Y}(\delta_0+\delta_2-\delta_1)]^2 \quad (22)$$
$$\text{MSE}(\bar{y}_{PR})\cong\bar{Y}^2\text{E}[(\delta_0+\delta_2-\delta_1)]^2 \quad (23)$$

Expanding the right of equation (23), and retaining the terms up to order two, we have

$$\text{MSE}(\bar{y}_{PR})=\bar{Y}^2E(\delta_0^2+\delta_1^2+\delta_2^2+2\delta_0\delta_2-2\delta_0\delta_1-2\delta_1\delta_2) \quad (24)$$

Thus the approximate Mean square error of the proposed estimator $(\bar{y}_{PR})$ to the first order of approximation is given as:

$$\text{MSE}(\bar{y}_{PR})=\bar{Y}^2\left[\frac{V(\bar{s}_n)}{\bar{Y}^2}+\frac{V(\bar{w}_n)}{\bar{X}^2}+\frac{V(\bar{x}')}{\bar{X}^2}+2\frac{cov(\bar{s}_n,\bar{x}')}{\bar{X}\bar{Y}}-2\frac{cov(\bar{s}_n,\bar{w}_n)}{\bar{X}\bar{Y}}-2\frac{cov(\bar{w}_n,\bar{x}')}{\bar{X}^2}\right] \quad (25)$$

5. EFFICIENCY COMPARISON

$\bar{y}_{PR}$ will be more efficient than $\bar{y}_1^*$ if it satisfies the following conditions:

(i) $\text{MSE}(\bar{y}_{PR}) < \text{MSE}(\bar{y}_1^*)$

(ii) $\frac{\text{MSE}(\bar{y}_1^*)}{\text{MSE}(\bar{y}_{PR})}>1$

6. NUMERICAL ILLUSTRATIONS

In this section, a numerical illustration was carried out to demonstrate the utility of the proposed sampling scheme and compare it with the existing sampling scheme. The data on Yield of Tobacco (metric tons) and Area of land (Hectares) by one hundred and six (106) countries published by Agricultural Statistics (1999) Washington, D.C., available in Singh ([Sin03b]), Appendix, pp.1119-1121, were extracted and used in this paper. These countries are from twenty two regions in the world and each region formed a cluster. The yield of tobacco and the area of land is taking as Y and X respectively.

The population consists of 106 countries (SSU) divided into preliminary sample of size n' = 10 regions (FSU). The number of countries (M_i) in 10 regions are 6, 6, 8, 10, 12, 4, 30, 17, 10 and 3. The Lahiri's method was used to select a PPSWR sample of n=6 FSU which is 60% of the preliminary sample of size n'. Within each selected FSU, 60% sub-samples of size m_i SSU were selected by SRSWOR.Let the first stage sample size be n= 6 i.e. (6, 6, 10, 30, 10 and 3) and the sizes of the second stage sample i.e. m_i(i = 1, 2, …, 6) as 4, 4, 6, 18, 6 and 2 respectively.For comparison of Mean square error (MSE) of $\bar{y}_{PR}$ *and* $\bar{y}_1^*$, we consider 60% sampling fraction by using SRSWOR at both stages.

Table 1: Statistical Computations for PPS sampling

M_i	m_i	P_i	s_i	w_i	S_{2yi}^2	S_{2xi}^2	S_{2xyi}	f_{2i}	λ_{2i}	λ_{3i}
6	4	0.09	2.06	2456.7	0.006	1397191.6	-31.34	0.67	0.023	0.014
6	4	0.092	1.47	23776	0.106	254181258.3	-4208.0	0.67	0.02	0.01
10	6	0.15385	1.00425	5377.13	0.06986	34088631.9	-940.99	0.60000	0.05051	0.03030
30	18	0.46154	0.23990	3752.51	0.79626	2265067540	12191.01	0.60000	0.15152	0.09001
10	6	0.15385	1.32383	36150.83	1.84006	7559964815	-39751.96	0.60000	0.05051	0.03030
3	2	0.04615	3.98667	3683.33	0.17602	568888.9	316.44	0.66667	0.01136	0.00681

Table 2: Statistical Computations for SRS Sampling

M_i	m_i	f_{2i}	λ_{2i}	λ_{3i}	S^2_{2xi}	S_{2xyi}	S_{1xy}	S^2_{2Ri}
6	4	0.6667	0.02273	0.01364	1397191.6	-31.33778	2.128E+06	0.01122
10	6	0.6000	0.05051	0.03030	48513502.3	-1138.81111	2.147E+06	0.39680
12	7	0.5833	0.06494	0.03896	12404493.3	0.16900	2.186E+06	-0.10828
4	2	0.5000	0.03030	0.01818	568888.9	0.00124	1.458E+06	0.00154
30	18	0.6000	0.15152	0.09091	1062967433	1.34954	2.125E+06	3.85333
3	2	0.6667	0.01136	0.00682	405000	0.00315	2.126E+06	-0.00117

Table 3: Summary of Estimated Parameters under PPS sampling

Parameters	Estimates	Parameters	Estimates
$\bar{Y}$	1.5510337736	$V(\bar{s}_n)$	0.440007708
$\bar{X}$	22169.72642	$V(\bar{w}_n)$	758200517.6
f'	0.454545455	$V(\bar{x}')$	930313274.6
λ_3	0.054545455	$Cov(\bar{w}_n, \bar{s}_n)$	327.8120457
$\bar{x}'$	234999.1	$Cov((\bar{s}_n, \bar{x}')$	-25073.05072
$\bar{s}_n$	1.680149691	$Cov(\bar{w}_n, \bar{x}')$	374952466.9
$\bar{w}_n$	12532.83156	$\bar{y}_{PR}$	31.50394733

Table 4: Summary of Estimated Parameters under SRS Sampling

Parameters	Estimates	Parameters	Estimates
f_1	0.6	S^2_{1y}	43.56910395
f'	0.0454545455	S^2_{1x}	1364236412
λ'	0.066666667	S_{1xy}	241504.3141
λ^3	0.0545454545	S^2_{1R}	16.4543657
R	6.9962E-05	$S^2_{1R'}$	-27.11473825
$\sum_{i=1}^{n'} \lambda_{2i} S^2_{2Ri}$	0.597137057	$\sum_{i=1}^{n'} \lambda_{3i} S^2_{2Ri}$	0.358282234

Calculated Results:

Taking N=22, n'=10, n=6.

The estimate of the proposed estimator under PPS is given as:

$$\bar{y}_{PR} = \left[\frac{\bar{s}_n}{\bar{w}_n}\bar{x}'\right] = \left[\frac{1.680149691}{12532.83156} * 234999.1\right] = 31.504$$

An estimate of the MSE $(\bar{y}_{PR})$ under PPS is given as:

$$MSE(\bar{y}_{PR}) = \bar{Y}^2\left[\frac{V(\bar{s}_n)}{\bar{Y}^2} + \frac{V(\bar{w}_n)}{\bar{X}^2} + \frac{V(\bar{x}')}{\bar{X}^2} + 2\frac{cov(\bar{s}_n, \bar{x}')}{\bar{X}\bar{Y}} - 2\frac{cov(\bar{s}_n, \bar{w}_n)}{\bar{X}\bar{Y}} - 2\frac{cov(\bar{w}_n, \bar{x}')}{\bar{X}^2}\right]$$

$$= (1.5510337736)^2\left[\frac{0.440007708}{(1.5510337736)^2} + \frac{758200517.6}{(22169.72642)^2} + \frac{930313274.6}{(22169.72642)^2} + 2*\frac{(-25073.05072)}{(22169.72642*1.5510337736)} - 2*\frac{327.8120457}{(22169.72642*1.5510337736)} - 2*\frac{374952466.9}{(22169.72642)^2}\right]$$

=1.480

An estimate of the Bias $(\bar{y}_{PR})$ is given as:

$$B(\bar{y}_{PR}) = \bar{Y}\left[\frac{V(\bar{w}_n)}{\bar{X}^2} + \frac{cov(\bar{s}_n, \bar{x}')}{\bar{X}\bar{Y}} - \frac{cov(\bar{s}_n, \bar{w}_n)}{\bar{X}\bar{Y}} - \frac{cov(\bar{w}_n, \bar{x}')}{\bar{X}^2}\right]$$

$$= 1.5510337736\left[\frac{758200517.6}{(22169.72642)^2} + \frac{(-25073.05072)}{(22169.72642*1.5510337736)} - \frac{327.8120457}{(22169.72642*1.5510337736)} - \frac{374952466.9}{(22169.72642)^2}\right] = 0.064$$

The MSE of the existing estimator under SRS is given as:

$$MSE(\bar{y}_1^*) = \lambda' S^2_{1R} + \sum_{i=1}^{n'} \lambda_{2i} S^2_{Ri} - \lambda_3 S^2_{1R'} - \sum_{i=1}^{n'} \lambda_{3i} S^2_{2Ri}$$

=(0.066666667*16.4543657)+0.597137057-(0.0545454545*(-27.11473825))-0.358282234

= 2.815

The Relative Efficiency of the PPS estimator with respect to the SRS estimator is given as:

$$Efficiency = \frac{MSE(\bar{y}_1^*)}{MSE(\bar{y}_{PR})} = \frac{2.815}{1.480} = 1.902 > 1$$

Table 5: Comparison of Mean Square Errors (MSE)

Estimator	Existing $\bar{y}_1^*$	Proposed $\bar{y}_{PR}$	Efficiency
MSE	2.815	1.480	1.902

6. DISCUSSION OF RESULTS

The statistical computations for PPS and SRS sampling are given in Table 1 and 2 respectively. The estimated parameters for PPS and SRS sampling are given in Table 3 and 4 respectively while Table 5 gives the comparison of the MSE.
The numerical illustration shows that the Mean Square Error (MSE) of the existing and proposed estimators was 2.815 and 1.480 respectively. It can be seen that there was gain in efficiency as the proposed estimator has minimum mean square error as compared to the existing estimator. Furthermore, the relative efficiency of the PPS estimator ($\bar{y}_{PR}$) with respect to the estimator ($\bar{y}_1^*$) under SRS was 1.902 which is greater than one, shows that the PPS estimator is more efficient than the estimator under SRS design proposed by Singh *et al.* ([S+13b]).

7. CONCLUSION

The estimator in two-stage sampling using Probability Proportional to Size (PPS) sampling method has been evaluated for its comparison with the existing estimator using Simple Random Sampling (SRS). From the comparison of the results obtained, it can be concluded the PPS enhances the efficiency of the derived estimator whose MSE was also found to be smaller. The derived estimator is therefore preferred in the estimation of heterogeneous population parameter, hence, recommended for samplers.

REFERENCES

[AHR14] **Abdulla F., Hossain M., Rahman M.** – *On the Selection in Probability to Size Sampling: Cumulative Relative Frequency Method*. Mathematical Theory and Modelling.4:6, 2014.

[Coc40] **Cochran W. G.** – *The Estimation of the Yields of the Cereal Experiments by Sampling for the Ratio of Grain to Total Produce*, Jour. Agri. Sci. 59, 1225-1226, 1940.

[Coc77] **Cochran W. G.** – *Sampling techniques*. 3rd edition, John Wiley, New York, 1977.

[Iva15] **Ivanov L.** – *General Geography and History of Livingston Island: Bulgarian Antartic Research: A Synthesis*. Eds.: C. Pimpirev and N. Chipev. Sofia: St. Kliment Ohridiki University Press, 17-28, 2015.

[Kha03] **Khare B. B.** – *Estimation of Ratio and Product of Two Population Means using Auxiliary Characters in the Presence of Non-Response*. Sampling Stategies for Finite Population using Auxiliary Information. 31-38, 2003.

[Lah51] **Lahiri D. B.** – *A Method of Sample Selection Providing Unbiased Ratio Estimates*. Bulletin of the International Statistical Institute 33(2): 133-146, 1951.

[Nob17] **Nobin C. P.** – *Sample Selection with Probability Proportional to Size Sampling using SAS and R software*. International Journal of Application or Innovation in Engineering and Management (IJAIEM), 6(1), 2017.

[Sam01] **Sampath S.** – *Sampling Theory and methods*. Narosa Publishing House, New Delhi-110 017, India, 2001.

[Sin03a] **Singh G. N.** – *On the Improvement of Product Method of Estimation in Sample Surveys*. Jóur. Ind. Stat. App. Sci., 56, 267-275, 2003.

[Sin03b] **Singh S.** – *Advanced Sampling Theory with applications: How Michael "Selected" Amy*. Vol 1&2, 1-1247. Kluwer Academic Publishers, The Netherlands, 2003.

[Suk54] **Sukhatme P. V.** – *Sampling Theory of Surveys with Applications*. The Indian society of Agricultural statistics, New Delhi, India, 1954.

[SC86] **Singh D., Chaudhary F. S.** – *Theory and Analysis of Sample Survey Designs*, New Age International (P) Ltd., Publishers, New Delhi, 1986.

[SG09] **Srivastava M., Garg N.** – *A General Class of Estimators of a Finite Population Mean using Multi-Auxiliary Information under Two-stage Sampling Scheme*. JRSS, 2(1), 103-118, 2009.

[SS12] **Singh H. P., Solanki R. S.** – *Improved estimation of population mean in simple random sampling using information on auxiliary attribute*. Applied Mathematics and Computation, 218(15), 7798-7812, 2012.

[SS13] **Singh H. P., Solanki R. S.** – *A new procedure for variance estimation in simple random sampling using auxiliary information*. Statistical Papers, 54(2), pp. 479-497, 2013.

[SSS14] **Singh H. P. Solanki R. S., Singh A. K.** – *Predictive Estimation of finite population mean using Exponential Estimators*. Statistika 94(1): 41-53, 2014.

[S+13a] **Singh R., Vishwakarma G. K., Gupta P. C., Pareek S.** – *An Alternative Approach to Estimation of Population Mean in Two-Stage Sampling.* Mathematical Theory and Modelling, 3(13), 48-53, 2013.

[S+13b] **Singh R., Gupta P.C., Pareek S., Vishwakarma G. K.** – *An Alternative approach to estimation of population mean in Two-stage sampling*. Journal of culture, society and Development. Vol 2, 72-76, 2013.

[TDK94] **Tripathi T. P., Das A. K., Khare B. B.** – *Use of Auxiliary Information in Sample Surveys - A review*. Research Gate Gmbh, 1994.

[***14] **Statistical Institute for Asia and Pacific** – *Regional Training Course on Sampling Methods for Producing Core Data Items for Agricultural and Rural statistics*. Jakarta,Indonesia, 2014.