Journal of Reliability and Statistical Studies; ISSN (Print): 0974-8024, (Online): 2229-5666
Vol. 11, Issue 2 (2018): 99-112

USE OF EXTREME VALUES TO ESTIMATE FINITE POPULATION MEAN UNDER PPS SAMPLING SCHEME

Sohaib Ahmad and Javid Shabbir*
Department of Statistics, Quaid-e-Azam University Islamabad, Pakistan
E Mail: sohaibahmad@stat.qau.edu.pk; javidshabbir@gmail.com
*Corresponding Author:

Received February 26, 2018
Modified August 09, 2018
Accepted October 05, 2018

Abstract

We propose ratio, product and regression type estimators for estimation of finite population mean under probability proportional to size (PPS) sampling design, when there exist some extreme values regarding the study and the auxiliary variables respectively. The biases and mean squared errors are derived up to first order of approximation. Theoretical and empirical studies show that the proposed estimators perform better as compared to usual estimators.

Key Words: PPS, Auxiliary Information, Bias, Mean Square Error.

1. Introduction

It is well known that at large scale survey sampling, use of several auxiliary variables may improve the precision of the estimators. The auxiliary information can be used either at the selection stage or at estimation stage, or at both stages. Many researchers have already attempted to obtain the estimates of population parameters such as mean, total and median etc., that possess maximum statistical properties. For that purpose, a representative part of population is needed. When characteristics of the population are homogeneous then one can use simple random sampling (SRS) for selecting the units. On the other hand, when units vary considerably in size, then one can use probability proportional to size (PPS) sampling scheme**.**

Agarwal and Kumar (1980) suggested the combination of ratio and PPS estimators for population mean $(\bar{Y})$. Rao (1966) proposed alternative estimators in PPS sampling for multiple characteristics. Tripathi (1969) introduced the regression type estimator with PPS for estimate of population mean $(\bar{Y})$. The readers are also referred to the papers by Pandey et al (1984), Srivenkataramana (1979), and Singh (2017).
Many research material exists for the construction of ratio, product and regression type estimators in estimating $\bar{Y}$. The readers can explore these research findings by looking the references Kadilar and Cingi (2004, 2006), Gupta and Shabbir (2008), Haq and Shabbir (2013), Grover and Kaur (2011), and Haq et al. (2013). Khanand Shabbir (2013), and Al-Hossainand Khan (2014)used the extreme values in estimating $\bar{Y}$.

In this article, we propose ratio, product and regression type estimators for estimating the finite population mean in the presence of extreme values under PPS sampling scheme. The performance of the proposed estimators, relative to their usual counter parts has been studied. We have obtained the MSE values by using 4 data sets in Table 1.

2. Notations and Symbols

Consider a finite population $\Omega=\{1,2,...,N\}$. Let y_i and $(x_i,\ z_i)$ be the values of the study variable y and the auxiliary variables (x,z) respectively. Let $P_i=\frac{z_i}{\sum_{i=1}^{N} z_i}$ be the probability proportional to size sampling for i^{th} unit. We draw a sample of size n by adopting the PPS sampling with replacement.

Let

$\bar{u}=\frac{1}{n}\sum_{i=1}^{n} u_i=\bar{y}_{pps}$ and $\bar{v}=\frac{1}{n}\sum_{i=1}^{n} v_i=\bar{x}_{pps}$, be the sample means corresponding to the population $\bar{Y}=\sum_{i=1}^{N}\frac{y_i}{N}$ and $\bar{X}=\sum_{i=1}^{N}\frac{x_i}{N}$ respectively,

where

$$u_i=\frac{y_i}{NP_i}\ ,\quad v_i=\frac{x_i}{NP_i},S_u^2=\sum_{i=1}^{N} P_i\left(u_i-\bar{Y}\right)^2,\ S_v^2=\sum_{i=1}^{N} P_i\left(v_i-\bar{X}\right)^2,$$

$$\rho_{uv}=\frac{\sum_{i=1}^{N} P_i\left(u_i-\bar{Y}\right)\left(v_i-\bar{X}\right)}{S_u S_v},\ S_u=\sqrt{P_i\left(u_i-\bar{Y}\right)^2},S_v=\sqrt{P_i\left(v_i-\bar{X}\right)^2}.$$

We use the following error terms for obtaining the properties of the estimators.

Let $e_0=\frac{\bar{u}}{\bar{Y}}-1$ and $e_1=\frac{\bar{v}}{\bar{X}}-1$, such that

$$E\left(e_0\right)=E\left(e_1\right)=0,\ E\left(e_0^2\right)=\lambda C_u^2,\ E\left(e_1^2\right)=\lambda C_v^2\ ,\ E\left(e_0e_1\right)=\lambda r_{uv}C_uC_v,$$

where $\lambda=\frac{1}{n}$.

Many actual data sets may contain unexpected large $(y_{\max})$ or small $(y_{\min})$ observations. For example, in estimating the mean income of a family, earning of the wealthy person (maximum) in the community of people living in a particular country or region is well known and that of poorest (minimum) can easily be assessed. Likewise, many organizations organized different types of surveys regularly after specific interval of time; information about maximum and minimum values can easily be obtained.

Mean per unit estimator for finite population mean is very sensitive to unexpected values, if there exists unexpected very large say $(y_{\max})$ and very small say $(y_{\min})$ units in the population. In addition, mean per unit estimator is very sensitive to these unexpected observations and as a result population mean will either be underestimated in the case of sample contains $(y_{\min})$ or overestimated in the case of sample contains $(y_{\max})$. So, in such situations, the estimator can produce misleading results if any of the unexpected value is selected in the sample. To overcome this situation, Sarndal (1972) suggested the following an unbiased estimator:

$$\bar{y}_S = \begin{cases} \bar{y} + c, \text{ if sample contains } y_{\min} \text{ but not } y_{\max} \\ \bar{y} - c, \text{ if sample contains } y_{\max} \text{ but not } y_{\min} \\ \bar{y}, \text{ for all samples,} \end{cases} \tag{1}$$

where c is the constant.

The variance of $\bar{y}_S$ is given by

$$V(\bar{y}_S) = \lambda S_y^2 - \frac{2\lambda nc}{N-1}(y_{\max} - y_{\min} - nc), \tag{2}$$

The optimum value of c is given by

$$c_{opt} = \frac{(y_{\max} - y_{\min})^2}{2n} .$$

The minimum variance of $\bar{y}_S$ is:

$$V(\bar{y}_S)_{\min} = V(\bar{y}) - \frac{\lambda (y_{\max} - y_{\min})^2}{2(N-1)} , \tag{3}$$

where $V(\bar{y}) = \lambda S_y^2.$

The minimum variance of $\bar{y}_S$ is always smaller than the variance of $\bar{y}$.

Singh et al. (2018) introduced the following class of estimators under PPS sampling for population total:

$$\hat{Y}_{pps(S)} = \frac{1}{n}\sum_{i=1}^{n}\frac{(y_i+\theta)}{p_{i(n)}} - N\theta$$

$$= \frac{(C_x + N\eta)}{n}\sum_{i=1}^{n}\frac{(y_i+\theta)}{(Cx_i+\eta)} - N\theta,$$

where (C, η, θ) are suitably chosen scales and $p_{i(n)} = \frac{Cx_i + \eta}{Cx_i + N\eta}$.

For estimation of population mean, above estimator becomes:

$$\hat{\bar{Y}}_{pps(S)} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i+\theta}{Np_{i(n)}} - \theta\right)$$

$$= \left(\frac{Cx + N\eta}{Nn}\right)\sum\frac{(y_i+\theta)}{(Cx_i+\eta)} - \theta.$$

For $C = 1,\ \eta = 0$ and $\theta = 0,$ we have the usual mean estimator

$$\hat{\bar{Y}}_{pps(S)} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i}{Np_i}\right) = \frac{1}{n}\sum_{i=1}^{n}u_i = \bar{u} = \bar{y}_{pps}$$

The variance of $\bar{y}_{pps}$ is given by

$$V\left(\bar{y}_{pps}\right) = \lambda\bar{Y}^2 C_u^2. \tag{4}$$

The usual ratio and product estimators in PPS sampling are given as:

$$\bar{y}_{R(pps)} = \bar{u}\left(\frac{\bar{X}}{\bar{v}}\right), \tag{5}$$

and $\bar{y}_{P(pps)} = \bar{u}\left(\frac{\bar{v}}{\bar{X}}\right).$ (6)

The bias and MSE of $\bar{y}_{R(pps)}$ and $\bar{y}_{P(pps)}$ are given by:

$$Bias\left(\bar{y}_{R(pps)}\right) \cong \lambda\bar{Y}\left(C_v^2 - r_{uv}C_uC_v\right), \tag{7}$$

$$Bias\left(\bar{y}_{P(pps)}\right) \cong \lambda\bar{Y}\rho_{uv}C_uC_v, \tag{8}$$

$$MSE\left(\bar{y}_{R(pps)}\right) \cong \lambda\bar{Y}^2\left(C_u^2 + C_v^2 - 2r_{uv}C_uC_v\right). \tag{9}$$

$$MSE\left(\bar{y}_{P(pps)}\right) \cong \lambda\bar{Y}^2\left(C_u^2 + C_v^2 + 2r_{uv}C_uC_v\right). \tag{10}$$

The usual regression estimator in PPS sampling, is given by:

$$\bar{y}_{lr(pps)} = \bar{u} + b\left(\bar{X} - \bar{v}\right), \tag{11}$$

where b is the sample regression coefficient. If b is the least square estimator of B, then the variance of the estimator $\bar{y}_{lr(pps)}$, is given by:

$$V\left(\bar{y}_{lr(pps)}\right) = \lambda S_u^2\left(1 - r_{uv}\right). \tag{12}$$

3. Suggested Estimator

On the lines of Sarndal (1972) and Khan and Shabbir (2013), we suggest the usual ratio, product and regression types estimators under PPS sampling in the presence of extreme values. Here we suggest an improved estimator in different situation separately as follows.

3.1 Case (1): The usual mean per unit estimator

The usual mean per unit estimator is given by:

$$\hat{\bar{Y}}_{0(pps)} = \begin{cases} \bar{u} + c, & \text{if sample contains } u_{\min} \\ \bar{u} - c, & \text{if sample contains } u_{\max} \\ \bar{u}, & \text{for all other samples} \end{cases} \tag{13}$$

The minimum variance of $\hat{\bar{Y}}_{0(pps)}$ at optimum values of c i.e:

$$c_{(opt)} = \frac{\lambda\left(u_{\max} - u_{\min}\right)^2}{2(N-1)}$$, is given by:

$$V(\hat{\bar{Y}}_{0(pps)}) = V(\bar{y}_{pps}) - \frac{\lambda\left(u_{\max} - u_{\min}\right)^2}{2(N-1)}. \tag{14}$$

3.2 Case(2): When correlation between u and v is positive

When we have positive correlation between u and v, then selection of larger value of v, a larger value of u is assume to be selected and when smaller value of u is selected, then selection of smaller value v is assumed. Sowe define the following ratio type estimator:

$$\hat{\bar{Y}}_{R(pps)} = \frac{\bar{u}_{c_{11}}}{\bar{v}_{c_{21}}}\bar{X},$$

or

$$\hat{\hat{Y}}_{R(pps)} = \begin{cases} \dfrac{\overline{u}+c_1}{\overline{v}-c_2}\overline{X} \\ \dfrac{\overline{u}-c_1}{\overline{v}-c_2}\overline{X} \\ \dfrac{\overline{u}}{\overline{v}}\overline{X} \end{cases} \quad (15)$$

The regression type estimator is:

$$\hat{\hat{Y}}_{lr1(pps)} = \overline{u}_{c_{11}} + b\left(\overline{X} - \overline{v}_{c_{21}}\right), \quad (16)$$

where $\left(\overline{u}_{c_{11}} = \overline{u} + c_1,\ \overline{v}_{c_{21}} = \overline{v} + c_2\right)$ if the sample contains $u_{\min}$ and $v_{\min}$; $\left(\overline{u}_{c_{11}} = \overline{u} - c_1,\ \overline{v}_{c_{21}} = \overline{v} - c_2\right)$ if the sample contains $u_{\max}$ and $v_{\max}$ and $\left(\overline{u}_{c_{11}} = \overline{u},\ \overline{v}_{c_{21}} = \overline{v}\right)$ for all the other combination of sample.

3.3 Case(3): When correlation between u and v is negative

When u and v are negatively correlated then with selection of large value of v, a smaller value of u is expected to be selected , and when smaller value of v is selected, a large value of u is expected to be selected. Keeping this point of view, we suggest the following product type estimator:

$$\hat{\hat{Y}}_{P(pps)} = \overline{u}_{c_{12}} \frac{\overline{v}_{c_{22}}}{\overline{X}} \quad (17)$$

or

$$\hat{\hat{Y}}_{P(pps)} = \begin{cases} \dfrac{\left(\overline{u}+c_1\right)\left(\overline{v}-c_2\right)}{\overline{X}} \\ \dfrac{\left(\overline{u}-c_1\right)\left(\overline{v}+c_2\right)}{\overline{X}} \\ \dfrac{\overline{u}}{\overline{v}}\overline{X} \end{cases} \quad (18)$$

The regression type estimator is:

$$\hat{\hat{Y}}_{lr2(pps)} = \overline{u}c_{12} + b\left(\overline{X} - \overline{v}c_{22}\right), \quad (19)$$

where $\left(\overline{u}c_{12} = \overline{u} + c_1, \overline{v}c_{22} = \overline{v} - c_2\right)$ if the sample contains $u_{\min}$ and $v_{\max}$; $\left(\overline{u}c_{12} = \overline{u} - c_1,\ \overline{v}c_{22} = \overline{v} + c_2\right)$ if the sample contains $u_{\max}$ and $v_{\max}$, and $\left(\overline{u}c_{12} = \overline{u},\ \overline{v}c_{22} = v\right)$ for all the other combination of sample.

To get the properties of the suggested estimators in form of bias and mean squared error, we define the relative error terms and their expectations below.

Let $e_0 = \dfrac{\bar{u}c_1 - \bar{Y}}{\bar{Y}}$ and $e_1 = \dfrac{\bar{v}c_2 - \bar{X}}{\bar{X}}$,

$$E\left(e_0^2\right) = \left(\frac{\lambda}{\bar{Y}^2}\right)\left[S_u^2 - \frac{2nc_1}{N-1}\left(u_{\max} - u_{\min} - nc_1\right)\right],$$

$$E\left(e_1^2\right) = \left(\frac{\lambda}{\bar{X}^2}\right)\left[S_v^2 - \frac{2nc_2}{N-1}\left(v_{\max} - v_{\min} - nc_2\right)\right],$$

$$E\left(e_0 e_1\right) = \left(\frac{\lambda}{\bar{Y}\bar{X}}\right)\left[S_{uv} - \frac{n}{N-1}\left\{c_2\left(u_{\max} - u_{\min}\right) + c_1\left(v_{\max} - v_{\min}\right) - 2nc_1c_2\right\}\right],$$

Expressing $\hat{\bar{Y}}_{R(pps)}$ in terms of e's we have

$$\hat{\bar{Y}}_{R(pps)} = \bar{Y}\left(1 + e_0\right)\left(1 + e_1\right)^{-1} \tag{20}$$

or

$$\left(\hat{\bar{Y}}_{R(pps)} - \bar{Y}\right) = \bar{Y}\left(e_0 - e_1 - e_0e_1 + e_1^2\right). \tag{21}$$

The bias of $\hat{\bar{Y}}_{R(pps)}$, to first order of approximation is given by:

$$Bias\left(\hat{\bar{Y}}_{R(pps)}\right) \cong \frac{\lambda}{\bar{X}}\left[R\left\{S_v^2 - \frac{2nc_2}{N-1}\left(u_{\max} - u_{\min} - nc_2\right)\right\} - \left\{S_{uv} - \frac{n}{N-1}\left(c_2\left(u_{\max} - u_{\min}\right)\right.\right.\right.$$

$$\left.\left.\left. + c_1\left(v_{\max} - v_{\min}\right) - 2nc_1c_2\right)\right\}\right], \tag{22}$$

where $R = \dfrac{\bar{Y}}{\bar{X}}$.

The MSE of $\hat{\bar{Y}}_{R(pps)}$ up to the first order of approximation, is given by:

$$MSE\left(\hat{\bar{Y}}_{R(pps)}\right) \cong \lambda\left[S_u^2 - \frac{2nc_1}{N-1}\left(u_{\max} - u_{\min} - nc_1\right) + R^2\left\{S_v^2 - \frac{2nc_2}{N-1}\left(v_{\max} - v_{\min} - nc_2\right)\right\}\right.$$

$$\left. -2R\left\{S_{uv} - \frac{n}{N-1}\left(c_2\left(u_{\max} - u_{\min}\right) + c_1\left(v_{\max} - v_{\min}\right) - 2nc_1c_2\right)\right\}\right]. \tag{23}$$

or

$$MSE\left(\hat{\bar{Y}}_{R(pps)}\right) \cong MSE\left(\bar{y}_{R(pps)}\right) - \frac{2\lambda n}{N-1}\left[\left(c_1 - Rc_2\right)\left\{\left(u_{\max} - u_{\min}\right)\right.\right.$$

$$\left.\left. - R\left(v_{\max} - v_{\min}\right) - n\left(c_1 - Rc_2\right)\right\}\right]. \tag{24}$$

To find out the optimum values of c_1 and c_2 , we differentiate (23) with respect to c_1 and c_2 as:

$$\frac{\partial M\left(\hat{\hat{Y}}_{R(pps)}\right)}{\partial c_1}=\left(u_{\max}-u_{\min}\right)-R\left(v_{\max}-v_{\min}\right)-2n\left(c_1-Rc_2\right)=0,$$

$$\frac{\partial M\left(\hat{\hat{Y}}_{R(pps)}\right)}{\partial c_2}=\left(u_{\max}-u_{\min}\right)-R\left(v_{\max}-v_{\min}\right)-2n\left(c_1-Rc_2\right)=0.$$

Here we have one equation with two unknowns,so unique solution is not possible. The optimum values of c_1 and c_2 are

$$c_{1opt}=\frac{\left(u_{\max}-u_{\min}\right)}{2n} \text{ and } c_{2opt}=\frac{\left(v_{\max}-v_{\min}\right)}{2n}.$$

For optimum values of c_1 and c_2, the minimum MSE of $\hat{\hat{Y}}_{R(pps)}$, is given as:

$$MSE\left(\hat{\hat{Y}}_{R(pps)}\right)_{\min}\cong MSE\left(\bar{y}_{R(pps)}\right)-\frac{\lambda}{2(N-1)}\left[\left(u_{\max}-u_{\min}\right)-R\left(v_{\max}-v_{\min}\right)\right]^2. \tag{25}$$

Similarly, the bias and *MSE*of $\hat{\hat{Y}}_{P(pps)}$ are respectively, given by:

$$Bias\left(\hat{\hat{Y}}_{P(pps)}\right)\cong\frac{\lambda}{\bar{X}}\left[S_{uv}-\frac{n}{N-1}\left\{\left(c_2\left(u_{\max}-u_{\min}\right)+c_1\left(v_{\max}-v_{\min}\right)-2nc_1c_2\right)\right\}\right] \tag{26}$$

and

$$MSE\left(\hat{\hat{Y}}_{P(pps)}\right)\cong MSE\left(\bar{y}_{P(pps)}\right)-\frac{2\lambda n}{N-1}\left[\left(c_1+Rc_2\right)\left\{\left(u_{\max}-u_{\min}\right)\right.\right.$$

$$\left.\left.+R\left(v_{\max}-v_{\min}\right)-n\left(c_1+Rc_2\right)\right\}\right]. \tag{27}$$

For optimum values of c_1 and c_2, the minimum mean squared error of $\hat{\hat{Y}}_{P(pps)}$ is given as:

$$MSE\left(\hat{\hat{Y}}_{P(pps)}\right)_{\min}\cong MSE\left(\bar{y}_{P(pps)}\right)-\frac{\lambda}{2(N-1)}\left[\left(u_{\max}-u_{\min}\right)-R\left(v_{\max}-v_{\min}\right)\right]^2. \tag{28}$$

The variance of regression estimator $\hat{\hat{Y}}_{lr1(pps)}$ in case of positive correlation, is given by:

$$V\left(\hat{\bar{Y}}_{lr1(pps)}\right) \cong V\left(\bar{y}_{lr(pps)}\right) - \frac{2\lambda n}{N-1}\Big[\left(c_1 - \beta c_2\right)\Big\{\left(u_{\max} - u_{\min}\right) - \beta\left(v_{\max} - v_{\min}\right)$$
$$-2n\left(c_1 - \beta c_2\right)\Big\}\Big], \tag{29}$$

where $\beta = r_{uv}\dfrac{S_u}{S_v}$ is the population regression coefficient of u on v. For

$c_{1opt} = \dfrac{\left(u_{\max} - u_{\min}\right)}{2n}$, $c_{2opt} = \dfrac{\left(v_{\max} - v_{\min}\right)}{2n}$,

the minimum variance of $\hat{\bar{Y}}_{lr1(pps)}$, is given as:

$$V\left(\hat{\bar{Y}}_{lr1(pps)}\right)_{\min} = V\left(\bar{y}_{lr(pps)}\right) - \frac{\lambda}{2(N-1)}\left[\left(u_{\max} - u_{\min}\right) - \beta\left(v_{\max} - v_{\min}\right)\right]^2. \tag{30}$$

For negative correlation, the variance of regression estimator $\hat{\bar{Y}}_{lr2(pps)}$, is given by:

$$V\left(\hat{\bar{Y}}_{lr2(pps)}\right) \cong V\left(\bar{y}_{lr(pps)}\right) - \frac{2\lambda n}{N-1}\Big[\left(c_1 + \beta c_2\right)\Big\{\left(u_{\max} - u_{\min}\right) + \beta\left(v_{\max} - v_{\min}\right)$$
$$-2n\left(c_1 + \beta c_2\right)\Big\}\Big]. \tag{31}$$

For optimum values of c_1 and c_2 i.e:

$c_{1opt} = \dfrac{\left(u_{\max} - u_{\min}\right)}{2n}$, $c_{2opt} = \dfrac{\left(v_{\max} - v_{\min}\right)}{2n}$,

the minimum variance of $\hat{\bar{Y}}_{lr2(pps)}$, is given by:

$$V\left(\hat{\bar{Y}}_{lr2(pps)}\right)_{\min} \cong V\left(\bar{y}_{lr(pps)}\right) - \frac{\lambda}{2(N-1)}\left[\left(u_{\max} - u_{\min}\right) + \beta\left(v_{\max} - v_{\min}\right)\right]. \tag{32}$$

In general, we can write:

$$V\left(\hat{\bar{Y}}_{lr(g)pps}\right)_{\min} \cong V\left(\bar{y}_{lr(pps)}\right) - \frac{\lambda}{2(N-1)}\left[\left(u_{\min} - u_{\min}\right) - |\beta|\left(v_{\max} - v_{\min}\right)\right]^2. \tag{33}$$

4. Comparison of Estimators

The proposed usual mean, ratio, product and regression types estimators $\hat{\bar{Y}}_{0(pps)}, \hat{\bar{Y}}_{R(pps)}, \hat{\bar{Y}}_{P(pps)}$ and $(\hat{\bar{Y}}_{lr1(pps)}, \hat{\bar{Y}}_{lr2(pps)})$ perform better than their counter parts i.e. $\bar{y}_{pps}$, $\bar{y}_{R(pps)}$, $\bar{y}_{P(pps)}$ and $\bar{y}_{lr(pps)}$.

Condition (i): By (4) and (14)

The proposed estimator $\hat{\bar{Y}}_{0(pps)}$ will perform better than usual mean per unit estimator (also suggested by Singh et al.,2017) if

$\left[V(\bar{y}_{pps})-V(\hat{\bar{Y}}_{0(pps)})_{\min}\right]\geq 0$, or if

$$\frac{\lambda\left(u_{\max}-u_{\min}\right)^{2}}{2\left(N-1\right)}\geq 0\,, \tag{34}$$

which is always true.

Condition (ii): By (9) and (25**)**

The proposed ratio type estimator $\hat{\bar{Y}}_{R(pps)}$ in PPS sampling will perform better than the traditional ratio type estimator $\bar{y}_{R(pps)}$ if

$\left[MSE\left(\bar{y}_{R(pps)}\right)-MSE\left(\hat{\bar{Y}}_{R(pps)}\right)_{\min}\right]\geq 0$ or if

$$\min\left[Rc_2,\ Rc_2-\left\{\frac{R\left(v_{\max}-v_{\min}\right)-\left(u_{\max}-u_{\min}\right)}{n}\right\}\right]$$

$$c_1<\max\left[Rc_2,\ Rc_2-\left\{\frac{R\left(v_{\max}-v_{\min}\right)-\left(u_{\max}-u_{\min}\right)}{n}\right\}\right] \tag{35}$$

Condition (iii): By (10) and (28)

The proposed product type estimator $\hat{\bar{Y}}_{P(pps)}$ will perform better than the usual product type estimator $\bar{y}_{P(pps)}$ if

$\left[MSE\left(\bar{y}_{P(pps)}\right)-MSE\left(\hat{\bar{Y}}_{P(pps)}\right)_{\min}\right]\geq 0$ or if

$$\min\left[-Rc_2,\ -Rc_2+\left\{\frac{R\left(v_{\max}-v_{\min}\right)+\left(u_{\max}-u_{\min}\right)}{n}\right\}\right]$$

$$<c_1<\max\left[\frac{-Rc_2,\ -Rc_2+\left\{R\left(v_{\max}-v_{\min}\right)+\left(u_{\max}-u_{\min}\right)\right\}}{n}\right] \tag{36}$$

Condition (iv): By (12) and (30)

The proposed regression type estimator (positive correlation) will perform better than usual regression estimator $\bar{y}_{lr(pps)}$ if

$$\left[V\left(\bar{y}_{lr(pps)}\right)-V\left(\hat{\bar{Y}}_{lr1(pps)}\right)\right]\geq 0 \text{ or if}$$

$$\min\left[\beta c_2,\ \beta c_2-\left\{\frac{\beta\left(v_{\max}-v_{\min}\right)-\left(u_{\max}-u_{\min}\right)}{n}\right\}\right]$$

$$<c_1<\max\left[\beta c_2,\ \beta c_2-\left\{\frac{\beta\left(v_{\max}-v_{\min}\right)-\left(u_{\max}-u_{\min}\right)}{n}\right\}\right] \quad (37)$$

Condition (v): By (12) and (32)

The proposed regression type estimator will perform better than the usual regression type estimator $\bar{y}_{lr(pps)}$ if

$$\left[V\left(\bar{y}_{lr(pps)}\right)-V\left(\hat{\bar{Y}}_{lr2(pps)}\right)\right]\geq 0 \text{ or if}$$

$$\min\left[-\beta c_2,\ \beta c_2+\left\{\frac{\beta\left(v_{\max}-v_{\min}\right)-\left(u_{\max}-u_{\min}\right)}{n}\right\}\right]$$

$$<c_1<\max\left[-\beta c_2,\ \beta c_2+\left\{\frac{\beta\left(v_{\max}-v_{\min}\right)-\left(u_{\max}-u_{\min}\right)}{n}\right\}\right]. \quad (38)$$

We observed that the proposed estimators perform better than the existing estimators if above condition (i)-(v) are satisfied.

5. Numerical Study

In this section, we use the four different populations for numerical comparisons of the estimators.

Population 1: [Source: Murthy (1967)].
y =Output for 80 factories in a region,
x=Fixed capital in region,
z=Number of workers.

$$N=80,\ n=8\ \bar{Y}=5182.637,\ \bar{X}=1126.463,\ \bar{Z}=285.125,$$
$$S_u^2=10568817,\ C_u^2=0.6011057,\ S_v^2=63257.12,\ S_v=251.5097,$$
$$C_v=0.2818878,\ C_V^2=0.07946076,\ r_{uv}=0.5800092,\ S_{uv}=740038.3,$$
$$C_{uv}=0.126713,\ u_{\max}=15586.83,\ u_{\min}=2408.59,\ v_{\max}=1944.034,$$
$$v_{\min}=592.6127.$$

Population 2: [Source: Campilho, Engineering Faculty, Oporto University, http://archive.ics.uci.edu/ml/datasets/cork stoppers].
y =Total number of defects,
x=Total perimeter of the defects(in pixels),
z =Total area of the defects (in pixels).

$N=150,\ n=20,\ \bar{Y}=78.47333,\ \bar{X}=710.3867,\ \bar{Z}=324.033,$
$S_u^2=954.0456,\ C_u=0.4562833,\ C_u^2=0.2081944,\ S_v^2=10950.96,$
$C_v=0.1638402,\ C_v^2=0.0268436,\ r_{uv}=0.7431098,\ S_{uv}=3096.833,$
$u_{\max}=176.1051,\ u_{\min}=35.93486,\ v_{\max}=1028.454,\ v_{\min}=493.175.$

Population 3: [Source: Singh (2003)]
y = Number of estimated fish caught during 1995,
x= Number of estimated fish caught during 1994,
z = Number of estimated fish caught during 1994.

$N=80,\ n=8,\ \bar{Y}=4514.899,\ \bar{X}=4954.435,\ \bar{Z}=4591.072,$
$S_u^2=2420387,\ S_u=1555.759,\ C_u=0.4720461,\ C_u^2=0.2228275,$
$S_v^2=2007238,\ S_v=1416.77,\ C_v=0.50490751,\ C_v^2=0.2549316,$
$r_{uv}=0.2660536,\ S_{uv}=1418429,\ C_{uv}=0.06341113,\ u_{\max}=11873.89,$
$u_{\min}=467.0002,\ v_{\max}=18850.12,\ v_{\min}=894.5356.$

Population 4: [Source:Paulo Herbert, DEQ, Faculty of Engineering, Porto University, Portugal]
http://archive.ics.uci.edu/ml/datasets/wine.
y = Aspartame,
x= Leucine,
z = Isoleucine.

$N=67,\ n=8,\ \bar{Y}=20.59851,\ \bar{Z}=9.792537,\ S_u^2=93.83272,$
$C_u=0.6072445,\ C_u^2=0.3687459,\ S_v^2=107.6524,\ C_v=0.7587807,$
$C_v^2=0.575748,\ r_{uv}=0.3171296,\ S_{uv}=71.13702,\ C_{uv}=0.1461223,$
$u_{\max}=122.2254,\ u_{\min}=5.092119,\ v_{\max}=131.6552,\ v_{\min}=0.$

Estimators	Population 1	Population 2	Population 3	Population 4
$\bar{y}_{pps} = \hat{\bar{Y}}_{0(pps)}$	2018192	64.103	567773.2	25.7483
$\bar{y}_{R(pps)}$	1433786	38.159	894200.7	45.5418
$\hat{\bar{Y}}_{R(pps)}$	1395455	37.057	871627.7	44.4519
$\bar{y}_{P(pps)}$	3136173	106.579	1540497	86.3524
$\hat{\bar{Y}}_{P(pps)}$	3097842	105.477	1517924	85.2625
$\bar{y}_{lr(pps)}$	876669	21.361	281132	10.5494
$\hat{\bar{Y}}_{lr(g)pps}$	838338	20.259	258570	9.4596

Table 1: MSE of the different estimators

In Table 1, we observed that MSE values of the suggested estimators are smaller than the existing counterparts estimators for all 4 data sets. The regression estimator $\hat{\bar{Y}}_{lr(g)pps}$ outperform, among all other considered estimators and are preferable.

6. Conclusion

We proposed some usual mean, ratio, product and regression type estimators under PPS sampling scheme when using maximum and minimum values. The proposed estimators under certain conditions are more efficient than the usual mean ratio, product and regression types estimators. In Table 1, we observed that the performances of the suggested estimators are better than the usual estimators in all 4 populations. Thus the proposed estimators may be preferred over the existing estimators. Among all estimators, the performances of the regression estimators are the best.

Acknowledgement

The authors are thankful to the learned referees for their valuable suggestions.

References

1. Agarwal, S. and Kumar, P. (1980). Combination of ratio and pps estimators, Journal of Indian Society of Agricultural Statistics, 32, p. 81–86.
2. Al-Hossain, A. Y. and Khan, M. (2014). Efficiency of ratio, product, and regression estimators under maximum and minimum values, using two auxiliary variables, Journal of Applied Mathematics, 2014, Article ID 693782, 6 pages.
3. Grover, L. K. and Kaur, P. (2011). An improved estimator of the finite population mean in simple random sampling. Model Assisted Statistics and Applications, 6(1), p. 47–55.
4. Gupta, S. and Shabbir, J. (2008). On improvement in estimating the population mean in simple random sampling, Journal of Applied Statistics, 35(5), p. 559–566.

5. Haq, A. and Shabbir, J. (2013). Improved family of ratio estimators in simple and stratified random sampling,m Communications in Statistics-Theory and Methods, 42(5), p.782–799.
6. Haq, A., Shabbir, J., and Gupta, S. (2013). Improved exponential ratio type estimators in stratified sampling, Pakistan Journal of Statistics, 29(1), p. 13-32.
7. Kadilar, C. and Cingi, H. (2004). Ratio estimators in simple random sampling, Applied Mathematics and Computation, 151(3), p. 893–902.
8. Kadilar, C. and Cingi, H. (2006). Improvement in estimating the population mean in simple random sampling, Applied Mathematics Letters, 19(1), p. 75–79.
9. Khan, M. and Shabbir, J. (2013). Some improved ratio, product, and regression estimators of finite population mean when using minimum and maximum values, The Scientific World Journal, 2013, Article ID 431868, 7 pages.
10. Murthy, M. N. (1967). Sampling Theory and Methods, 2nd Edition, Statistical Publishing Society, Calcutta.
11. Pandey, S. and Singh, R. K. (1984). On combination of ratio and pps estimators, Biometrical Journal, 26(3), p. 333–336.
12. Rao, J. (1966). Alternative estimators in pps sampling for multiple characteristics, Sankhya: The Indian Journal of Statistics, Series A, 28, p. 47–60.
13. Sarndal, C. E. (1972). Sample survey theory vs. general statistical theory: Estimation of the population mean. International Statistical Review/Revue Internationale de Statistique, 1–12.
14. Singh, H. P., Mishra, A. C., and Pal, S. K. (2018). Improved estimator of population total in PPS sampling, Communications in Statistics-Theory and Methods, 47 (4), p. 912-934.
15. Singh, M. (1967). Ratio cum product method of estimation. Metrika, 12(1). p. 34–42.
16. Singh, S. (2003). Advanced Sampling Theory with Applications: How MichaelSelected Amy, Volume 2. Kluwer Academic Publishers.
17. Srivenkataramana, T., and Tracy, D.S. (1979). Transforming the study variate after PPS Sampling. Metron, 37(1), p.175-181.
18. Tripathi, T.P. (1968). A regression-type estimator in sampling with PPS and with replacement, Australian Journal of Statistics, 11(3), p.140-148.