Hacettepe Journal of Mathematics and Statistics
Volume 47 (2) (2018), 383 – 402

Using extreme values and fractional raw moments for mean estimation in stratified random sampling

Shoaib Ali*, Manzoor Khan† and Javid Shabbir ‡§

Abstract

Unusual observations can occur in sample survey data. Mean estimator is sensitive to very large and/or small values, if included in sample. It can provide biased results and ultimately, tempted to delete from the sample data. Extreme values, if known, can be retained in data and used as the auxiliary information to increase the precision of estimate. Similarly, a known auxiliary variable is always source of improvement in precision of estimates. A transformation can be used for the auxiliary variable to get even more precised estimates. In this article, we have suggested modified estimators for finite population mean when a sample is drawn under stratified random sampling design. We used extreme values and fractional raw moments of the auxiliary variable and suggested improved ratio, product and regression type estimators. By theoretical comparison, efficiency of proposed estimators is established and numerical and simulation studies are conducted to support the theoretical results.

Keywords: Use of Extreme Values, Fraction Raw Moments, Estimation of Mean, Stratified Random Sampling.

2000 AMS Classification: 62D05

Received : 27.11.2015 *Accepted* : 14.06.2016 *Doi* : 10.15672/HJMS.201613018798

* Department of Statistics, Faculty of Natural Sciences, Quaid-i-Azam University, Islamabad 45320, Pakistan. Email: ali.s001@hotmail.com

† Department of Statistics, Faculty of Natural Sciences, Quaid-i-Azam University, Islamabad 45320, Pakistan. Email: manzoor_hum01@yahoo.com

‡ Department of Statistics, Faculty of Natural Sciences, Quaid-i-Azam University, Islamabad 45320, Pakistan. Email: js@qau.edu.pk

§ Corresponding Author

1. Introduction

The purpose of survey sampling is to utilize the maximum information about the characteristic of interest. Many fields of study require estimation of the finite population mean for variable of interest. For example, average wheat production per acre, average income of households, mean weight of meat producing animals etc. Mean per unit estimator is base line estimator to estimate finite population mean.

When the variable of interest is dependent on an extraneous source, the variance of the estimator can be inflated. To avoid this problem and to get precise estimates, it is important to use the stratified random sampling. To improve the precision of estimates, use of the auxiliary information has been in practice. [2] was pioneer to use the auxiliary information for the estimation of population mean. It was established that when the study variable and the auxiliary variable are positively correlated then ratio estimator provides more efficient estimates as compared to sample mean estimator and if there is negative relationship between the study and the auxiliary variable then product estimator provides better estimate. When regression line between the study and the auxiliary variables does not pass through origin then regression estimator dominates over ratio and product estimators. In stratified random sampling, [3] proposed two different methods for constructing the ratio estimators. In our study we use both combined and separate estimators when using maximum and minimum values.

2. Sampling scheme

Let a population of size N is divided into L mutually exclusive strata of sizes $N_h(h = 1, 2, 3, ..., L)$ such that $\sum_{h=1}^{L} N_h = N$. Let Y_{hi} and X_{hi} be the values of the study and the auxiliary variables at i^{th} unit$(i = 1, 2..., N_h)$ in the h^{th} $(h = 1, 2, 3, ..., L)$ stratum respectively. Let a sample of size n_h $(h = 1, 2, 3, ..., L)$ is drawn from each stratum independently by simple random sampling without replacement (SRSWOR) such that $\sum_{h=1}^{L} n_h = n$, where n is total number of units in a sample. Let y_{hi} and x_{hi} be the values of the study and the auxiliary variables of the i^{th} unit $(i = 1, 2..., n_h)$ in a sample.
Define:
$\bar{Y} = \frac{\sum_{i=1}^{N} Y_i}{N}$, $\bar{X} = \frac{\sum_{i=1}^{N} X_i}{N} =$
Population means of the study and the auxiliary variables respectively,
$\bar{Y}_h = \frac{\sum_{i=1}^{N_h} Y_{hi}}{N_h}$, $\bar{y}_h = \frac{\sum_{i=1}^{n_h} y_{hi}}{n_h} =$
Population and sample means of the study variable in the h^{th} stratum,
$\bar{X}_h = \frac{\sum_{i=1}^{N_h} X_{hi}}{N_h}$, $\bar{x}_h = \frac{\sum_{i=1}^{n_h} x_{hi}}{n_h} =$
Population and sample means of the auxiliary variable in the h^{th} stratum,
$S_{hx}^2 = \frac{\sum_{i=1}^{N_h} (X_{hi} - \bar{X}_h)^2}{N_h - 1} =$
Population variance of the auxiliary variable in the h^{th} stratum,
$S_{hy}^2 = \frac{\sum_{i=1}^{N_h} (Y_{hi} - \bar{Y}_h)^2}{N_h - 1} =$ Population variance of the study variable in the h^{th} stratum,
$S_{hyx} = \frac{\sum_{i=1}^{N_h} (X_{hi} - \bar{X}_h)(Y_{hi} - \bar{Y}_h)}{N_h - 1} =$ Population covariance of Y and X in the h^{th} stratum,
$\beta_h = \frac{S_{hyx}}{S_{hx}^2} =$ Population regression coefficient for the h^{th} stratum,
$\beta_c = \frac{\sum_{h=1}^{L} W_h^2 (\frac{N_h - n_h}{N_h n_h}) S_{hyx}}{\sum_{h=1}^{L} W_h^2 (\frac{N_h - n_h}{N_h n_h}) S_{hx}^2} =$ Population regression coefficient across the strata,
$W_h = \frac{N_h}{N} =$ Stratum weight in the h^{th} stratum,
$f_h = \frac{n_h}{N_h} =$ Sampling fraction in the h^{th} stratum,
$R = \frac{\bar{Y}}{\bar{X}} =$ Population ratio,

$R_h = \frac{\bar{Y}_h}{\bar{X}_h}$ = Population ratio in the h^{th} stratum.
The mean per unit estimator and its variance under stratified random sampling, are given by

$$(2.1)\qquad \bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h; V(\bar{y}_{st}) = \sum_{h=1}^{L} W_h^2 \lambda_h S_{hy}^2, \quad \text{where} \quad \lambda_h = \frac{1-f_h}{n_h}.$$

Combined ratio, product and regression estimators of finite population mean ($\bar{Y}$) in stratified random sampling with their biases and mean square errors are given as follows:

$$(2.2)\qquad \hat{\bar{Y}}_{RC_0} = \frac{\bar{y}_{st}}{\bar{x}_{st}} \bar{X},$$

$$(2.3)\qquad \hat{\bar{Y}}_{PC_0} = \frac{\bar{y}_{st}\bar{x}_{st}}{\bar{X}},$$

$$(2.4)\qquad \hat{\bar{Y}}_{lrC_0} = \bar{y}_{st} + b_c(\bar{X} - \bar{x}_{st}),$$

where $b_c = \frac{\sum_{h=1}^{L} W_h^2 \lambda_h s_{hyx}}{\sum_{h=1}^{L} W_h^2 \lambda_h s_{hx}^2}$ is the combined sample regression coefficient, across the strata.

$$(2.5)\qquad B(\hat{\bar{Y}}_{RC_0}) \cong \sum_{h=1}^{L} W_h \lambda_h \frac{(RS_{hx}^2 - S_{hyx})}{\bar{X}},$$

$$(2.6)\qquad B(\hat{\bar{Y}}_{PC_0}) \cong \sum_{h=1}^{L} W_h \lambda_h \frac{S_{hyx}}{\bar{X}},$$

$$(2.7)\qquad B(\hat{\bar{Y}}_{lrC_0}) \cong -Cov(\bar{x}_{st}, b_c),$$

$$(2.8)\qquad MSE(\hat{\bar{Y}}_{RC_0}) \cong \sum_{h=1}^{L} W_h^2 \lambda_h (S_{hy}^2 + R^2 S_{hx}^2 - 2RS_{hyx}),$$

$$(2.9)\qquad MSE(\hat{\bar{Y}}_{PC_0}) \cong \sum_{h=1}^{L} W_h^2 \lambda_h (S_{hy}^2 + R^2 S_{hx}^2 + 2RS_{hyx}),$$

$$(2.10)\qquad MSE(\hat{\bar{Y}}_{lrC_0}) \cong \sum_{h=1}^{L} W_h^2 \lambda_h S_{hy}^2 (1 - \rho_c^2),$$

where $\rho_c = \frac{\sum_{h=1}^{L} W_h^2 \lambda_h S_{hyx}}{\sqrt{\sum_{h=1}^{L} W_h^2 \lambda_h S_{hy}^2 \sum_{h=1}^{L} W_h^2 \lambda_h S_{hx}^2}}$ is the population correlation coefficient between the study and the auxiliary variables.

Separate ratio, product and regression estimators for finite population mean in stratified random sampling are given as follows:

$$(2.11)\qquad \hat{\bar{Y}}_{RS_0} = \sum_{h=1}^{L} W_h \bar{y}_{hR}, \text{ where } \quad \bar{y}_{hR} = \frac{\bar{y}_h}{\bar{x}_h} \bar{X}_h$$

$$(2.12)\qquad \hat{\bar{Y}}_{PS_0} = \sum_{h=1}^{L} W_h \bar{y}_{hP}, \text{ where } \quad \bar{y}_{hP} = \frac{\bar{y}_h}{\bar{X}_h} \bar{x}_h$$

$$\hat{\bar{Y}}_{lrS_0} = \sum_{h=1}^{L} W_h \bar{y}_{hlr}, \text{ where } \bar{y}_{hlr} = \left\{\bar{y}_h + b_h(\bar{X}_h - \bar{x}_h)\right\} \text{ and } b_h = \frac{s_{hyx}}{s_{hx}^2}. \tag{2.13}$$

Biases and mean square errors for separate estimators are given as:

$$B(\hat{\bar{Y}}_{RS_0}) \cong \sum_{h=1}^{L} W_h \lambda_h \frac{1}{\bar{X}_h}(R_h S_{hx}^2 - S_{hyx}), \tag{2.14}$$

$$B(\hat{\bar{Y}}_{PS_0}) \cong \sum_{h=1}^{L} W_h \lambda_h \frac{1}{\bar{X}_h}(R_h S_{hx}^2 + S_{hyx}), \tag{2.15}$$

$$B(\hat{\bar{Y}}_{lrS_0}) \cong -\sum_{h=1}^{L} W_h Cov(\bar{x}_h, b_h), \tag{2.16}$$

$$MSE(\hat{\bar{Y}}_{RS_0}) \cong \sum_{h=1}^{L} W_h^2 \lambda_h (S_{hy}^2 + R_h^2 S_{hx}^2 - 2R_h S_{hyx}), \tag{2.17}$$

$$MSE(\hat{\bar{Y}}_{PS_0}) \cong \sum_{h=1}^{L} W_h^2 \lambda_h (S_{hy}^2 + R_h^2 S_{hx}^2 + 2R_h S_{hyx}), \tag{2.18}$$

$$MSE(\hat{\bar{Y}}_{lrS_0}) \cong \sum_{h=1}^{L} W_h^2 \lambda_h S_{hy}^2 (1 - \rho_h^2), \tag{2.19}$$

where $\rho_h = \frac{S_{hyx}}{S_{hy}S_{hx}}$ is the correlation coefficient between the study and auxiliary variables in the h^{th} stratum.

In many situations real data sets contain unusual large and/or small values. Various hybrid seed production companies introduce new varieties of seeds and also specify the range of production per acre that farmer would benefit from. Maximum and minimum values can easily be read from the specified ranges. For estimating average income of households, income of the richest persons (maximum) in a society is well-known, and that of poorest (minimum) can easily be assessed. Similarly in various surveys which are conducted 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 unusual values. In such situation, this estimator can produce misleading results if any of the unusual values is selected in the sample. [5] suggested an unbiased estimator to overcome this problem of extreme values in the data set. Let y_{max} and y_{min} be the maximum and minimum values in the data set respectively. The estimator defined by [5], is given by

$$\bar{y}_s = \begin{cases} \bar{y} + c, & \text{if sample contains } y_{min} \text{ but not } y_{max} \\ \bar{y} - c, & \text{if sample contains } y_{max} \text{ but not } y_{min} \\ \bar{y}, & \text{In other cases} \end{cases} \tag{2.20}$$

where c is an arbitrary constant. Variance of the estimator $\bar{y}_s$, is given by:

$$Var(\bar{y}_s) = \lambda S_y^2 - \frac{2\lambda nc}{N-1}(y_{max} - y_{min} - nc). \tag{2.21}$$

At the optimum value i.e. $c_{(opt)} = \frac{y_{max}-y_{min}}{2n}$, the minimum variance of $\bar{y}_s$, is given by:

$$(2.22) \qquad Var(\bar{y}_s)_{(min)} = Var(\bar{y}) - \frac{\lambda}{2(N-1)}(y_{max} - y_{min})^2.$$

[4] and [1] suggested ratio, product and regression type estimators using extreme values of data in simple random sampling using one and two auxiliary variables respectively.

3. Proposed estimators using extreme values of data

Let $y_{h_{max}}(y_{h_{min}})$ and $x_{h_{max}}(x_{h_{min}})$ be the maximum(minimum) values of the study and the auxiliary variables respectively in the stratum h. The proposed estimator for finite population mean $(\bar{Y})$ under stratified random sampling is given as follows

$$(3.1) \qquad \bar{y}_{st.c} = \sum_{h=1}^{L} W_h \bar{y}_{h_c},$$

where

$$\bar{y}_{hc} = \begin{cases} \bar{y}_h + c_h, & \text{if a sample from the } h^{th} \text{ stratum contains } y_{h_{min}} \text{ but not } y_{h_{max}}, \\ \bar{y}_h - c_h, & \text{if a sample from the } h^{th} \text{ stratum contains } y_{h_{max}} \text{ but not } y_{h_{min}}, \\ \bar{y}_h, & \text{in other cases,} \end{cases}$$

where $c_h(h = 1, 2, 3, ..., L)$ are arbitrary constants.

3.1. Proposed estimators when positive correlation between Y and X. When relationship between the study variable Y and the auxiliary variable X is positive then for large value of X in a sample, a large value of Y is expected to be selected in the sample. Similarly for small value of X in a sample, a small value of Y is expected to be selected. Therefore ratio estimators can be defined as:

Combined ratio estimator.

$$(3.2) \qquad \hat{\bar{Y}}_{RC_1} = \frac{\bar{y}_{st.c_{11}}}{\bar{x}_{st.c_{21}}}\bar{X}.$$

Separate ratio estimator.

$$(3.3) \qquad \hat{\bar{Y}}_{RS_1} = \sum_{h=1}^{L} W_h \frac{\bar{y}_{h.c_{11}}}{\bar{x}_{h.c_{21}}}\bar{X}_h.$$

Combined regression estimator.

$$(3.4) \qquad \hat{\bar{Y}}_{lrC_{11}} = \bar{y}_{st.c_{11}} + b_c(\bar{X} - \bar{x}_{st.c_{21}}).$$

Separate regression estimator.

$$(3.5) \qquad \hat{\bar{Y}}_{lrS_{11}} = \sum_{h=1}^{L} W_h\{\bar{y}_{h.c_{11}} + b_h(\bar{X}_h - \bar{x}_{h.c_{21}})\}.$$

In the estimators from (3.2) to (3.5); $\bar{y}_{st.c_{11}} = \sum_{h=1}^{L} W_h \bar{y}_{h.c_{11}}$ and $\bar{x}_{st.c_{21}} = \sum_{h=1}^{L} W_h \bar{x}_{h.c_{21}}$. Here, $\bar{y}_{h.c_{11}}$ and $\bar{x}_{h.c_{21}}$ are defined as:

$$(\bar{y}_{h.c_{11}}, \bar{x}_{h.c_{21}}) = \begin{cases} (\bar{y}_h + c_{1h}, \bar{x}_h + c_{2h}), & \text{if sample contains } y_{h_{min}} \text{ but not } y_{h_{max}}, \\ (\bar{y}_h - c_{1h}, \bar{x}_h - c_{2h}), & \text{if sample contains } y_{h_{max}} \text{ but not } y_{h_{min}}, \\ (\bar{y}_h, \bar{x}_h), & \text{in other cases,} \end{cases}$$

where $c_{1h}, c_{2h} (h = 1, 2, 3, ..., L)$ are arbitrary constants.

3.2. Proposed estimators when negative correlation between Y and X. When relationship between the study variable Y and the auxiliary variable X is negative then for large value of X in a sample, a small value of Y is expected to be selected in a sample. Similarly for small value of X in a sample, a large value of Y is expected to be selected. Therefore product estimators can be defined as:

Combined product estimator.

$$\hat{\bar{Y}}_{PC_1} = \frac{\bar{y}_{st.c_{12}}}{\bar{X}} \bar{x}_{st.c_{22}}. \tag{3.6}$$

Separate product estimator.

$$\hat{\bar{Y}}_{PS_1} = \sum_{h=1}^{L} \frac{\bar{y}_{h.c_{12}}}{\bar{X}_h} \bar{x}_{h.c_{22}}. \tag{3.7}$$

Combined regression estimator.

$$\hat{\bar{Y}}_{lrC_{12}} = \bar{y}_{st.c_{12}} + b_c(\bar{X} - \bar{x}_{st.c_{22}}). \tag{3.8}$$

Separate regression estimator.

$$\hat{\bar{Y}}_{lrS_{12}} = \sum_{h=1}^{L} W_h \{\bar{y}_{h.c_{12}} + b_h(\bar{X} - \bar{x}_{h.c_{22}})\}. \tag{3.9}$$

In the estimators from (3.6) to (3.9); $\bar{y}_{st.c_{11}} = \sum_{h=1}^{L} W_h \bar{y}_{h.c_{11}}$ and $\bar{x}_{st.c_{21}} = \sum_{h=1}^{L} W_h \bar{x}_{h.c_{21}}$. Here, $\bar{y}_{h.c_{12}}$ and $\bar{x}_{h.c_{22}}$ are defined as follows:

$$(\bar{y}_{h.c_{12}}, \bar{x}_{h.c_{22}}) = \begin{cases} (\bar{y}_h + c_{1h}, \bar{x}_h - c_{2h}), & \text{if sample contains } y_{h_{min}} \text{ but not } y_{h_{max}}, \\ (\bar{y}_h - c_{1h}, \bar{x}_h + c_{2h}), & \text{if sample contains } y_{h_{max}} \text{ but not } y_{h_{min}}, \\ (\bar{y}_h, \bar{x}_h), & \text{in other cases,} \end{cases}$$

where $c_{1h}, c_{2h} (h = 1, 2, 3, ..., L)$ are arbitrary constants.

4. Properties of the proposed estimators

4.1. Some useful results. To find out the biases and mean square errors of the proposed estimators, we prove following two theorems.

4.1. Theorem. *Let $n_h (h = 1, 2, 3, ..., L)$ is drawn from sub-populations $N_h (h = 1, 2, 3, ..., L)$ such that $\sum n_h = n$.*

- *The estimator $\bar{y}_{hc}$ is an unbiased estimator of population mean $\bar{Y}_h$ in the h^{th} stratum.*

- *Variance of $\bar{y}_{hc}$ is given as:*

$$(4.1) \qquad Var(\bar{y}_{hc}) = \lambda_h S_{hy}^2 - \frac{2n_h c_h \lambda_h}{N_h - 1}(y_{h_{max}} - y_{h_{min}} - n_h c_h).$$

Proof. To prove first and second parts of Theorem 4.1, see [5] and the results in (2.21) respectively. □

4.2. Theorem. *Let a sample of sizes $n_h (h = 1, 2, 3, ..., L)$ are drawn from bi-variate sub-populations $N_h (h = 1, 2, 3, ..., L)$ respectively which constitute a stratified random sample of size $n (= \sum_{h=1}^{L} n_h)$ from a population of size N.*

- *The covariance between $\bar{y}_{h.c_{11}}$ and $\bar{x}_{h.c_{21}}$, when they are positively correlated, is given as:*

$$(4.2) \qquad \begin{aligned} Cov(\bar{y}_{h.c_{11}}, \bar{x}_{h.c_{21}}) =& \lambda_h S_{hyx}^2 - \frac{\lambda_h n_h}{N_h - 1}\{c_{1h}(x_{h_{max}} - x_{h_{min}}) \\ &+ c_{2h}(y_{h_{max}} - y_{h_{min}}) - 2n_h c_{1h} c_{2h}\}. \end{aligned}$$

- *The covariance between $\bar{y}_{h.c_{12}}$ and $\bar{x}_{h.c_{22}}$, when they are negatively correlated, is given as:*

$$(4.3) \qquad \begin{aligned} Cov(\bar{y}_{h.c_{12}}, \bar{x}_{h.c_{22}}) =& \lambda_h S_{hyx}^2 - \frac{\lambda_h n_h}{N_h - 1}\{c_{1h}(x_{h_{max}} - x_{h_{min}}) \\ &+ c_{2h}(y_{h_{max}} - y_{h_{min}}) - 2n_h c_{1h} c_{2h}\}. \end{aligned}$$

Proof. Proof of the results in Theorem 4.2 can be easily derived using the results from [4] □

Using the results from Theorem 4.1, it can be shown that $\bar{y}_{st.c}$ is an unbiased estimator of population mean $\bar{Y}$.

$$E(\bar{y}_{st.c}) = E(\sum_{h=1}^{L} W_h \bar{y}_{hc}) = \sum_{h=1}^{L} W_h E(\bar{y}_{hc}) = \sum_{h=1}^{L} W_h \bar{Y}_h = \bar{Y}.$$

And expression for variance of $\bar{y}_{st.c}$ can be established using result in (4.1),

$$(4.4) \qquad \begin{aligned} Var(\bar{y}_{st.c}) &= Var(\sum_{h=1}^{L} W_h \bar{y}_{hc}) = \sum_{h=1}^{L} W_h^2 Var(\bar{y}_{hc}). \\ &= \sum_{h=1}^{L} W_h^2 \lambda_h \{S_{hy}^2 - \frac{2n_h c_h}{N_h - 1}(y_{h_{max}} - y_{h_{min}} - n_h c_h)\}. \end{aligned}$$

Differentiating $V(\bar{y}_{st.c})$ with respect to $c_h (h = 1, 2, 3, ..., L)$ respectively and equate to zero, we have

$$\frac{\partial Var(\bar{y}_{st.c})}{\partial c_h} = 0 \Rightarrow y_{h_{max}} - y_{h_{min}} - 2n_h c_h = 0.$$

Optimum value of the constant c_h can be obtained from its corresponding equation as follows

$$(4.5) \qquad c_{h_{opt}} = \frac{y_{h_{max}} - y_{h_{min}}}{2n_h} \quad \text{for} \quad (h = 1, 2, 3, ..., L).$$

The minimum variance $Var(\bar{y}_{st.c})$, is given by

$$(4.6) \qquad Var(\bar{y}_{st.c})_{min} = \sum_{h=1}^{L} W_h^2 \lambda_h \{S_{hy}^2 - \frac{1}{2(N_h - 1)}(y_{h_{max}} - y_{h_{min}})^2\},$$

or

$$(4.7)\qquad Var(\bar{y}_{st.c})_{min} = V(\bar{y}_{st}) - \sum_{h=1}^{L} W_h^2 \lambda_h \frac{1}{2(N_h-1)}(y_{h_{max}} - y_{h_{min}})^2.$$

$$e_{st.c_{1j}} = (\bar{y}_{st.c_{1j}} - \bar{Y})/\bar{Y}, \quad e_{st.c_{2j}} = (\bar{x}_{st.c_{2j}} - \bar{X})/\bar{X}, \quad j = 1,2$$
$$e_{h.c_{1j}} = (\bar{y}_{h.c_{1j}} - \bar{Y}_h)/\bar{Y}_h, \quad e_{h.c_{2j}} = (\bar{x}_{h.c_{2j}} - \bar{X}_h)/\bar{X}_h, \quad j = 1,2$$

To the first order approximation, we have

$$E(e_{st.c_{1j}}) = E(e_{st.c_{2j}}) = E(e_{h.c_{1j}}) = E(e_{h.c_{2j}}) = 0,$$
$$E(e^2_{st.c_{1j}}) = \frac{1}{\bar{Y}} \sum_{h=1}^{L} W_h^2 \lambda_h \{S_{hy}^2 - \frac{2n_h c_{1h}}{N_h-1}(y_{h_{max}} - y_{h_{min}} - n_h c_{1h})\},$$
$$E(e^2_{st.c_{2j}}) = \frac{1}{\bar{X}} \sum_{h=1}^{L} W_h^2 \lambda_h \{S_{hx}^2 - \frac{2n_h c_{2h}}{N_h-1}(x_{h_{max}} - x_{h_{min}} - n_h c_{2h})\},$$
$$E(e^2_{h.c_{1j}}) = \frac{\lambda_h}{\bar{Y}_h} \{S_{hy}^2 - \frac{2n_h c_{1h}}{N_h-1}(y_{h_{max}} - y_{h_{min}} - n_h c_{1h})\},$$
$$E(e^2_{h.c_{2j}}) = \frac{\lambda_h}{\bar{X}_h} \{S_{hx}^2 - \frac{2n_h c_{2h}}{N_h-1}(x_{h_{max}} - x_{h_{min}} - n_h c_{2h})\},$$

for $j = 1, 2$.

Using the results from the Theorem 4.2, we have following expressions for the expectation of the relative errors

$$E(e_{st.c_{1j}} \times e_{st.c_{2j}}) = \frac{1}{\bar{Y}\bar{X}} \sum_{h=1}^{L} W_h^2 \lambda_h [S_{hyx} - \frac{n_h}{N_h-1}\{c_{1h}(x_{h_{max}} - x_{h_{min}}) + c_{2h}(y_{h_{max}} - y_{h_{min}}) - 2n_h c_{1h} c_{2h}\}],$$
$$E(e_{h.c_{1j}} \times e_{h.c_{2j}}) = \frac{\lambda_h}{\bar{Y}_h \bar{X}_h} [S_{hyx} - \frac{n_h}{N_h-1}\{c_{1h}(x_{h_{max}} - x_{h_{min}}) + c_{2h}(y_{h_{max}} - y_{h_{min}}) - 2n_h c_{1h} c_{2h}\}],$$

for $j = 1, 2$.

4.2. Properties of combined estimators. Using (3.2), the combined ratio estimator $\hat{\bar{Y}}_{RC1}$ in terms of relative errors, we have

$$\hat{\bar{Y}}_{RC_1} = \bar{Y}(1 + e_{st.c_{11}})(1 + e_{st.c_{21}})^{-1},$$

Approximating up to first order, we have

$$(\hat{\bar{Y}}_{RC_1} - \bar{Y}) \cong \bar{Y}(e_{st.c_{11}} + e_{st.c_{21}} - e_{st.c_{11}} e_{st.c_{21}} + e^2_{st.c_{21}}),$$

The bias of $\hat{\bar{Y}}_{RC_1}$ is given by

$$(4.8)\qquad B(\hat{\bar{Y}}_{RC_1}) \cong \sum_{h=1}^{L} W_h^2 \frac{\lambda_h}{\bar{X}} \Big[R\{S_{hx}^2 - \frac{2n_h c_{2h}}{N_h-1}(x_{h_{max}} - x_{h_{min}} - n_h c_{2h})\} - \{S_{hyx} - \frac{n_h}{N_h-1}(c_{1h}(x_{h_{max}} - x_{h_{min}}) + c_{2h}(y_{h_{max}} - y_{h_{min}}) - 2n_h c_{1h} c_{2h})\}\Big],$$

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

The mean square error of $\hat{\bar{Y}}_{RC_1}$ up to first degree approximation is given by

$$MSE(\hat{\bar{Y}}_{RC_1}) \cong MSE(\hat{\bar{Y}}_{RC_0}) - \sum_{h=1}^{L} W_h^2 \frac{2\lambda_h n_h}{N_h - 1}(c_{1h} - Rc_{2h}) \times \{(y_{h_{max}} - y_{h_{min}}) - R(x_{h_{max}} - x_{h_{min}}) - n_h(c_{1h} - Rc_{2h})\}. \tag{4.9}$$

Differentiating $MSE(\hat{\bar{Y}}_{RC_1})$ with respect to c_{1h} and c_{2h} $(h = 1, 2, 3, ..., L)$ respectively and equate to zero, we have

$$\frac{\partial MSE(\hat{\bar{Y}}_{RC_1})}{\partial c_{1h}} = 0,$$
$$(y_{h_{max}} - y_{h_{min}}) - R(x_{h_{max}} - x_{h_{min}}) - 2n_h(c_{1h} - Rc_{2h}) = 0,$$

where $h = 1, 2, \ldots, L$.

Here we have L equations with $2L$ unknowns, therefore unique solution for the constants is not possible. We suppose that $c_{1h} = (y_{h_{max}} - y_{h_{min}})/2n_h$ and it implies that $c_{2h} = (x_{h_{max}} - x_{h_{min}})/2n_h$ where $(h = 1, 2, 3, ..., L)$. For the optimum values of c_{1h} and $c_{2h}(h = 1, 2, 3, ..., L)$, the mean square error of $(\hat{\bar{Y}}_{RC_1})$, is given by

$$MSE(\hat{\bar{Y}}_{RC_1})_{min} = MSE(\hat{\bar{Y}}_{RC_0}) - \sum_{h=1}^{L} W_h^2 \frac{\lambda_h}{2(N_h - 1)} \times \{(y_{h_{max}} - y_{h_{min}}) - R(x_{h_{max}} - x_{h_{min}})\}^2. \tag{4.10}$$

Similarly bias and mean square error of combined product estimator $\hat{\bar{Y}}_{PC_1}$, are give by

$$B(\hat{\bar{Y}}_{PC_1}) \cong \sum_{h=1}^{L} W_h^2 \frac{\lambda_h}{\bar{X}} \big[S_{hyx} - \frac{n_h}{N_h - 1} \times \{c_{1h}(x_{h_{max}} - x_{h_{min}}) + c_{2h}(y_{h_{max}} - y_{h_{min}}) - 2n_h c_{1h} c_{2h}\}\big]. \tag{4.11}$$

$$MSE(\hat{\bar{Y}}_{PC_1}) = MSE(\hat{\bar{Y}}_{PC_1}) - \sum_{h=1}^{L} W_h^2 \frac{2\lambda_h n_h}{N_h - 1}(c_{1h} - Rc_{2h}) \times \{(y_{h_{max}} - y_{h_{min}}) + R(x_{h_{max}} - x_{h_{min}}) - n_h(c_{1h} + Rc_{2h})\}. \tag{4.12}$$

Using optimum values of c_{1h} and c_{2h}, mean square error of $\hat{\bar{Y}}_{PC_1}$, is given by

$$MSE(\hat{\bar{Y}}_{PC_1})_{min} = MSE(\hat{\bar{Y}}_{PC_0}) - \sum_{h=1}^{L} W_h^2 \frac{\lambda_h}{2(N_h - 1)} \times \{(y_{h_{max}} - y_{h_{min}}) - R(x_{h_{max}} - x_{h_{min}})\}^2. \tag{4.13}$$

The bias and mean square error of combined regression estimator $\hat{\bar{Y}}_{lrC_{11}}$ in case positive correlation between Y and X are given by

$$B(\hat{\bar{Y}}_{lrC_{11}}) = -Cov(\bar{x}_{st.c_{21}}, b_c), \tag{4.14}$$

and

$$MSE(\hat{\bar{Y}}_{lrC_{11}}) = MSE(\hat{\bar{Y}}_{lrC_0}) - \sum_{h=1}^{L} W_h^2 \frac{2\lambda_h n_h}{N_h - 1}(c_{1h} - \beta_c c_{2h}) \times \{(y_{h_{max}} - y_{h_{min}}) - \beta_c(x_{h_{max}} - x_{h_{min}}) - n_h(c_{1h} - \beta_c c_{2h})\}, \tag{4.15}$$

where $\beta_c = Cov(\bar{y}_{st}, \bar{x}_{st})/Var(\bar{x}_{st})$ is population regression coefficient.

For optimum values of constants $c_{1h} = (y_{h_{max}} - y_{h_{min}})/2n_h$ and $c_{2h} = (x_{h_{max}} - x_{h_{min}})/2n_h$ $(h = 1, 2, 3, ..., L)$, mean square error of $\hat{\bar{Y}}_{lrC_{11}}$, is given by:

$$MSE(\hat{\bar{Y}}_{lrC_{11}})_{min} = MSE(\hat{\bar{Y}}_{lrC_0}) - \sum_{h=1}^{L} W_h^2 \frac{\lambda_h}{2(N_h - 1)}. \times \{(y_{h_{max}} - y_{h_{min}}) - \beta_c(x_{h_{max}} - x_{h_{min}})\}^2 \tag{4.16}$$

When there is negative correlation between Y and X, the variance of combined regression estimator $\hat{\bar{Y}}_{lrC2}$, is given by

$$MSE(\hat{\bar{Y}}_{lrC_{12}}) = MSE(\hat{\bar{Y}}_{lrC_0}) - \sum_{h=1}^{L} W_h^2 \frac{2\lambda_h n_h}{N_h - 1}(c_{1h} + \beta_c c_{2h}) \times \{(y_{h_{max}} - y_{h_{min}}) + \beta_c(x_{h_{max}} - x_{h_{min}}) - n_h(c_{1h} + \beta_c c_{2h})\}, \tag{4.17}$$

At optimum values of constants $c_{1h} = (y_{h_{max}} - y_{h_{min}})/2n_h$ and $c_{2h} = (x_{h_{max}} - x_{h_{min}})/2n_h$ for $(h = 1, 2, 3, ..., L)$. The *MSE* of $\hat{\bar{Y}}_{lrC_{12}}$, is given by

$$MSE(\hat{\bar{Y}}_{lrC_{12}})_{min} = MSE(\hat{\bar{Y}}_{lrC_0}) - \sum_{h=1}^{L} W_h^2 \frac{\lambda_h}{2(N_h - 1)} \times \{(y_{-} y_{h_{min}}) + \beta_c(x_{h_{max}} - x_{h_{min}})\}^2. \tag{4.18}$$

Using results for variances of combined regression estimators in Equations (4.16) and (4.18), the minimum mean square error of combined regression estimator can be written as

$$MSE(\hat{\bar{Y}}_{lrC_1})_{min} = MSE(\hat{\bar{Y}}_{lrC_0}) - \sum_{h=1}^{L} W_h^2 \frac{\lambda_h}{2(N_h - 1)} \times \{(y_{h_{max}} - y_{h_{min}}) - |\beta_c|\,(x_{h_{max}} - x_{h_{min}})\}^2. \tag{4.19}$$

4.3. Properties of separate estimators. Consider separate ratio estimator $\hat{\bar{Y}}_{RS_1}$ in term of $e's$

$$\hat{\bar{Y}}_{RS_1} = \sum_{h=1}^{L} W_h \bar{Y}_h (1 + e_{h.c_{11}})(1 + e_{h.c_{21}})^{-1}.$$

Expanding right hand side of above expression up to first order approximation, we have

$$(\hat{\bar{Y}}_{RS_1} - \bar{Y}) \cong \sum_{h=1}^{L} W_h \bar{Y}_h (e_{h.c_{11}} + e_{h.c_{21}} - e_{h.c_{11}} e_{h.c_{21}} + e_{h.c_{21}}^2),$$

The bias of $\hat{\bar{Y}}_{RS_1}$, is given by

$$B(\hat{\bar{Y}}_{RS_1}) \cong \sum_{h=1}^{L} W_h^2 \frac{\lambda_h}{\bar{X}_h} \Big[R_h\{S_{hx}^2 - \frac{2n_h c_h}{N_h - 1}(x_{h_{max}} - x_{h_{min}} - n_h c_{2h})\} - \{S_{hyx} - \frac{n_h}{N_h - 1}(c_{1h}(x_{h_{max}} - x_{h_{min}}) + c_{2h}(y_{h_{max}} - y_{h_{min}}) - 2n_h c_{1h} c_{2h})\}\Big], \tag{4.20}$$

where $R_h = \bar{Y}_h/\bar{X}_h$.
The mean square error of $\hat{\bar{Y}}_{RS_1}$, is given as

$$
\begin{aligned}
MSE(\hat{\bar{Y}}_{RS_1}) \cong & \sum_{h=1}^{L} W_h^2 \lambda_h \big[S_{hy}^2 - \frac{2n_h c_{1h}}{N_h - 1}(y_{h_{max}} - y_{h_{min}} - n_h c_{1h}) \\
& + R_h^2\{S_{hx}^2 - \frac{2n_h c_{2h}}{N_h - 1}(x_{h_{max}} - x_{h_{min}} - n_h c_{2h})\} \\
& - 2R_h\{S_{hyx} - \frac{n_h}{N_h - 1}(c_{1h}(x_{h_{max}} - x_{h_{min}}) + c_{2h}(y_{h_{max}} - y_{h_{min}}) \\
& - 2n_h c_{1h} c_{2h})\}\big],
\end{aligned} \tag{4.21}
$$

or

$$
\begin{aligned}
MSE(\hat{\bar{Y}}_{RS_1}) \cong & MSE(\hat{\bar{Y}}_{RS_0}) - \sum_{h=1}^{L} W_h^2 \frac{2\lambda_h n_h}{N_h - 1}(c_{1h} - R_h c_{2h}) \\
& \times \{(y_{h_{max}} - y_{h_{min}}) - R_h(x_{h_{max}} - x_{h_{min}}) - n_h(c_{1h} - R_h c_{2h})\}.
\end{aligned} \tag{4.22}
$$

Differentiating $MSE(\hat{\bar{Y}}_{RS_1})$ with respect to each constant and equate to zero. Optimum values of constants can be obtained from resulting equations as $c_{1h} = (y_{h_{max}} - y_{h_{min}})/2n_h$ and $c_{2h} = (x_{h_{max}} - x_{h_{min}})/2n_h$ and minimum mean square error of $\hat{\bar{Y}}_{RS_1}$, is given by

$$
\begin{aligned}
MSE(\hat{\bar{Y}}_{RS_1})_{min} \cong & MSE(\hat{\bar{Y}}_{RS_0}) - \sum_{h=1}^{L} W_h^2 \frac{\lambda_h}{2(N_h - 1)} \\
& \times \{(y_{h_{max}} - y_{h_{min}}) - R_h(x_{h_{max}} - x_{h_{min}})\}^2.
\end{aligned} \tag{4.23}
$$

Similarly bias and mean square error of separate product estimator, are given by

$$
\begin{aligned}
B(\hat{\bar{Y}}_{PS_1}) \cong & \sum_{h=1}^{L} W_h^2 \frac{\lambda_h}{\bar{X}_h} \big[S_{hyx} - \frac{n_h}{N_h - 1} \\
& \times \{c_{1h}(x_{h_{max}} - x_{h_{min}}) + c_{2h}(y_{h_{max}} - y_{h_{min}}) - 2n_h c_{1h} c_{2h}\}\big],
\end{aligned} \tag{4.24}
$$

$$
\begin{aligned}
MSE(\hat{\bar{Y}}_{PS_1}) \cong & MSE(\hat{\bar{Y}}_{PS_0}) - \sum_{h=1}^{L} W_h^2 \frac{2\lambda_h n_h}{N_h - 1}(c_{1h} + R_h c_{2h}) \\
& \times \{(y_{h_{max}} - y_{h_{min}}) + R_h(x_{h_{max}} - x_{h_{min}}) - n_h(c_{1h} + R_h c_{2h})\}.
\end{aligned} \tag{4.25}
$$

At optimum values of constants $c_{1h} = (y_{h_{max}} - y_{h_{min}})/2n_h$ and $c_{2h} = (x_{h_{max}} - x_{h_{min}})/2n_h$, minimum mean square error, is given by

$$
\begin{aligned}
MSE(\hat{\bar{Y}}_{PS_1})_{min} \cong & MSE(\hat{\bar{Y}}_{PS_0}) - \sum_{h=1}^{L} W_h^2 \frac{\lambda_h}{2(N_h - 1)} \\
& \times \{(y_{h_{max}} - y_{h_{min}}) + R_h(x_{h_{max}} - x_{h_{min}})\}^2.
\end{aligned} \tag{4.26}
$$

When there is positive correlation between Y_h and X_h, the bias and mean square error of separate regression estimator, are given by

$$
B(\hat{\bar{Y}}_{lrS_{11}}) = -\sum_{h=1}^{L} W_h Cov(\bar{x}_{h.c_{21}}, b_h), \tag{4.27}
$$

$$
\begin{aligned}
MSE(\hat{\bar{Y}}_{lrS_{11}}) \cong & MSE(\hat{\bar{Y}}_{lrS_0}) - \sum_{h=1}^{L} W_h^2 \frac{2\lambda_h n_h}{N_h - 1}(c_{1h} - \beta_h c_{2h}) \\
& \{(y_{h_{max}} - y_{h_{min}}) - \beta_h(x_{h_{max}} - x_{h_{min}}) - n_h(c_{1h} - \beta_h c_{2h})\},
\end{aligned} \tag{4.28}
$$

where $\beta_h = S_{hyx}/S_{hx}^2$ is population regression coefficient in the h^{th} stratum.
At optimum values of constant $c_{1h} = (y_{h_{max}} - y_{h_{min}})/2n_h$ and $c_{2h} = (x_{h_{max}} - x_{h_{min}})/2n_h$, minimum mean square error, is given by

$$MSE(\hat{\bar{Y}}_{lrS_{11}})_{min} \cong MSE(\hat{\bar{Y}}_{lrS_0}) - \sum_{h=1}^{L} W_h^2 \frac{\lambda_h}{2(N_h-1)} \times \{(y_{h_{max}} - y_{h_{min}}) - \beta_h(x_{h_{max}} - x_{h_{min}})\}^2. \tag{4.29}$$

The *MSE* of separate regression estimator in case of negative correlation between Y_h and X_h, is given by

$$MSE(\hat{\bar{Y}}_{lrS_{12}}) \cong MSE(\hat{\bar{Y}}_{lrS_0}) - \sum_{h=1}^{L} W_h^2 \frac{2\lambda_h n_h}{N_h-1}(c_{1h} + \beta_h c_{2h}) \{(y_{h_{max}} - y_{h_{min}}) + \beta_h(x_{h_{max}} - x_{h_{min}}) - n_h(c_{1h} + \beta_h c_{2h})\}, \tag{4.30}$$

At optimum values of constant $c_{1h} = (y_{h_{max}} - y_{h_{min}})/2n_h$ and $c_{2h} = (x_{h_{max}} - x_{h_{min}})/2n_h$, minimum mean square error of $\hat{\bar{Y}}_{lrS_{12}}$, is given by

$$MSE(\hat{\bar{Y}}_{lrS_{12}})_{min} \cong MSE(\hat{\bar{Y}}_{lrS_0}) - \sum_{h=1}^{L} W_h^2 \frac{\lambda_h}{2(N_h-1)} \times \{(y_{h_{max}} - y_{h_{min}}) + \beta_h(x_{h_{max}} - x_{h_{min}})\}^2. \tag{4.31}$$

Generally mean square error of separate regression estimator can be written as

$$MSE(\hat{\bar{Y}}_{lrS_1})_{min} \cong MSE(\hat{\bar{Y}}_{lrS_0}) - \sum_{h=1}^{L} W_h^2 \frac{\lambda_h}{2(N_h-1)} \times \{(y_{h_{max}} - y_{h_{min}}) - |\beta_h|\,(x_{h_{max}} - x_{h_{min}})\}^2. \tag{4.32}$$

4.3. Remark. It is obvious from MSE expressions of the proposed estimators based on the extreme values, given in equations (4.7) - (4.32), that the proposed estimators are having smaller MSE/variances than the existing estimators, mentioned in equations (2.1) - (2.19) along with their MSE/variances.

5. Proposed estimators using fractional raw moments

Generally auxiliary information is utilized to enhance the precision of estimates of finite population parameters. If population parameters of the auxiliary variable are known then some common estimators like ratio, product, regression and their modifications are used. When information of the auxiliary variable is known it can be transformed to get more precise estimates. Suppose that we have transformation on the auxiliary variable X in the form of raw moment, is given by

$$U_i = X_i^p, \quad \text{where } p > 0 \text{ and } (i = 1, 2, 3, ..., N) \quad \Rightarrow \bar{U} = \sum_{i=1}^{N} \frac{X_i^p}{N}. \tag{5.1}$$

Here further assume that $X \in \Re^+$.
In stratified random sampling after transformation, let U_{hi} and u_{hi} be the population and sample values of transformed auxiliary variable at the i^{th} unit in the h^{th} stratum respectively. Let $\bar{U}_h = \frac{\sum U_{hi}}{N_h}$ and $\bar{u}_h = \frac{\sum u_{hi}}{n_h}$ be the population and sample means, $S_{hu}^2 = \frac{\sum (U_{hi} - \bar{U}_h)^2}{N_h - 1}$ be the variance and $S_{hyu} = \frac{\sum (U_{hi} - \bar{U}_h)(Y_{hi} - \bar{Y}_h)}{N_h - 1}$ be the covariance

with transformed auxiliary variable. Stratified ratio, product and regression estimators along with their biases and mean square errors can be written as

$$(5.2)\qquad \hat{\bar{Y}}_{RC_2} = \frac{\bar{y}_{st}}{\bar{u}_{st}}\bar{U},$$

$$(5.3)\qquad B(\hat{\bar{Y}}_{RC_2}) = \sum_{h=1}^{L} W_h\lambda_h \frac{(R_u S_{hu}^2 - S_{hyu})}{\bar{U}},$$

$$(5.4)\qquad MSE(\hat{\bar{Y}}_{RC_2}) = \sum_{h=1}^{L} W_h^2\lambda_h (S_{hy}^2 + R_u^2 S_{hu}^2 - 2R_u S_{hyu}).$$

$$(5.5)\qquad \hat{\bar{Y}}_{PC_2} = \frac{\bar{y}_{st}\bar{u}_{st}}{\bar{U}},$$

$$(5.6)\qquad B(\hat{\bar{Y}}_{PC_2}) = \sum_{h=1}^{L} W_h\lambda_h \frac{S_{hyu}}{\bar{U}},$$

$$(5.7)\qquad MSE(\hat{\bar{Y}}_{PC_2}) = \sum_{h=1}^{L} W_h^2\lambda_h (S_{hy}^2 + R_u^2 S_{hu}^2 + 2R_u S_{hyu}).$$

$$(5.8)\qquad \hat{\bar{Y}}_{lrC_2} = \bar{y}_{st} + b_c(\bar{U} - \bar{u}_{st}),$$

where $b_c = \frac{\sum_{h=1}^{L} W_h^2\lambda_h s_{hyu}}{\sum_{h=1}^{L} W_h^2\lambda_h s_{hu}^2}$ is the combined sample regression coefficient and $R_u = \frac{\bar{Y}}{\bar{U}}$.

$$(5.9)\qquad B(\hat{\bar{Y}}_{lrC_2}) = -Cov(\bar{u}_{st}, b_c),$$

$$(5.10)\qquad MSE(\hat{\bar{Y}}_{lrC_2}) = \sum_{h=1}^{L} W_h^2\lambda_h S_{hy}^2 (1 - \rho_{yu}^2),$$

where $\rho_{yu} = \frac{\sum_{h=1}^{L} W_h^2\lambda_h S_{hyu}}{\sqrt{\sum_{h=1}^{L} W_h^2\lambda_h S_{hy}^2 \sum_{h=1}^{L} W_h^2\lambda_h S_{hu}^2}}$. Separate ratio, product and regression estimators for finite population mean in stratified random sampling, are given as follows:

$$(5.11)\qquad \hat{\bar{Y}}_{RS_2} = \sum_{h=1}^{L} W_h\bar{y}_{hR_u}, \quad \text{where} \quad \bar{y}_{hR_u} = \frac{\bar{y}_h}{\bar{u}_h}\bar{U}_h,$$

$$(5.12)\qquad \hat{\bar{Y}}_{PS_2} = \sum_{h=1}^{L} W_h\bar{y}_{hP_u}, \quad \text{where} \quad \bar{y}_{hP_u} = \frac{\bar{y}_h}{\bar{U}_h}\bar{u}_h,$$

$$(5.13)\qquad \hat{\bar{Y}}_{lrS_2} = \sum_{h=1}^{L} W_h\bar{y}_{hlr_u},$$

where $\bar{y}_{hlr_u} = \left\{\bar{y}_h + b_{hu}(\bar{U}_h - \bar{u}_h)\right\}$ and $b_h = \frac{s_{hyu}}{s_{hu}^2}$.
The biases and mean square errors for separate estimators are given as:

$$(5.14)\qquad B(\hat{\bar{Y}}_{RS_2}) = \sum_{h=1}^{L} W_h\lambda_h \frac{1}{\bar{U}_h}(R_{hu}S_{hu}^2 - S_{hyu}),$$

$$(5.15)\quad B(\hat{\bar{Y}}_{PS_2}) = \sum_{h=1}^{L} W_h \lambda_h \frac{1}{\bar{U}_h}(R_{hu}S_{hu}^2 + S_{hyu}),$$

$$(5.16)\quad B(\hat{\bar{Y}}_{lrS_2}) = -\sum_{h=1}^{L} W_h Cov(\bar{u}_h, b_h),$$

$$(5.17)\quad MSE(\hat{\bar{Y}}_{RS_2}) = \sum_{h=1}^{L} W_h^2 \lambda_h (S_{hy}^2 + R_{hu}^2 S_{hu}^2 - 2R_{hu}S_{hyu}),$$

$$(5.18)\quad MSE(\hat{\bar{Y}}_{PS_2}) = \sum_{h=1}^{L} W_h^2 \lambda_h (S_{hy}^2 + R_{hu}^2 S_{hu}^2 + 2R_{hu}S_{hyu}),$$

$$(5.19)\quad MSE(\hat{\bar{Y}}_{lrS_2}) = \sum_{h=1}^{L} W_h^2 \lambda_h S_{hy}^2 (1 - \rho_{hyu}^2),$$

where $R_{hu} = \frac{\bar{Y}_h}{\bar{U}_h}$ and $\rho_{hyu} = \frac{S_{hyu}}{S_{hy}S_{hu}}$.

Let $u_{h_{max}}$ and $x_{h_{min}}$ be the maximum and minimum values of transformed auxiliary variable respectively. The proposed estimators along with their mean square errors are given as follows

Combined ratio estimator.

$$(5.20)\quad \hat{\bar{Y}}_{RC_3} = \frac{\bar{y}_{st.c_{11}}}{\bar{u}_{st.c_{21}}} \bar{U},$$

$$(5.21)\quad \begin{aligned} MSE(\hat{\bar{Y}}_{RC_3})_{min} = MSE(\hat{\bar{Y}}_{RC_2}) - \sum_{h=1}^{L} W_h^2 \frac{\lambda_h}{2(N_h - 1)} \\ \times \{(y_{h_{max}} - y_{h_{min}}) - R_u(u_{h_{max}} - u_{h_{min}})\}^2. \end{aligned}$$

Separate ratio estimator.

$$(5.22)\quad \hat{\bar{Y}}_{RS_3} = \sum_{h=1}^{L} W_h \frac{\bar{y}_{h.c_{11}}}{\bar{u}_{h.c_{21}}} \bar{U}_h,$$

$$(5.23)\quad \begin{aligned} MSE(\hat{\bar{Y}}_{RS_3})_{min} = MSE(\hat{\bar{Y}}_{RS_2}) - \sum_{h=1}^{L} W_h^2 \frac{\lambda_h}{2(N_h - 1)} \\ \times \{(y_{h_{max}} - y_{h_{min}}) - R_{hu}(u_{h_{max}} - u_{h_{min}})\}^2. \end{aligned}$$

Combined product estimator.

$$(5.24)\quad \hat{\bar{Y}}_{PC_3} = \frac{\bar{y}_{st.c_{12}}}{\bar{U}} \bar{u}_{st.c_{22}},$$

$$(5.25)\quad \begin{aligned} MSE(\hat{\bar{Y}}_{PC_3})_{min} = MSE(\hat{\bar{Y}}_{PC_2}) - \sum_{h=1}^{L} W_h^2 \frac{\lambda_h}{2(N_h - 1)} \\ \times \{(y_{h_{max}} - y_{h_{min}}) + R_{hu}(u_{h_{max}} - u_{h_{min}})\}^2. \end{aligned}$$

Separate product estimator.

$$(5.26)\quad \hat{\bar{Y}}_{PS_3} = \sum_{h=1}^{L} \frac{\bar{y}_{h.c_{11}}}{\bar{U}_h} \bar{u}_{h.c_{21}},$$

$$(5.27)\quad MSE(\hat{\bar{Y}}_{PS_3})_{min} = MSE(\hat{\bar{Y}}_{PS_2}) - \sum_{h=1}^{L} W_h^2 \frac{\lambda_h}{2(N_h-1)} \times \{(y_{h_{max}} - y_{h_{min}}) + R_{hu}(u_{h_{max}} - u_{h_{min}})\}^2.$$

Combined regression estimator.

$$(5.28)\quad \hat{\bar{Y}}_{lrC_{31}} = \bar{y}_{st.c_{11}} + b_c(\bar{U} - \bar{u}_{st.c_{21}}),$$

$$(5.29)\quad MSE(\hat{\bar{Y}}_{lrC_3})_{min} = MSE(\hat{\bar{Y}}_{lrC_2}) - \sum_{h=1}^{L} W_h^2 \frac{\lambda_h}{2(N_h-1)} \times \{(y_{h_{max}} - y_{h_{min}}) - |\beta_c|\,(u_{h_{max}} - u_{h_{min}})\}^2.$$

Separate regression estimator.

$$(5.30)\quad \hat{\bar{Y}}_{lrS_{31}} = \sum_{h=1}^{L} W_h\{\bar{y}_{h.c_{11}} + b_h(\bar{U}_h - \bar{u}_{h.c_{21}})\},$$

$$(5.31)\quad MSE(\hat{\bar{Y}}_{lrS_3})_{min} = MSE(\bar{\bar{Y}}_{lrS_2}) - \sum_{h=1}^{L} W_h^2 \frac{\lambda_h}{2(N_h-1)}$$

$$(5.32)\quad \times \{(y_{h_{max}} - y_{h_{min}}) - |\beta_h|\,(u_{h_{max}} - u_{h_{min}})\}^2.$$

5.1. Remark. The results are obvious from the MSE expressions of proposed estimators based on fractional raw moments and the extreme values in equations (5.20) - (5.31) that the proposed estimators have smaller MSE/variances than the existing estimators in the equations (2.2) - (2.19) along with their MSE/variances.

6. Numerical study

We have conducted a numerical study for three real data sets. Percentage efficiencies of the proposed estimators relative to usual stratified estimator ($\bar{y}_{st}$) are calculated for the three data sets.

6.1. Population-I: [6, page 218]**.**

Y: Juice quantity per cane (grams)
X: Weight of cane (grams)

Table 1. Summary of Population-I

	N_h	$\bar{X}_h$	$\bar{Y}_h$	S_{hx}^2	S_{hy}^2	S_{hyx}	$y_{h_{max}}$	$y_{h_{min}}$	$x_{h_{max}}$	$x_{h_{min}}$
Stratum-1	6	366.67	135.00	2706.67	80.00	440.00	150	125	450	300
Stratum-2	12	310.83	99.17	1881.06	226.52	618.93	135	80	410	260
Stratum-3	7	317.14	80.71	2890.48	120.24	444.05	100	70	420	250

A stratified sample of size $n = 12$ is selected by using proportional allocation such that $n_1 = 3, n_2 = 6, n_3 = 3$.

6.2. Population-II: [6, page 194].

Y: Amount of pocket money spent by students (in rupees)
X: Annual income of students' parents (in '000 rupees)

Table 2. Summary of Population-II

	N_h	$\bar{X}_h$	$\bar{Y}_h$	S^2_{hx}	S^2_{hy}	S_{hyx}	$y_{h_{max}}$	$y_{h_{min}}$	$x_{h_{max}}$	$x_{h_{min}}$
Stratum-1	4	92.50	925.00	851.00	50833.33	6266.70	1250	750	135	70
Stratum-2	10	57.20	535.00	31.07	11694.44	486.67	700	350	66	50
Stratum-3	13	38.00	303.85	35.00	9775.64	445.83	450	150	47	28

Students are divided into three stratum: poor, middle and rich. A sample of $15(= n)$ students is selected by using proportional allocation. Sample sizes from the three stratum are $n_1 = 2, n_2 = 6, n_3 = 7$.

6.3. Population-III: [6, page 212].

Y: Area of leaf (in sq cm)
X: Weight of leaf (in mg)

Table 3. Summary of Population-III

	N_h	$\bar{X}_h$	$\bar{Y}_h$	S^2_{hx}	S^2_{hy}	S_{hyx}	$y_{h_{max}}$	$y_{h_{min}}$	$x_{h_{max}}$	$x_{h_{min}}$
Stratum-1	12	103.42	25.752	133.900	40.157	67.506	36.61	17.76	123	84
Stratum-2	13	110.92	28.940	66.244	30.334	41.034	41.07	21.00	130	101
Stratum-3	14	104.29	25.777	154.990	46.628	82.082	39.06	16.07	129	81

Using proportional allocation method, $n_1 = 6, n_2 = 7, n_3 = 7$ are sample sizes from the three strata which constitute a sample of size $n = 20$.

The results are given in Tables 4, 5, 6 and 7.

7. Results and discussion

In Table 4, percentage efficiencies of proposed ratio, product and regression estimators $\bar{y}_{st.c}$, $\hat{\bar{Y}}_{RC_1}$, $\hat{\bar{Y}}_{RS_1}$, $\hat{\bar{Y}}_{PC_1}$, $\hat{\bar{Y}}_{PS_1}$, $\hat{\bar{Y}}_{lrC_1}$ and $\hat{\bar{Y}}_{lrS_1}$, based on extreme values, relative to the usual stratified estimator $\bar{y}_{st}$ are given for three populations. Also percentage efficiencies of usual ratio, product and regression estimators $\hat{\bar{Y}}_{RC_0}, \hat{\bar{Y}}_{RS_0}, \hat{\bar{Y}}_{PC_0}, \hat{\bar{Y}}_{PS_0}, \hat{\bar{Y}}_{lrC_0}$ and $\hat{\bar{Y}}_{lrS_0}$ relative to the usual stratified estimator $\bar{y}_{st}$ are given for the same three populations. Results based on these data sets indicate that the all the proposed estimators outperform their respective competitors. As the study variable and auxiliary variable are positively correlated in the data sets for all three populations, therefore product estimators are not performing well.

In Tables 5, 6 and 7, for different values of p, percentage efficiencies of proposed estimators, using fractional moments of the auxiliary variable, relative to the usual stratified estimator $\bar{y}_{st}$ are given for Populations I, II and III respectively. For $p = 1$, the proposed estimators $\hat{\bar{Y}}_{RC_2}, \hat{\bar{Y}}_{RS_2}, \hat{\bar{Y}}_{PC_2}, \hat{\bar{Y}}_{PS_2}, \hat{\bar{Y}}_{lrC_2}, \hat{\bar{Y}}_{lrS_2}$ reduce to the usual estimators $\hat{\bar{Y}}_{RC_0}, \hat{\bar{Y}}_{RS_0}, \hat{\bar{Y}}_{PC_0}, \hat{\bar{Y}}_{PS_0}, \hat{\bar{Y}}_{lrC_0}, \hat{\bar{Y}}_{lrS_0}$ and the estimators $\hat{\bar{Y}}_{RC_3}, \hat{\bar{Y}}_{RS_3}, \hat{\bar{Y}}_{PC_3}, \hat{\bar{Y}}_{PS_3}, \hat{\bar{Y}}_{lrC_3}$

and $\hat{\bar{Y}}_{lrS_3}$ reduce to the previously proposed estimators in this article $\hat{\bar{Y}}_{RC_1}, \hat{\bar{Y}}_{RS_1}, \hat{\bar{Y}}_{PC_1}, \hat{\bar{Y}}_{PS_1}, \hat{\bar{Y}}_{lrC_1}, \hat{\bar{Y}}_{lrS_1}$.

In Tables 5 and 6, results show that as the value of p decreases, the percentage relative efficiencies for the proposed ratio and product estimators increase up to a specific value of p and then tend to decrease. But in Table 7, as the the value of p decreases, the relative efficiencies of the proposed estimators decrease too.

Table 4. Percent relative efficiencies of proposed estimators with respect to $\bar{y}_{st}$

Estimators	**Population-I**	**Population-II**	**Population-III**
$\bar{y}_{st}$	100.00	100.00	100.00
$\bar{y}_{st.c}$	270.80	286.08	182.43
Ratio Estimators			
$\hat{\bar{Y}}_{RC_0}$	228.41	337.58	272.61
$\hat{\bar{Y}}_{RC_1}$	381.39	447.58	412.59
$\hat{\bar{Y}}_{RS_0}$	243.52	295.16	270.21
$\hat{\bar{Y}}_{RS_1}$	407.12	423.42	408.82
Product Estimators			
$\hat{\bar{Y}}_{PC_0}$	21.84	29.55	49.40
$\hat{\bar{Y}}_{PC_1}$	78.09	115.88	96.92
$\hat{\bar{Y}}_{PS_0}$	22.32	27.57	49.66
$\hat{\bar{Y}}_{PS_1}$	77.81	112.72	97.43
Regression Estimators			
$\hat{\bar{Y}}_{lrC_0}$	201.28	284.28	820.46
$\hat{\bar{Y}}_{lrC_1}$	356.66	411.50	896.52
$\hat{\bar{Y}}_{lrS_0}$	532.92	426.33	853.88
$\hat{\bar{Y}}_{lrS_1}$	552.93	453.70	882.53

Table 5. Percent relative efficiencies, calculated from Population-I, for different values of p

Estimators	$p=2$	$p=1.5$	$p=1$	$p=0.90$	$p=0.80$	$p=0.70$	$p=0.60$	$p=0.50$	$p=0.40$	$p=0.30$
$\bar{y}_{st}$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
$\hat{\bar{Y}}_{RC_2}$	31.17	73.83	228.41	279.84	323.35	341.60	325.69	284.57	235.23	189.58
$\hat{\bar{Y}}_{RS_2}$	34.23	80.96	243.52	293.05	293.05	331.06	341.95	320.79	278.33	230.15
$\hat{\bar{Y}}_{PC_2}$	8.46	13.00	21.84	24.58	24.58	27.83	31.71	36.39	42.09	49.09
$\hat{\bar{Y}}_{PS_2}$	8.71	13.33	22.32	25.10	25.10	28.38	32.30	37.02	42.75	49.76
$\hat{\bar{Y}}_{lrC_2}$	204.18	203.61	201.28	200.60	200.60	199.85	199.03	198.15	197.20	196.19
$\hat{\bar{Y}}_{lrS_2}$	579.94	559.39	532.92	527.08	527.08	521.10	514.99	508.77	502.46	496.06
$\bar{y}_{st.c}$	270.80	270.80	270.80	270.80	270.80	270.80	270.80	270.80	270.80	270.80
$\hat{\bar{Y}}_{RC_3}$	97.95	193.99	381.39	419.64	419.64	449.08	464.93	464.49	448.27	419.73
$\hat{\bar{Y}}_{RS_3}$	101.09	204.60	407.12	446.38	446.38	474.63	487.01	481.52	459.89	426.64
$\hat{\bar{Y}}_{PC_3}$	31.21	47.58	78.09	87.16	87.16	97.68	109.90	124.15	140.77	160.17
$\hat{\bar{Y}}_{PS_3}$	30.62	47.06	77.81	86.96	86.96	97.54	109.84	124.15	140.82	160.24
$\hat{\bar{Y}}_{lrC_3}$	379.99	371.31	356.66	353.10	353.10	349.35	345.44	341.37	337.16	332.83
$\hat{\bar{Y}}_{lrS_3}$	597.97	578.26	552.93	547.36	547.36	541.65	535.82	529.90	523.89	517.82

Table 6. Percent relative efficiencies, calculated from Population-II, for different values of p

Estimators	$p=2$	$p=1.5$	$p=1$	$p=0.90$	$p=0.80$	$p=0.70$	$p=0.60$	$p=0.50$	$p=0.40$	$p=0.30$
$\bar{y}_{st}$	100.00	100.00	100	100	100	100	100	100	100	100
$\hat{\bar{Y}}_{RC_2}$	20.22	82.25	337.58	351.94	327.72	284.31	239.43	200.86	169.99	145.83
$\hat{\bar{Y}}_{RS_2}$	53.15	123.93	295.16	308.45	320.80	311.57	284.75	249.01	212.16	178.72
$\hat{\bar{Y}}_{PC_2}$	7.15	14.62	29.547	31.62	36.12	41.16	46.76	52.96	59.82	67.38
$\hat{\bar{Y}}_{PS_2}$	11.44	17.05	27.575	29.09	32.46	36.39	40.98	46.37	52.73	60.26
$\hat{\bar{Y}}_{lrC_2}$	150.43	197.97	284.28	294.80	316.01	336.57	355.13	370.20	380.39	384.66
$\hat{\bar{Y}}_{lrS_2}$	430.10	429.51	426.33	425.86	424.82	423.66	422.38	420.97	419.44	417.78
$\bar{y}_{st.c}$	286.08	286.08	286.08	286.08	286.08	286.08	286.08	286.08	286.08	286.08
$\hat{\bar{Y}}_{RC_3}$	73.92	219.67	447.58	457.88	466.66	461.46	445.86	423.77	398.40	372.01
$\hat{\bar{Y}}_{RS_3}$	148.81	272.90	423.42	431.98	442.20	442.77	434.27	418.35	397.15	372.86
$\hat{\bar{Y}}_{PC_3}$	32.46	64.15	115.88	122.16	135.34	149.34	164.21	180.01	196.81	214.69
$\hat{\bar{Y}}_{PS_3}$	46.91	71.08	112.72	118.26	130.24	143.51	158.16	174.24	191.78	210.75
$\hat{\bar{Y}}_{lrC_3}$	298.79	353.49	411.5	416.06	423.88	429.68	433.21	434.32	433.00	429.36
$\hat{\bar{Y}}_{lrS_3}$	459.95	458.12	453.7	453.09	451.80	450.39	448.84	447.18	445.39	443.47

Table 7. Percent relative efficiencies, calculated from Population-III, for different values of p

Estimators	$p=2$	$p=1.5$	$p=1$	$p=0.90$	$p=0.80$	$p=0.70$	$p=0.60$	$p=0.50$	$p=0.40$	$p=0.30$
$\bar{y}_{st}$	100.00	100.00	100	100	100	100	100	100	100	100
$\hat{\bar{Y}}_{RC_2}$	884.81	519.12	272.61	242.15	216.04	193.58	174.19	157.38	142.74	129.94
$\hat{\bar{Y}}_{RS_2}$	882.19	515.98	270.21	240.05	214.23	192.06	172.95	156.39	141.99	129.40
$\hat{\bar{Y}}_{PC_2}$	29.14	37.31	49.40	52.51	55.92	59.66	63.78	68.34	73.38	78.98
$\hat{\bar{Y}}_{PS_2}$	29.12	37.42	49.66	52.80	56.23	59.99	64.12	68.67	73.69	79.26
$\hat{\bar{Y}}_{lrC_2}$	918.56	877.40	820.46	807.71	794.62	781.25	767.65	753.86	739.94	725.92
$\hat{\bar{Y}}_{lrS_2}$	929.68	898.89	853.88	843.53	832.82	821.79	810.47	798.91	787.14	775.19
$\bar{y}_{st.c}$	182.43	182.43	182.43	182.43	182.43	182.43	182.43	182.43	182.43	182.43
$\hat{\bar{Y}}_{RC_3}$	934.99	654.94	412.59	376.74	344.60	315.84	290.10	267.04	246.37	227.80
$\hat{\bar{Y}}_{RS_3}$	933.72	650.31	408.82	373.38	341.68	313.35	288.02	265.37	245.08	226.87
$\hat{\bar{Y}}_{PC_3}$	58.82	74.46	96.92	102.56	108.68	115.33	122.57	130.46	139.07	148.49
$\hat{\bar{Y}}_{PS_3}$	58.76	74.68	97.43	103.12	109.28	115.96	123.21	131.09	139.67	149.01
$\hat{\bar{Y}}_{lrC_3}$	971.53	943.34	896.52	885.34	873.66	861.54	849.01	836.14	822.96	809.54
$\hat{\bar{Y}}_{lrS_3}$	963.62	931.03	882.53	871.34	859.76	847.83	835.60	823.10	810.38	797.49

8. Simulation study

Simulation study is important to evaluate performances of the suggested estimators by repeated sampling. In this section, three populations given in Section 6 are considered and different sample sizes ($20\%, 30\%$ and 40% of N) are selected from each population. It can be seen from Table 8 that sample sizes selected from Populations I , II and III are ($n = 5, 8, 10$), ($n = 5, 8, 11$) and ($n = 8, 12, 16$) respectively. The sampling process is repeated 1000 times and mean square errors of the estimators are calculated using these samples as follows

$$MSE(\hat{\theta}) = \frac{1}{1000}\sum_{k=1}^{1000}(\hat{\theta}_k - \theta)^2 \tag{8.1}$$

where $\hat{\theta}$ is estimator of the parameter θ. Overall, mean square errors of all the estimators decreases by increasing the sample sizes, however some results in Table 8 indicate that decrease in mean square errors of ratio and regression estimators may be slower for larger sample sizes. Furthermore, proposed estimators have smaller mean square errors than the existing estimators. Therefore, the proposed estimators perform better than the conventional estimators while handling the extreme values in the data.

Table 8. Mean square errors of the estimators based on simulation study

	Population-I			Population-II			Population-III		
Estimators	$n=5$	$n=8$	$n=10$	$n=5$	$n=8$	$n=11$	$n=8$	$n=12$	$n=16$
$\bar{y}_{st}$	59.10	22.16	12.90	3588.33	2275.59	1232.95	4.33	3.41	2.03
$\bar{y}_{st.c}$	35.29	13.18	7.00	2351.09	1023.06	683.83	3.19	2.77	1.59
Ratio Estimators									
$\hat{\bar{Y}}_{RC_0}$	11.15	4.63	4.30	1264.47	716.00	381.26	1.90	1.39	0.84
$\hat{\bar{Y}}_{RC_1}$	7.35	4.15	3.08	891.61	467.42	255.33	1.35	1.09	0.64
$\hat{\bar{Y}}_{RS_0}$	11.14	4.62	4.30	1264.03	715.36	382.16	1.90	1.39	0.84
$\hat{\bar{Y}}_{RS_1}$	7.36	4.15	3.08	891.82	467.16	256.25	1.34	1.10	0.64
Product Estimators									
$\hat{\bar{Y}}_{PC_0}$	229.39	80.87	49.28	10956.23	7640.00	3794.54	8.30	6.77	3.94
$\hat{\bar{Y}}_{PC_1}$	162.06	52.33	26.57	6094.73	2565.30	1762.97	6.20	5.55	3.12
$\hat{\bar{Y}}_{PS_0}$	229.32	80.83	49.26	10960.84	7634.88	3802.23	8.30	6.78	3.94
$\hat{\bar{Y}}_{PS_1}$	162.06	52.30	26.58	6089.40	2566.04	1765.51	6.20	5.56	3.12
Regression Estimators									
$\hat{\bar{Y}}_{lrC_0}$	10.55	4.43	4.18	1419.75	812.56	424.45	0.89	0.54	0.33
$\hat{\bar{Y}}_{lrC_1}$	6.60	3.81	3.01	983.50	498.93	272.55	0.56	0.37	0.23
$\hat{\bar{Y}}_{lrS_0}$	10.55	4.43	4.18	1419.27	811.76	424.95	0.89	0.54	0.33
$\hat{\bar{Y}}_{lrS_1}$	6.60	3.81	3.01	985.41	498.46	272.60	0.56	0.37	0.23

9. Conclusions

In the present study, it is established that when the study population contain extreme values (substantially large or small relative to the other values of data) then the proposed estimators can perform efficiently. Also it can be concluded that extreme values may be used to enhance the efficiency of the estimator. Therefore, the proposed estimators can be used in place of their competitor estimators in real life applications. It is shown that using fractional raw moments of the auxiliary variable, efficiency of ratio and product estimators can be improved by decreasing value of p up to a specific value for some populations.

References

[1] Al-Hossain, A. Y. and Khan, M. *Efficiency of ratio, product, and regression estimators under maximum and minimum values, using two auxiliary variables*, Journal of Applied Mathematics, 6 pages, 2014.

[2] Cochran, W. *The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce*, The Journal of Agricultural Science **30**(02), 26-275, 1940.

[3] Hansen, M. H., Hurwitz, W. N., and Gurney, M. *Problems and methods of the sample survey of business*, Journal of the American Statistical Association **41**(234), 173-189, 1946.

[4] Khan, M. and Shabbir, J. *Some improved ratio, product, and regression estimators of finite population mean when using minimum and maximum values*, The Scientific World Journal, 7 pages, 2013.

[5] Särndal, C.-E. *Sample survey theory vs. general statistical theory: Estimation of the population mean*, International Statistical Review/Revue Internationale de Statistique **40**, 1-12, 1972.

[6] Singh, R. and Mangat, N. S. *Elements of Survey Sampling*, Kluwer Academic Publishers, Netherlands, 1996.