A Difference-Cum-Exponential Type Estimator for Estimating the Population Mean Under Stratification

H.S. Jhajj and L. Kusam Lata

Abstract—In survey sampling, stratification is helpful in improving precision of estimators over simple random sampling in case of heterogeneous population. In the present paper, a difference-cum exponential type estimator of population mean under two phase stratified random sampling design has been proposed for the case of heterogeneous population. The expressions for bias and mean squared error of the proposed estimator have been obtained up to first order of approximation. It has been shown that proposed estimator is efficient than linear regression estimator under the same sampling design for some range of variation in the values of constants involved. The results obtained have also been illustrated numerically as well as graphically by taking data from the population considered in the literature.

Keywords—Auxiliary Variable, Efficiency, Mean Square Error, Population Mean, Stratified Random Sampling

I. INTRODUCTION

In survey sampling, whenever there is auxiliary information available, we want to utilize it in method of estimation to obtain the efficient estimators. When there is positive correlation between the auxiliary variable x and variable of interest y, ratio-type estimators are preferred. On the other hand in case of negative correlation between study variable and auxiliary variable, product-type estimators are used to estimate the population parameters.

But in general regression estimator is more efficient than the ratio and product estimators except when the regression line of study variable on auxiliary variable passes through the neighborhood of the origin. In this case both the ratio and regression estimators are almost equally efficient. For estimating the parameters of heterogeneous population, stratified random sampling design is generally preferred. In literature several estimators of population mean, variance have been defined by various authors using stratified random sampling. Singh et al. [5] suggested exponential ratio & product type estimators of population mean in stratified random sampling. Koyuncu and Kadilar [6] suggested family of estimators using stratified random sampling to estimate the population mean. Recently Malik and Singh [7] has defined multivariate ratio type estimators using geometric and harmonic means in stratified random sampling. They have shown that estimators based on geometric and harmonic mean are less biased than Olkin [1] and Singh [3] estimators under certain conditions. Motivated from the literature, In present paper, considering bivariate population consisting of highly correlated variables y and x, we propose a difference-cum-exponential type estimator for estimating the population mean of variable y using two phase stratified random sampling design. The comparison has been made with the existing ones theoretically as well as numerically.

II. NOTATIONS AND RESULTS

Let a population of size N is divided into k homogeneous strata of size $N_h, \ h=1,2,\ldots, k \text{ with } \sum_{h=1}^{k} N_h = N$ under two phase sampling design, first phase simple random sample of size $n'_h$ is drawn independently from the $h^{th}$ stratum of size $N_h$ and variable y and x are observed on the sample units. Now second phase simple random sample of size n is drawn from a first phase sample. Let $Y_{hi}$ and $X_{hi}$ denote the respective values of variables y and x on the $i^{th}; i=1,2,\ldots, N_h$ unit of the population in the $h^{th}$ stratum respectively. The corresponding small letters denote the values in the sample. Denoting

$$\bar{Y}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} Y_{hi} \quad \bar{X}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} X_{hi}$$

$$\bar{Y} = \frac{1}{N} \sum_{h=1}^{k} N_h \bar{Y}_h \quad \bar{X} = \frac{1}{N} \sum_{h=1}^{k} N_h \bar{X}_h$$

$$\bar{Y}_h = \frac{1}{n'_h} \sum_{i=1}^{n'_h} Y_{hi} \quad \bar{X}_h = \frac{1}{n'_h} \sum_{i=1}^{n'_h} X_{hi}$$

$$\bar{Y}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} Y_{hi} \quad \bar{X}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} X_{hi}$$

$$S_{yh}^2 = N_h - 1 \sum_{i=1}^{N_h} Y_{hi} - \bar{Y}_h^2$$

$$S_{xh}^2 = N_h - 1 \sum_{i=1}^{N_h} X_{hi} - \bar{X}_h^2$$

$$\rho_h = \frac{S_{yh}^2}{S_{yh} S_{xh}}$$
\[ S_{y_{sh}} = N_h - 1 \sum_{i=1}^{N_h} X_{ha} - \bar{X}_h \]

\[ C^2_{sh} = \frac{S^2_{y_{sh}}}{\bar{X}_h} \]

Where \( \rho_h \) is the correlation coefficient between \( x \) and \( y \) in the \( h \)th stratum.

Defining

\[ \varepsilon_0 = \frac{\bar{Y}_h}{\bar{Y}_h} - 1 \]
\[ \varepsilon_1 = \frac{\bar{Y}_h}{\bar{Y}_h} - 1 \]
\[ \varepsilon_2 = \frac{\bar{Y}_h}{\bar{Y}_h} - 1 \]
\[ \varepsilon_3 = \frac{\bar{Y}_h}{\bar{Y}_h} - 1 \]

Such that

\[ E \varepsilon_0 = E \varepsilon_1 = E \varepsilon_2 = E \varepsilon_3 = 0 \]

And

\[ E \varepsilon_0^2 = \left( \frac{1}{n_h} - \frac{1}{N_h} \right) C^2_{sh} \]
\[ E \varepsilon_1^2 = \left( \frac{1}{n_h} - \frac{1}{N_h} \right) C^2_{sh} \]
\[ E \varepsilon_2^2 = \left( \frac{1}{n_h} - \frac{1}{N_h} \right) C^2_{sh} \]
\[ E \varepsilon_3^2 = \left( \frac{1}{n_h} - \frac{1}{N_h} \right) C^2_{sh} \]

III. PROPOSED ESTIMATOR AND ITS RESULTS

For the case of heterogeneous population using the sampling design defined in section 2, we propose an estimator of population mean as

\[ \hat{Y}_st = \sum_{h=1}^{k} p_h \bar{Y}_h \left[ \frac{1 + \theta \bar{X}_h - \bar{Y}_h}{\bar{X}_h} \right] \left[ 1 - \theta \frac{\bar{X}_h - \bar{Y}_h}{\bar{X}_h} \right] \left[ \exp \left( \frac{\bar{X}_h - \bar{Y}_h}{\bar{X}_h} \right) \right] \]  

(3.1)

Here \( p_h = \frac{N_h}{N} \); \( \theta > 0 \) and \( \alpha_h \) are any real constants.

To obtain the bias and mean squared error of estimator \( \hat{Y}_st \), we expand \( \hat{Y}_st \) in terms of \( \varepsilon \)'s as

\[ \hat{Y}_st = \sum_{h=1}^{k} p_h \bar{Y}_h \left[ 1 + \varepsilon \right] \varepsilon \left[ \frac{1 - \theta \varepsilon - \varepsilon^2}{\varepsilon} \right] \left[ \exp \left( \frac{\varepsilon - \varepsilon^2}{\varepsilon} \right) \right] \]  

(3.2)

By substituting \( \bar{X}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} X_{ha} \), \( \bar{Y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} Y_{ha} \) and \( \varepsilon = \frac{X_{ha} - \bar{X}_h}{\bar{X}_h} \), we get

\[ \hat{Y}_st = \sum_{h=1}^{k} p_h \bar{Y}_h \left[ 1 + \varepsilon + \alpha \right] \left( \frac{1}{2 - \theta} \varepsilon - \frac{1}{2} \varepsilon^2 - \varepsilon^3 \right) \]  

(3.3)

By differentiating (3.3) w.r.t \( \alpha \) and equating it to zero, we get

\[ \alpha_0 = -\frac{1}{2} \left( 2 - \theta \right) \left( \frac{1}{2 - \theta} \right)^2 \]  

(3.4)

The expression (3.4) depends upon two unknown constants \( \alpha_h \) and \( \theta \); so keeping the value of \( \theta \) fixed, we differentiate (3.4) w.r.t \( \alpha_h \) and equating it to zero, we get

\[ \alpha_0 = \frac{C_{sh}}{C^2_{sh}} \]  

(3.5)

Substituting the optimum value of \( \alpha_h \) from (3.5) in (3.4), minimum mean square error of \( \hat{Y}_st \) obtained is

\[ \text{Min. MSE} \hat{Y}_st = \sum_{h=1}^{k} p_h \bar{Y}_h \left[ \frac{1 - \theta \varepsilon - \varepsilon^2}{\varepsilon} \right] \left[ \exp \left( \frac{\varepsilon - \varepsilon^2}{\varepsilon} \right) \right] \]  

(3.6)
From (3.6) we see that \( \text{MSE} \hat{Y}_{st} \) depends upon the value of \( \theta \), so choosing the appropriate value of \( \theta \), the \( \text{MSE} \hat{Y}_{st} \) can be decreased. The optimum value of \( \theta \) is equal to one and \( \text{min MSE} \) of \( \hat{Y}_{st} \) is given by

\[
\text{Min. MSE } \hat{Y}_{st} = \sum_{h=1}^{d} p_h^2 \left( \frac{1}{n_h} - \frac{1}{N_h} \right) C_{yh} \left( 1 - \theta^2 \right)
\]

Theorem 1: Upto first order of approximation, the bias of estimator \( \hat{Y}_{st} \) is

\[
\text{Bias } \hat{Y}_{st} = \sum_{h=1}^{d} p_h^2 \left[ \frac{1}{n_h} - \frac{1}{n'_h} \right] \alpha_h \left( 1 - \theta^2 \right)
\]

And its mean squared error is given by

\[
\text{MSE } \hat{Y}_{st} = \sum_{h=1}^{d} p_h^2 \left[ \frac{1}{n_h} - \frac{1}{n'_h} \right] C_{yh} \left( 1 - \theta^2 \right) \alpha_h^2 C_{yh}^2 - 2 \alpha_h C_{yh}
\]

Theorem 2: Upto first order of approximation, keeping the value of \( \theta \) fixed, the MSE of estimator \( \hat{Y}_{st} \) is minimized for

\[
\alpha_h \text{ opt} = \frac{C_{yh}}{C_{yh}^2}
\]

And its minimum value is given by

\[
\text{Min. MSE } \hat{Y}_{st} = \sum_{h=1}^{d} p_h^2 \left[ \frac{1}{n_h} - \frac{1}{n'_h} \right] C_{yh} \left( 1 - \theta^2 \right) \alpha_h^2 C_{yh} - 2 \alpha_h C_{yh}
\]

For optimum value at \( \theta = 1 \), the MSE of \( \hat{Y}_{st} \) is

\[
\text{Min. MSE } \hat{Y}_{st} = \sum_{h=1}^{d} p_h^2 \left[ \frac{1}{n_h} - \frac{1}{n'_h} \right] C_{yh} \left( 1 - \theta^2 \right)
\]

Theorem 3: Upto first order of approximation, the bias of optimum estimator of \( \hat{Y}_{st} \) is

\[
\text{Bias } \hat{Y}_{st} = \sum_{h=1}^{d} p_h^2 \left[ \frac{1}{n_h} - \frac{1}{n'_h} \right] \alpha_h \left( 1 - \theta^2 \right)
\]

\[
\left[ \frac{\alpha_h - 1}{2} + \frac{1}{1 - \theta} \right] C_{yh}^2 \rho_{yh} C_{yh} C_{yh}
\]

Which are function of \( \theta \), so for optimum value of \( \theta \) bias is given by

\[
\text{Bias } \hat{Y}_{st} \text{ opt} = 0
\]

IV. COMPARISON

For comparing the proposed estimator with the linear regression estimator in stratified random sampling under the considered sampling design, we first write the expression for mean square error of linear regression estimator (\( \hat{Y}_{lr} \)), upto first order of approximation

\[
\text{MSE } \hat{Y}_{lr} = \sum_{h=1}^{d} p_h^2 \left( \frac{1}{n_h} - \frac{1}{n'_h} \right) S_{yh} \left( 1 - \rho_{yh}^2 \right)
\]

(4.1)

Using (3.6) and (4.1) and after some algebra, we obtain

\[
\text{MSE } \hat{Y}_{lr} - \text{MSE}_{\text{min}} \hat{Y}_{st} = \sum_{h=1}^{d} p_h^2 \left( \frac{1}{n_h} - \frac{1}{n'_h} \right) S_{yh} \left( \theta^2 - 2 \theta \rho_{yh}^2 - 1 \right)
\]

(4.2)

The right hand side of (4.2) always holds true, which implies that

\[
\text{MSE } \hat{Y}_{lr} \geq \text{MSE}_{\text{min}} \hat{Y}_{st} \text{ if } 0 < \theta < 2
\]

(4.3)

From (4.3), we see that proposed estimator can be made efficient than the linear regression estimator by choosing any value of \( \theta \) lies between 0 and 2 and have maximum efficiency when \( \theta = 1 \).

V. EMPIRICAL STUDY

To take the rough idea about the gain in efficiency of the proposed estimator (\( \hat{Y}_{st} \)) over the linear regression estimator (\( \hat{Y}_{lr} \)) in stratified random sampling, we take the empirical population considered in literature (Source: Singh and Chaudhary 5, p-162). The values of population parameters obtained are given in Table 5.1. The mean square error and relative efficiency of proposed estimator (\( \hat{Y}_{st} \)) w.r.t linear
regression estimator \( (\hat{Y}_m) \) are given for some different values of \( \theta \) in table 5.2.

\[
y: \text{total number of trees, } x: \text{area under orchards in hectare}
\]

### Table 5.1: Values of Parameters

<table>
<thead>
<tr>
<th>Stratum number</th>
<th>Stratum Size ( N'_1 )</th>
<th>Sample Size ( n'_1 )</th>
<th>Population variance ( s^2_{p1} )</th>
<th>Correlation coefficient ( \rho_{p1} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>4</td>
<td>74775.467</td>
<td>0.92</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>5</td>
<td>259113.70</td>
<td>0.97</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>6</td>
<td>65885.60</td>
<td>0.88</td>
</tr>
<tr>
<td>Total</td>
<td>25</td>
<td>15</td>
<td>1237702</td>
<td>0.93</td>
</tr>
</tbody>
</table>

### Table 5.2: Mean Squared Error and Relative Efficiencies of Estimator

<table>
<thead>
<tr>
<th>( \theta )</th>
<th>MSE ((\bar{Y}_{ls}))</th>
<th>MSE ((\hat{Y}_{st}))</th>
<th>Efficiency ((\bar{Y}_{ls}))</th>
<th>Efficiency ((\hat{Y}_{st}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>3786.900</td>
<td>3786.900</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>0.5</td>
<td>3786.900</td>
<td>3432.20</td>
<td>100</td>
<td>110.3</td>
</tr>
<tr>
<td>1.0</td>
<td>3786.900</td>
<td>3314.35</td>
<td>100</td>
<td>114.3</td>
</tr>
<tr>
<td>1.5</td>
<td>3786.900</td>
<td>3432.20</td>
<td>100</td>
<td>110.3</td>
</tr>
<tr>
<td>2.0</td>
<td>3786.900</td>
<td>3786.900</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

---

### VI. CONCLUSION

From table 5.2, we observe that there is a significant gain in efficiency of proposed estimator \((\hat{Y}_{st})\) over the regression estimator \((\bar{Y}_{ls})\) for \(0 < \theta < 2\) under two phase stratified random sampling. The graphical representation also support that efficiency of proposed estimator \((\hat{Y}_{st})\) is more than the linear regression estimator \((\bar{Y}_{ls})\) under the same range of \( \theta \) and efficiency becomes maximum at \( \theta = 1 \). Hence we conclude that the proposed estimator should be recommended for estimation of population mean under stratified random sampling by choosing appropriate value of \( \theta \) between 0 and 2.

**REFERENCES**


---

**Dr. H.S.Jhajj** has done his Ph.D in Statistics from Punjabi University Patiala, Patiala. He is currently working as Professor, Department of Statistics, Punjabi University Patiala. He has working experience of more than 30 years at Punjabi University Patiala. Large number of research papers of Dr. Jhajj has been publishing in reputed journals from last many years. He has presented papers in seminars and conferences of National and International level. He has given a great contribution in development of sampling methods and new advancements in statistics.

**Mrs. Kusam Lata** has done her M.Sc Statistics with first class distinction at Department of Statistics, Punjabi University Patiala during 2005-2006. She is currently pursuing her Ph.D in Statistics in sampling theory and submitted her thesis under the guidance of Dr.H.S.Jhajj, Professor, Department of Statistics, Punjabi University Patiala. She has been awarded Gold medal for her performance in M.Sc Statistics from Punjabi University Patiala. She is currently working as Asst. Prof. at SDDCOE Panchkula.