I. INTRODUCTION

The use of auxiliary information at both selection and estimation stages to increase the efficiency of estimators has been employed with several improvements at both selection and estimation stages since the work of (Bahl, S., & Tuteja, R. 1991). Some estimation method that uses auxiliary variables includes ratio, product and regression estimators. Although, the first two methods give rise to biased estimators, the bias can be reduced by increasing the sample size. It is well known that to estimate any parameter, a suitable estimator is the corresponding statistic. Thus for estimating population mean, sample mean is the most appropriate estimator. Although it is unbiased, it has a large amount of variation. Therefore we seek an estimator which may be biased but has smaller mean squared error as compared to sample mean. This is achieved through the use of an auxiliary variable that has strong positive or negative correlation with the study variable. When there is strong positive correlation between the study variable and the auxiliary variable and the line of regression passes through origin, then the (Bahl, S., & Tuteja, R. 1991), estimator is used for improved estimation of population mean. Estimator by (Cochran, W. G. 1940) & (Kadilar, C., & ÇINGI, H. 2003), estimator are used when there is strong negative correlation. The regression type estimators are used for the improved estimation of population mean when the line of regression does not pass through the origin. This prompts the need for an appropriate transformation of the auxiliary variable to estimate the population total or mean of the study variable which may be asymptotically equal or close to the mean square error of the linear regression estimator. In light of the above shortcomings, several authors have studied to improve the existing classical ratio and product estimators to increase efficiency and also give better options in decision making. (Kadilar, C., & Cingi, H. 2004), (Murthy, M. N. 1964), (Robson, D. S. 1957), (Sharma, B., & Tailor, R. 2010), (Singh, B. K. et al., 2013) & (Singh, H. P., & Espejo, M. R. 2003), are amongst authors who have contributed in one way or the other to the modification of the classical ratio and product estimator for the population mean of the study variable under simple random sampling without replacement (SRSWOR) scheme with extensions to other sampling techniques.

II. NOTATIONS AND SOME EXISTING ESTIMATORS

Let \( U = \{U_1, \ldots, U_N\} \) be a finite population of size \( N \) and let \((y_i, x_i)\) be the values of the study variable \( Y \) and an auxiliary variable \( X \) on the \( i^{th} \) unit \( U_i, I = 1, \ldots, N \). Let a sample of size \( n \) be drawn from this population using a simple random sampling without replacement (SRSWOR). The notations used in this research are defined below:

\[
\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i \quad \text{: Population mean of the auxiliary variable}
\]

\[
\bar{Y} = \frac{1}{N} \sum_{i=1}^{N} Y_i \quad \text{: Population mean of the study variable}
\]
\[
\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i; \text{ Sample mean of the auxiliary variable}
\]

\[
\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i; \text{ Sample mean of the study variable}
\]

\[
C_x = \frac{S_x}{\bar{x}}; \text{ Coefficient of variation of the auxiliary variable}
\]

\[
C_y = \frac{S_y}{\bar{y}}; \text{ Coefficient of variation of the study variable}
\]

\[
C_{xy} = \frac{S_{xy}}{\bar{x}\bar{y}}; \text{ Coefficient of variation between the study variable and the auxiliary variable}
\]

\[
\lambda = \frac{1-f}{n} \text{ and;}
\]

\[
f = \frac{n}{N} \text{ is the sampling fraction}
\]

**Table 1: Some Existing Estimators With Their Mean Square Error (Mse)**

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Usual unbiased estimator</th>
<th>( V(\bar{y}) = \lambda C_y^2 \bar{Y}^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio estimator by (Cochran, W. G. 1940)</td>
<td>( T_{RI} = \bar{y}\left(\frac{\bar{X}}{x}\right) )</td>
<td>( MSE(T_{RI}) = \lambda \bar{Y}^2 \left( C_y^2 + C_x^2 - 2 \rho C_x C_y \right) )</td>
</tr>
<tr>
<td>Product estimator by (Murthy, M. N. 1964; &amp; Robson, D. S. 1957)</td>
<td>( T_{PI} = \bar{y}\left(\frac{\bar{x}}{\bar{X}}\right) )</td>
<td>( MSE(T_{PI}) = \lambda \bar{Y}^2 \left( C_y^2 + C_x^2 + 2 \rho C_x C_y \right) )</td>
</tr>
<tr>
<td>Chain ratio by (Sharma, B., &amp; Tailor, R. 2010)</td>
<td>( T_{chR} = y\left(\frac{\bar{X}}{x}\right)^2 )</td>
<td>( MSE(T_{chR}) = \lambda \bar{Y}^2 \left[ C_y^2 + 4 \left( C_x^2 - \rho C_x C_y \right) \right] )</td>
</tr>
<tr>
<td>Chain product</td>
<td>( T_{chP} = y\left(\frac{\bar{x}}{\bar{X}}\right)^2 )</td>
<td>( MSE(T_{chP}) = \lambda \bar{Y}^2 \left[ C_y^2 + 4 \left( C_x^2 + \rho C_x C_y \right) \right] )</td>
</tr>
<tr>
<td>Exponential ratio by (Bahl, S., &amp; Tuteja, R. 1991)</td>
<td>( T_{expR} = \bar{y}\exp\left(\frac{\bar{X}-\bar{x}}{x+\bar{X}}\right) )</td>
<td>( MSE(T_{expR}) = \lambda \bar{Y}^2 \left( C_y^2 + \frac{1}{4} C_x^2 - \rho C_x C_y \right) )</td>
</tr>
<tr>
<td>Exponential product by (Bahl, S., &amp; Tuteja, R. 1991)</td>
<td>( T_{expP} = \bar{y}\exp\left(\frac{\bar{x}-\bar{X}}{\bar{X}+x}\right) )</td>
<td>( MSE(T_{expP}) = \lambda \bar{Y}^2 \left( C_y^2 + \frac{1}{4} C_x^2 + \rho C_x C_y \right) )</td>
</tr>
</tbody>
</table>
III. PROPOSED ESTIMATOR

The proposed estimator is a linear combination of (Cochran, W. G. 1940; Murthy, M. N. 1964; & Robson, D. S. 1957) denoted by:

\[ T_{np} = y \left[ \alpha \left( \frac{\sqrt{X}}{\sqrt{x}} \right) + (1 - \alpha) \left( \frac{\sqrt{X}}{\sqrt{X}} \right) \right] \]  

(1)

Bias and Mean Square Error of (1)

Let \( \tilde{x} = \bar{X} (1 + e_i) \) and \( \tilde{y} = \bar{Y} (1 + e_0) \)

\[ E(e_0) = E(e_i) = 0, \ E(e_i^2) = \lambda C_y^2, \ E(e_i e_i) = \lambda C_x^2, \ E(e_0 e_i) = \rho C_x C_y, \text{ where } \lambda = \left( \frac{1 - f}{n} \right) \]

To obtain the Bias of the proposed estimator, we Substitute \( \tilde{y} \) and \( \tilde{x} \) in (Cochran, W. G. 1940)

\[ T_{np} = \bar{Y} (1 + e_0) \left[ \alpha \left( \frac{\sqrt{X}}{\sqrt{X (1 + e_i)}} \right) + (1 - \alpha) \left( \frac{\sqrt{X (1 + e_i)}}{\sqrt{X}} \right) \right] \]

\[ = \bar{Y} (1 + e_0) \left[ \alpha \left( 1 + e_i \right)^{1/2} + (1 - \alpha) \left( 1 + e_i \right)^{1/2} \right] \]  

(2)

Using Taylor’s series expansion up to second order approximation and assuming higher orders are negligible, we obtain:

\[ = \bar{Y} \left( 1 + e_0 + \frac{1}{2} e_i - \alpha e_i - \alpha e_i e_i + \frac{1}{2} e_0 e_i + \frac{1}{8} e_i^2 \right) \]  

(4)

The bias of the proposed estimator to its first order approximation is obtained from (4) as follows:

\[ \text{Bias} \left( T_{np} \right) = E \left( T_{np} - \bar{Y} \right) = \lambda \bar{Y} \left( \frac{1}{8} C_x^2 - \alpha \rho C_x C_y + \frac{1}{2} \rho C_x C_y \right) \]  

(5)

The expression for the MSE of the proposed estimator is also obtained from (4) as follows:

\[ MSE \left( T_{np} \right) = E \left( T_{np} - \bar{Y} \right)^2 \]

\[ = \bar{Y}^2 E \left( e_0^2 + \frac{1}{4} e_i^2 - \alpha e_i^2 + \alpha e_i e_i^2 - 2 \alpha e_0 e_i + e_0 e_i \right) \]

\[ = \lambda \bar{Y}^2 \left( C_y^2 + \frac{1}{4} C_x^2 - \alpha C_x^2 + \alpha^2 C_x^2 - 2 \alpha \rho C_x C_y + \rho C_x C_y \right) \]  

(6)

Optimality conditions for the proposed estimator

To get the optimal value of \( \alpha \) that will minimize the MSE, The partial derivative of (6) is taken with respect to \( \alpha \) equated to zero and the value of \( \alpha \) is obtained.

\[ \frac{\partial \text{MSE} \left( T_{np} \right)}{\partial \alpha} = \lambda \bar{Y}^2 \left( C_y^2 + \frac{1}{4} C_x^2 - \alpha C_x^2 + \alpha^2 C_x^2 - 2 \alpha \rho C_x C_y + \rho C_x C_y \right) = 0 \]

\[ \alpha = \frac{1}{2} + \rho \frac{C_y}{C_x} \]  

(7)

Substituting the value of \( \alpha \) in (6), we obtained the optimal MSE \( (T_{np}) \) as:

\[ MSE \left( T_{np} \right) = \lambda \bar{Y}^2 C_y^2 \left( 1 - \rho^2 \right) \]  

(8)

From (8), it is observed that the optimum MSE of the proposed estimator is the same as the MSE of the linear regression estimator.
IV. EFFICIENCY COMPARISON

The efficiency comparisons in this study are done using the mean square error of the proposed estimators and that of three existing estimators.

i. \( \text{MSE}(\bar{y}) < \text{MSE}(\bar{y}) \) if \( \lambda \bar{y}^2 C_y^2 \rho^2 > 0 \)

ii. \( \text{MSE}(\bar{y}) < \text{MSE}(T_{Ri}) \) if \( (C_x - \rho C_y)^2 > 0 \)

iii. \( \text{MSE}(\bar{y}) < \text{MSE}(T_{Pi}) \) if \( (C_x + \rho C_y)^2 > 0 \)

iv. \( \text{MSE}(\bar{y}) < \text{MSE}(T_{cbh}) \) if \( 4C_x^2 + \rho^2 C_y^2 - 4 \rho C_x C_y > 0 \)

v. \( \text{MSE}(\bar{y}) < \text{MSE}(T_{cbh}) \) if \( 4C_x^2 + \rho^2 C_y^2 + 4 \rho C_x C_y > 0 \)

vi. \( \text{MSE}(\bar{y}) < \text{MSE}(T_{bPR}) \) if \( \left(\frac{1}{2} C_x - \rho C_y\right)^2 > 0 \)

vii. \( \text{MSE}(\bar{y}) < \text{MSE}(T_{bPR}) \) if \( \left(\frac{1}{2} C_x + \rho C_y\right)^2 > 0 \)

VII. EMPIRICAL STUDY

In this section, we consider four (4) real data sets to numerically evaluate the performances of the proposed and the existing estimators considered here.

Population 1: [Source: Cochran (1977), pp. 196] Let \( y \) be the peach production in bushels in an orchard and \( x \) be the number of peach trees in the orchard in North Carolina in June 1946. The summary statistics for this data set are: \( N = 256, n = 100, \bar{Y} = 56.47, \bar{X} = 44.45, C_y = 1.42, C_x = 1.40, \rho_{yx} = 0.887. \)

Population 2: [Source: Murthy (1977), pp. 228] Let \( y \) be the output and \( x \) be the number of workers. The summary statistics for this data set are: \( N = 80, n = 10, \bar{Y} = 1.42, \bar{X} = 1.40, C_y = 0.7213, C_x = 0.3542, \rho_{yx} = 0.915. \)

Population 3: [Source: Das (1988)] Let \( y \) be the number of agricultural labourers for 1971 and \( x \) be the number of agricultural labourers for 1961. The summary statistics for this data set are: \( N = 278, n = 25, \bar{Y} = 39.068, \bar{X} = 25.111, C_y = 1.4451, C_x = 1.6198, \rho_{yx} = 0.7213. \)

Population 4: [Source: Steel, Torrie & Dickey (1960), pp. 282] Let \( y \) be the log of lef burn in sacs and \( x \) be the chlorine percentage. The summary statistics for this data set are: \( N = 70, n = 6, \bar{Y} = 0.6860, \bar{X} = 0.8077, C_y = 0.7001, C_x = 0.7493, \rho_{yx} = -0.4996. \)

| Table 2: The Percentage Relative Efficiency (PREs) of the Estimators |
|-----------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Population  | \( \bar{Y} \) | \( T_{RI} \) | \( T_{Pi} \) | \( T_{cbh} \) | \( T_{cbh} \) | \( T_{bPR} \) | \( T_{bPR} \) |
| 1          | 100          | 448.399       | 26.874         | 468.975         | 68.975           | 270.525         | 47.225          | 471.481         |
| 2          | 100          | 468.975       | 7.651          | 614.345         | 14.345           | 292.078         | 19.075          | 644.285         |
| 3          | 100          | 156.397       | 25.817         | 208.452         | 20.452           | 197.785         | 47.112          | 220.343         |
| 4          | 100          | 31.106        | 92.932         | 133.261         | 54.912           | 133.02           | 141.854         |

From the table above, it is shown that the proposed estimator \( T_{oup} \) has higher percentage relative efficiency than those estimators provided in this article.

VIII. CONCLUSION

The theoretical and empirical studies carried out reveals that the proposed estimators \( T_{oup} \) is better than those estimators considered in this article because the conditions are satisfied and has higher percentage relative efficiency, also for optimum value of \( \alpha \) the proposed estimators \( T_{oup} \) is equally efficient as the linear regression estimator. Hence the proposed estimators \( T_{oup} \) is recommended for its practical use for estimating population mean when the auxiliary information is available in survey sampling.

REFERENCES


