An Efficient Class of Two-Phase Exponential Ratio and Product – Type Estimators for Estimating the Finite Population Mean

Housila P. Singh, Neeraj Sharma and Tanveer A. Tarray*

School of Studies in Statistics, Vikram University, Ujjain, M.P., 456010 India.
*E-mail: tanveerstat@gmail.com

Received: 1 Dec. 2013, Revised: 5 Jan. 2014, Accepted: 9 Mar. 2014, Published online: 1 Jan. 2015

Abstract: This paper addresses the problem of estimating the population mean of the study variable using auxiliary information in double sampling. We have suggested a class of estimators of population mean in double sampling. It has been shown that the estimator due to Khatua and Mishra (2013) is a member of the proposed class. We have obtained the bias and mean square error (MSE) of the proposed class of estimators under large sample approximation. It is observed that the mean square error expression of the estimator due to Khatua and Mishra (2013) is not correct. So, we have obtained the correct expression of the mean square error of the estimator due to Khatua and Mishra (2013). We have compared the proposed class of estimators with that of usual unbiased estimator, usual double sampling ratio and product estimators, Singh and Vishwakarma (2007) estimators, usual regression estimator and usual double sampling ratio estimator and shown that the proposed estimator is better than existing estimators. Numerical illustrations are also given in support of the present study.

Keywords: Auxiliary variable, Exponential estimator, Mean squared error, Two-Phase sampling.

1. Introduction

In survey sampling, it is not uncommon to estimate the finite population mean \( \overline{Y} \) of the study variable of \( y \). When information on auxiliary variable \( x \), highly correlated with \( y \), is readily available on all units of population, it is well known that ratio estimator (for high positive correlation), product estimator (for high negative correlation) and regression estimator (for high correlation) can be used to increase the efficiency, incorporating the knowledge of population mean \( \overline{X} \). However, in certain practical situation when the population mean of auxiliary variable \( \overline{X} \) is not known a priori, the technique of two-phase sampling is successfully used in practice. The conventional method in such situation, a larger sample of size \( n' \) to furnish the good estimate of the population mean \( \overline{X} \) while a sub-sample size \( n \) is selected from the first phase sample 'n' to observe the characteristic under study.
Consider a finite population \( U = \{1, 2, 3, \ldots, N\} \). Let \( y \) and \( x \) be two real variables assuming the value \( y_i \) and \( x_i \) on the \( i \)th unit \( (i = 1, 2, 3, \ldots, N) \). Now consider \( y \) be the study variable and \( x \) be the auxiliary variable. We consider simple random sampling without replacement (SRSWOR) design to draw samples in each phase of two-phase sampling set-up. The first phase sample \( s' (s' \subset u) \) of fixed size \( n' \) is drawn to observe \( x \) only. The second phase sample \( s (s \subset s') \) of fixed size \( n' \) is drawn to observe \( y \) and \( x \) for given \( s \), \( (n < n') \).

Let, \( \bar{x} = \frac{1}{n} \sum_{i \in s} x_i \), \( \bar{y} = \frac{1}{n} \sum_{i \in s} y_i \) and \( \bar{x}' = \frac{1}{n'} \sum_{i \in s'} x_i \).

Now the usual two phase ratio, product and regression estimators are given by

\[
t_{dr} = \frac{\bar{y}}{\bar{x}} \bar{x}', \quad (1.1)
\]

\[
t_{dp} = \frac{\bar{y}x}{\bar{x}'}, \quad (1.2)
\]

\[
t_{dlr} = \bar{y} + b_{yx} (\bar{x}' - \bar{x}), \quad (1.3)
\]

where \( b_{yx} \) is the sample regression coefficient of 'y' on 'x', calculated from the data based on second phase sample of size 'n'.

The mean square error (MSE) of the estimators given in (1.1), (1.2) and (1.3) to first order of approximation are respectively given by

\[
\text{MSE}(t_{dr}) = \text{Var} \left[ \lambda C_y^2 + (\lambda - \lambda')(C_x^2 - 2C_{yx}) \right] \quad (1.4)
\]

\[
\text{MSE}(t_{dp}) = \text{Var} \left[ \lambda C_y^2 + (\lambda - \lambda')(C_x^2 - 2C_{yx}) \right] \quad (1.5)
\]

\[
\text{MSE}(t_{dlr}) = \text{Var} \left[ \lambda C_y^2 + (\lambda - \lambda')(C_x^2 - 2C_{yx}) + \lambda'(1 - \rho^2) + \rho \right] \quad (1.6)
\]

where \( \lambda = \left( \frac{1}{n} - \frac{1}{N} \right), \lambda' = \left( \frac{1}{n'} - \frac{1}{N} \right) \)

\[
C_y^2 = \frac{S_y^2}{Y^2}, \quad C_x^2 = \frac{S_x^2}{X^2}, \quad C_{yx} = \frac{S_{yx}}{YX} = \rho C_y C_x,
\]

\[
S_y^2 = \frac{1}{N-1} \sum_{i=1}^{N} (Y_i - \bar{Y})^2, \quad S_x^2 = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})^2,
\]

\[
S_{yx} = \frac{1}{N-1} \sum_{i=1}^{N} (Y_i - \bar{Y})(X_i - \bar{X}) = \rho S_y S_x,
\]
ρ being the population correlation coefficient between 'y' and 'x'.

Singh and Vishwakarma (2007) have proposed modified exponential ratio and product estimate to estimate finite population mean $\bar{Y}$ of study variable y in presence of auxiliary variable x. As Singh and Vishwakarma (2007) assumed that the population mean $\bar{X}$ of auxiliary variable x is not known, they used the two phase mechanism to estimate the population mean $\bar{Y}$.

The modified exponential ratio and product estimator suggested by Singh and Vishwakarma (2007) to estimate population mean $\bar{Y}$ under two phase sampling are given by

\[
t_{der} = \bar{y}.\exp\left(\frac{\bar{x}' - \bar{x}}{\bar{x}' + \bar{x}}\right)
\]

(1.7)

\[
t_{dep} = \bar{y}.\exp\left(\frac{\bar{x} - \bar{x}'}{\bar{x} + \bar{x}'}\right)
\]

(1.8)

The MSE of the estimators $t_{der}$ and $t_{dep}$ to first order of approximation respectively are given by

\[
\text{MSE}(t_{der}) = \bar{Y}^2\left[\lambda C_y^2 + (\lambda - \lambda')\left(\frac{C_x^2}{4} - C_{yx}\right)\right]
\]

(1.9)

\[
\text{MSE}(t_{dep}) = \bar{Y}^2\left[\lambda C_y^2 + (\lambda - \lambda')\left(\frac{C_x^2}{4} + C_{yx}\right)\right]
\]

(1.10)

Khatua and Mishra (2013) have suggested a generalized exponential estimator to estimate finite population mean $\bar{Y}$ under two-phase sampling scheme with assumption that the population mean auxiliary variable x, $\bar{X}$ is not known.

\[
t_{dge} = \bar{y}[d_1 + d_2(\bar{x}' - \bar{x})]\exp\left[\frac{\bar{x}' - \bar{x}}{\bar{x}' + \bar{x}}\right],
\]

(1.11)

where $d_1$ and $d_2$ are suitable chosen constants or suitable chosen constants or statistics and $d_1 + d_2$ are not necessarily equal to unity.

It is to be mentioned that the mean square error (MSE) expression of the estimator $t_{dge}$ obtained by Khatua and Mishra(2013) is not correct. The correct expression of the bias and MSE of the Khatua and Mishra (2013) estimator $t_{dge}$ are given in the following theorem.
**Theorem 1.1** To the first degree of approximation, the bias and MSE of the Khatua and Mishra (2013) estimator $t_{dge}$ are respectively given by

$$B(t_{dge}) = \bar{Y} \left[ d_1 \left( 1 + \frac{\lambda - \lambda'}{8} C_x^2 (3 - 4k) \right) + \frac{d_2 \bar{X}(\lambda - \lambda')}{2} C_x^2 (1 - 2k) - 1 \right]$$

(1.12)

and

$$\text{MSE}(t_{dge}) = \bar{Y}^2 \left[ d_1^2 \left( 1 + \lambda C_y^2 + (\lambda - \lambda') C_x^2 (1 - 2k) \right) + d_2^2 (\lambda - \lambda') \bar{X}^2 C_x^2 + 1 
+ 2d_1d_2 (\lambda - \lambda') \bar{X}(1 - 2k) C_x^2 - 2d_1 \left( 1 + \frac{\lambda - \lambda'}{8} C_x^2 (3 - 4k) \right) \right]$$

(1.13)

where $k = \rho C_y / C_x$.

**Proof.** To obtain the bias and MSE of $t_{dge}$ we write

$$\bar{y} = \bar{Y}(1 + e_0), \quad \bar{x} = \bar{X}(1 - e_1), \quad \bar{x}' = \bar{X}(1 + e_1)$$

Such that

$$E(e_0) = E(e_1) = E(e'_1) = 0$$

and

$$E(e_0^2) = \lambda C_y^2, \quad E(e_1^2) = \lambda C_x^2, \quad E(e_1^2) = \lambda' C_x^2, \quad E(e_0 e_1) = \lambda k C_x^2, \quad E(e_0 e_1') = \lambda' k C_x^2, \quad E(e_1 e_1') = \lambda' C_x^2$$

Now expressing (1.11) in terms of $e$’s we have

$$t_{dge} = \bar{Y}(1 + e_0) \left[ d_1 + d_2 \bar{X}(1 + e'_1 - 1 - e_1) \right] \left[ 1 + \frac{e'_1 - 1 - e_1}{1 + e'_1 + 1 + e_1} \right]$$

$$= \bar{Y}(1 + e_0) \left[ d_1 + d_2 \bar{X}(e_1 - e'_1) \right] \left[ 1 - \frac{(e_1 - e'_1)}{2} \left( 1 + \frac{e_1 + e'_1}{2} \right)^{-1} \right]$$

$$= \bar{Y}(1 + e_0) \left[ d_1 + d_2 \bar{X}(e_1 - e'_1) \right] \left[ 1 - \frac{(e_1 - e'_1)}{2} \left( 1 + \frac{e_1 + e'_1}{2} \right)^{-1} + \frac{(e_1 - e'_1)^2}{8} \left( 1 + \frac{e_1 + e'_1}{2} \right)^{-2} \ldots \right]$$
\[ Y = \sum \left[ d_1(1 + e_0) \left\{ 1 - \frac{(e_i - e_i')}{2} \left( 1 + \frac{e_i + e_i'}{2} \right)^{-1} + \frac{(e_i - e_i')^2}{8} \left( 1 + \frac{e_i + e_i'}{2} \right)^{-2} \right\} \right. \\
- d_2 (1 + e_0) (e_i - e_i') \left\{ 1 - \frac{(e_i - e_i')}{2} \left( 1 + \frac{e_i + e_i'}{2} \right)^{-1} + \frac{(e_i - e_i')^2}{8} \left( 1 + \frac{e_i + e_i'}{2} \right)^{-2} \right\} \right] \]

\[ = \sum \left[ \left\{ d_1 \left( 1 + e_0 \right) \left( 1 - \frac{(e_i - e_i')}{2} \left( 1 + \frac{e_i + e_i'}{2} \right)^{-1} + \frac{(e_i - e_i')^2}{8} \left( 1 + \frac{e_i + e_i'}{2} \right)^{-2} \right\} \right. \\
- d_2 (1 + e_0) (e_i - e_i') \left\{ 1 - \frac{(e_i - e_i')}{2} \left( 1 + \frac{e_i + e_i'}{2} \right)^{-1} + \frac{(e_i - e_i')^2}{8} \left( 1 + \frac{e_i + e_i'}{2} \right)^{-2} \right\} \right] \]

Neglecting terms of e's having power greater than two we have

\[ t_{dp} \approx \sum \left[ d_1 \left( 1 + e_0 \right) \left( 1 - \frac{(e_i - e_i')}{2} \right) + \frac{(e_0 e_i - e_0 e_i')}{2} + \frac{3(e_i^2 - e_i'^2 - 2e_i e_i')}{8} \right] \\
- d_2 (1 + e_0) \left( e_i - e_i' \right) \left( e_0 e_i - e_0 e_i' \right) - \frac{(e_i^2 - 2e_i e_i' + e_i'^2)}{2} \right] \]

Subtracting \( \bar{Y} \) from both side of the above expression we have
\[
(t_{\text{dge}} - \bar{Y}) \cong \bar{Y} \left[ d_1 \left\{ 1 + e_0 - \frac{(e_1 - e'_1)}{2} + \frac{(e_0 e_1 - e_0 e'_1)}{2} + \frac{3e_1^2 - e'_1^2 - 2e_1 e'_1}{8} \right\} \right]
- d_2 \bar{X} \left\{ (e_1 - e'_1) + (e_0 e_1 - e_0 e'_1) - \frac{(e_1^2 - 2e_1 e'_1 + e'_1^2)}{2} \right\} - 1
\]

(1.14)

Taking expectation of both sides of (1.14) we get the bias of \( t_{\text{dge}} \) to the first degree of approximation as

\[
\mathbb{B}(t_{\text{dge}}) = \mathbb{E}(t_{\text{dge}} - \bar{Y}) = \bar{Y} \left[ d_1 \left\{ 1 + \left( \frac{\lambda - \lambda'}{8} \right) C_x^2 (3 - 4k) \right\} \right]
+ d_2 \bar{X}(\lambda - \lambda') C_x^2 (1 - 2k) - 1
\]

(1.15)

Squaring both sides of (1.14) and neglecting terms of e’s having power greater than two we have

\[
(t_{\text{dge}} - \bar{Y})^2 = \bar{Y}^2 \left[ d_1^2 \left\{ 1 + e_0^2 + \frac{(e_1 - e'_1)^2}{4} + 2e_0 - (e_1 - e'_1) - (e_0 e_1 - e_0 e'_1) \right\} \right]
- (e_0 e_1 - e_0 e'_1) + \frac{3e_1^2 - e'_1^2 - 2e_1 e'_1}{4} + d_2^2 \bar{X}^2 (e_1 - e'_1)^2 + 1

- 2d_1 d_2 \bar{X} \left\{ (e_1 - e'_1) + (e_0 e_1 - e_0 e'_1) - \frac{(e_1 - e'_1)^2}{2} \right\}
+ (e_0 e_1 - e_0 e'_1) - \frac{(e_1^2 - 2e_1 e'_1 + e'_1^2)}{2}

- 2d_1 \left\{ 1 + e_0 - \frac{(e_1 - e'_1)}{2} - \frac{(e_0 e_1 - e_0 e'_1)}{2} + \frac{3e_1^2 - e'_1^2 - 2e_1 e'_1}{8} \right\}
+ 2d_2 \bar{X} \left\{ (e_1 - e'_1) + (e_0 e_1 - e_0 e'_1) - \frac{(e_1^2 - 2e_1 e'_1 + e'_1^2)}{2} \right\}
\]

or
$$\left( t_{dge} - \bar{Y} \right)^2 = \bar{Y}^2 \left[ 1 + d_1^2 \left( 1 + 2e_0 - (e_1 - e'_1) - 2(e_0e_1 - e_0e'_1) + \left( e_1^2 - e_1e'_1 \right) \right) + d_2^2 \bar{X}^2 \left( e_1 - e'_1 \right)^2 \right. \\
- 2d_1d_2 \bar{X} \left\{ (e_1 - e'_1) + 2(e_0e_1 - e_0e'_1) - \left( e_1^2 - 2e_1e'_1 + e'_1^2 \right) \right\} \\
- 2d_1 \left\{ 1 + e_0 - \left( e_1 - e'_1 \right) - \left( e_0e_1 - e_0e'_1 \right) + \left( 3e_1^2 - e_1^2 - 2e_1e'_1 \right) \right\} \right. \\
+ 2d_2 \bar{X} \left\{ (e_1 - e'_1) + (e_0e_1 - e_0e'_1) - \left( e_1^2 - 2e_1e'_1 + e'_1^2 \right) \right\} \right]\right]$$

(1.16)

Taking expectation of both sides of (1.16) we get the mean squared error of the estimator $t_{dge}$ to the first degree of approximation as

$$\text{MSE}(t_{dge}) = \bar{Y}^2 \left[ 1 + d_1^2 \left( 1 + \lambda T^2 + \left( \lambda - \lambda' \right) C_x \left( 1 + 2k \right) \right) + d_2^2 \bar{X}^2 \left( \lambda - \lambda' \right) C_x^2 \right. \\
+ 2d_1d_2 \bar{X} \left( \lambda - \lambda' \right) C_x^2 \left( 1 - 2k \right) \\
- 2d_1 \left\{ 1 + \left( \lambda - \lambda' \right) C_x \left( 3 - 4k \right) \right\} \\
+ d_2 \left( \lambda - \lambda' \right) C_x^2 \left( 1 - 2k \right) \right]$$

$$= \bar{Y}^2 \left[ 1 + d_1^2A_{1(1)} + d_2^2A_{2(1)} + 2d_1d_2A_{3(1)} - 2d_1A_{4(1)} - d_2A_{3(1)} \right]$$

(1.17)

which is minimum when

$$\begin{bmatrix} A_{1(1)} & A_{2(1)} \\ A_{3(1)} & A_{2(1)} \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} A_{4(1)} \\ A_{3(1)}/2 \end{bmatrix}$$

(1.18)

Solving (1.18) we get the optimum values of $d_1$ and $d_2$ as

$$d_1 = \frac{2A_{2(1)}A_{4(1)} - A_{3(1)}^2}{2(A_{1(1)}A_{2(1)} - A_{3(1)}^2)} = d_{10}$$

(1.19)

$$d_2 = \frac{A_{3(1)}(A_{1(1)} - 2A_{4(1)})}{2(A_{1(1)}A_{2(1)} - A_{3(1)}^2)} = d_{20}$$

(1.20)

where,
\[ A_{1(i)} = \left[ 1 + \lambda C_{x}^{2} + (\lambda - \lambda') (1 - 2k) C_{x}^{2} \right], \]

\[ A_{2(i)} = (\lambda - \lambda') \overline{X}^{2} C_{x}^{2}, \]

\[ A_{3(i)} = (\lambda - \lambda') \overline{X} (1 - 2k) C_{x}^{2}, \]

\[ A_{4(i)} = \left[ 1 + \frac{(\lambda - \lambda') C_{x}^{2}}{8} (3 - 4k) \right] \]

Substituting \((d_{10}, d_{20})\) in place of \((d_{1}, d_{2})\) in (1.17) we get the minimum MSE of \( t_{dge} \) as

\[
\min \text{ MSE} (t_{dge}) = \overline{Y}^{2} \left[ 1 - \left( \frac{4A_{2(i)}^{2}A_{4(i)}^{2} + A_{4(i)}^{2}A_{3(i)}^{2} - 4A_{3(i)}^{2}A_{4(i)}^{2}}{4(A_{4(i)}^{2} - A_{3(i)}^{2})} \right) \right] \] (1.21)

which is correct expression of MSE of \( t_{dge} \).

If we set \( d_{1} = 1 \) in (1.11), then the class of estimators \( t_{dge} \) reduces to:

\[ t_{dge(i)} = \overline{Y}[1 + d_{2}(\overline{X} - \overline{x})]\exp\left[\frac{\overline{x}’ - \overline{X}}{\overline{x}’ + \overline{X}}\right] \] (1.22)

Putting \( d_{1} = 1 \) in (1.15) and (1.17) the bias and MSE of \( t_{dge(i)} \) are respectively given by

\[ B(t_{dge(i)}) = \overline{Y}(\lambda - \lambda’) C_{x}^{2} \left[ (3 - 4k) + 8d_{2} \overline{X}(1 - 2k) \right] \] (1.23)

and

\[ \text{MSE}\left(t_{dge(i)}\right) = \overline{Y}^{2} \left[ 1 + A_{i(i)} - 2A_{4(i)} + d_{2}A_{2(i)} + d_{2}A_{3(i)} \right] \] (1.24)

The MSE \( t_{dge(i)} \) is minimized for

\[ d_{2} = - \frac{A_{3(i)}}{2A_{2(i)}} = d_{20pt} \] (1.25)

Thus the resulting (minimum) MSE of \( t_{dge(i)} \) is given by
\[
\text{MSE}(t_{\text{dge(1)}}) = \overline{Y}^2\left[ 1 + A_{l(1)} - 2A_{4(l)} + \frac{A_{3(l)}^2}{4A_{2(l)}} \right] \\
= S_y^2\left[ \lambda - (\lambda - \lambda')p^2 \right]
\] (1.26)

which is equal to the approximate variance of the usual regression estimator

\[
\bar{y}_{\text{ldr}} = \bar{y} + \hat{\beta}(\bar{x}' - \bar{x})
\] (1.27)

where \( \hat{\beta} = \frac{s_{xy}}{s_x^2} \) is the sample regression coefficient of \( y \) on \( x \).

we note that the MSE expression of the estimator of \( t_{\text{dge(1)}} \):

\[
\text{MSE}(t_{\text{dge(1)}}) = \overline{Y}^2\left[ 1 + d_1^2\left\{ 1 + \lambda C_y^2 + \frac{(\lambda - \lambda')C_y^2}{4}(1 - 4k) \right\} + d_2^2\overline{X}^2(\lambda - \lambda')C_x^2 \\
+ 2d_1d_2\overline{X}C_x^2(\lambda - \lambda')(1 - 2k) + 2d_1d_2\overline{X}C_x^2(\lambda - \lambda')(1 - 2k) - 2d_1 \right]
\] (1.28)

obtained by Khatua and Mishra (2013) is not corrected. So the other results of the paper are also incorrect. This led authors to derive the correct expression of MSE of the estimator \( t_{\text{dge(1)}} \), which is given in 1.28.

2. The Suggested Class of Estimators

For estimating the population mean \( \overline{Y} \), we define a class of estimators as

\[
t_s = \overline{Y}\left[ \delta_1 + \delta_2(\bar{x}' - \bar{x}) \right]\left( \frac{\bar{x}'}{\overline{X}} \right)^{\alpha} \exp\left[ \frac{\delta(\bar{x}' - \bar{x})}{(\bar{x}' + \overline{X})} \right],
\] (2.1)

where \( \left( \delta_1, \delta_2 \right) \) are suitably chosen constants such that the MSE of \( t \) is minimum, \( (\alpha, \delta) \) are suitably chosen classes scalars which may assume real number or the functions of \( \rho, C_y, C_x \) etc.

It is to be noted that for \( (\alpha, \delta) = (0,1) \) the proposed class of estimators \( t_s \) reduces to the class of estimators reported by Khatua and Mishra (2013).

To obtain the bias and MSE of \( t_s \) we express \( t_s \) in terms of e’s we have

\[
t_s = \overline{Y}(1 + e_0\left[ \delta_1 + \delta_2(\bar{e}_1' - \bar{e}_1) \right]\left( \frac{1 + e_1'}{1 + e_1} \right)^{\alpha} \exp\left[ \frac{\delta(\bar{e}_1' - \bar{e}_1)}{2 + \bar{e}_1 + e_1'} \right],
\] (2.2)
We assume that $|e_1| < 1, |e_1'| < 1$ so that the right hand side of (2.2) is expandable. Expanding the right hand side, multiplying out and neglecting terms of $e'$s having power greater than two we have some members of the proposed class of estimators $t_s$ are given in Table 2.1

**Table 2.1 Some members of the proposed class of estimators $t_s$**

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Estimators</th>
<th>Values of scalars</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$t_1 = \bar{y}$</td>
<td>$\delta_1 \quad 1 \quad 0 \quad 0 \quad 0$</td>
</tr>
<tr>
<td>2</td>
<td>$t_2 = \bar{y}(\bar{x}'/\bar{x})$</td>
<td>$\delta_1 \quad 1 \quad 0 \quad 1 \quad 0$</td>
</tr>
<tr>
<td>3</td>
<td>$t_3 = \bar{y}(\bar{x}/\bar{x}')$</td>
<td>$\delta_1 \quad 1 \quad 0 \quad -1 \quad 0$</td>
</tr>
<tr>
<td>4</td>
<td>$t_4 = \bar{y}(\bar{x}'/\bar{x})^\alpha$</td>
<td>$1 \quad 0 \quad \alpha \quad 0$</td>
</tr>
<tr>
<td>5</td>
<td>$t_5 = \bar{y}\exp\left(\frac{\bar{x}' - \bar{x}}{\bar{x}' + \bar{x}}\right)$</td>
<td>$\delta_1 \quad 1 \quad 0 \quad 0 \quad 1$</td>
</tr>
<tr>
<td>6</td>
<td>$t_6 = \bar{y}\exp\left(\frac{\bar{x} - \bar{x}'}{\bar{x} + \bar{x}'}\right)$</td>
<td>$\delta_1 \quad 1 \quad 0 \quad 0 \quad -1$</td>
</tr>
<tr>
<td>7</td>
<td>$t_7 = \bar{y}\exp\left(\frac{\delta (\bar{x}' - \bar{x})}{\bar{x}' + \bar{x}}\right)$</td>
<td>$\delta_1 \quad 1 \quad 0 \quad 0 \quad \delta$</td>
</tr>
<tr>
<td>8</td>
<td>$t_8 = \bar{y}[d_1 + d_2(\bar{x}' - \bar{x})]\exp\left(\frac{\bar{x}' - \bar{x}}{\bar{x} + \bar{x}'}\right)$</td>
<td>$d_1 \quad d_2 \quad 0 \quad 1$</td>
</tr>
<tr>
<td>9</td>
<td>$t_9 = \bar{y}[\delta_1 + \delta_2(\bar{x}' - \bar{x})]\left(\frac{\bar{x}}{\bar{x}'}\right)$</td>
<td>$\delta_1 \quad \delta_2 \quad -1 \quad 0$</td>
</tr>
<tr>
<td>10</td>
<td>$t_{10} = \bar{y}[\delta_1 + \delta_2(\bar{x}' - \bar{x})]\left(\frac{\bar{x}}{\bar{x}'}\right)\exp\left(\frac{(\bar{x} - \bar{x}')}{8(\bar{x} + \bar{x}')}\right)$</td>
<td>$\delta_1 \quad \delta_2 \quad -1 \quad -1/8$</td>
</tr>
<tr>
<td>11</td>
<td>$t_{11} = \bar{y}[\delta_1 + \delta_2(\bar{x}' - \bar{x})]\left(\frac{\bar{x}}{\bar{x}'}\right)\exp\left(\frac{(\bar{x} - \bar{x}')}{4(\bar{x} + \bar{x}')}\right)$</td>
<td>$\delta_1 \quad \delta_2 \quad -1 \quad -1/4$</td>
</tr>
</tbody>
</table>
\[
    t_s \cong Y \left[ \delta_1 \left\{ 1 + e_0 - \frac{\theta(e_1 - e_1')}{2} - \frac{\theta(e_0e_1 - e_0e_1')}{2} + \frac{\theta(e_1^2 - e_1'^2)}{4} + \frac{\theta^2(e_1 - e_1')^2}{8} \right\} 
    - \delta_2 \overline{X} \left\{ (e_1 - e_1') + (e_0e_1 - e_0e_1') - \frac{\theta(e_1 - e_1')^2}{2} \right\} \right]
\]

or

\[
    (t_s - \overline{Y}) \cong Y \left[ \delta_1 \left\{ 1 + e_0 - \frac{\theta(e_1 - e_1')}{2} - \frac{\theta(e_0e_1 - e_0e_1')}{2} + \frac{\theta(e_1^2 - e_1'^2)}{4} + \frac{\theta^2(e_1 - e_1')^2}{8} \right\} 
    - \delta_2 \overline{X} \left\{ (e_1 - e_1') + (e_0e_1 - e_0e_1') - \frac{\theta(e_1 - e_1')^2}{2} \right\} - 1 \right]
\]

where \( \theta = (2\alpha + \delta) \).

Taking expectation of both sides of (2.3) we get the bias of the proposed class of estimator \( t_s \) to the first degree of approximation as

\[
    B(t_s) = Y \left[ \delta_1 \left\{ 1 + \frac{(\lambda - \lambda')\theta C_X^2(\theta - 4k + 2)}{8} \right\} + \delta_2 \overline{X}(\lambda - \lambda')C_X^2(\theta - 2k) - 1 \right]
\]

Squaring both sides of (2.3) and neglecting terms of \( e \)'s having power greater than two we have

\[
    (t_s - \overline{Y})^2 = Y^2 \left[ 1 + d_1 \left\{ 1 + 2e_0 - \theta(e_1 - e_1') - 2\theta(e_0e_1 - e_0e_1') + e_0^2 
    + \frac{\theta}{2}(e_1^2 + e_1'^2 - 2e_1e_1') + e_1^2 - e_1'^2 \right\} + d_2^2\overline{X}^2(e_1 - e_1')^2 
    - 2d_1d_2\overline{X}(e_1 - e_1') + 2(e_0e_1 - e_0e_1') - \theta(e_1 - e_1')^2 \right] 
    - 2d_1 \left\{ 1 + e_0 - \frac{\theta(e_1 - e_1')}{2} - \frac{\theta(e_0e_1 - e_0e_1')}{2} + \frac{\theta(e_1 - e_1')^2}{8} \right\} 
    + \frac{\theta}{2}(e_1^2 + e_1'^2 - 2e_1e_1') + e_1^2 - e_1'^2 \right\} \right] 
    + 2d_2\overline{X} \left\{ (e_1 - e_1') + (e_0e_1 - e_0e_1') - \frac{\theta(e_1 - e_1')^2}{2} \right\} \right]
\]

Taking expectation of both sides of (2.5) we get the mean squared error of \( t_s \) to the first degree of approximation as

\[
    \text{MSE}(t_s) = Y^2 \left[ 1 + \delta_1^2A_1 + \delta_2^2A_2 + 2\delta_1\delta_2A_3 - 2\delta_1A_4 - \delta_2A_3 \right]
\]
where
\[
A_1 = \left[1 + \lambda C^2_x + \frac{(\lambda - \lambda')\theta C^2_x}{2} (\theta - 4k + 1)\right]
\]
\[
A_2 = (\lambda - \lambda')\bar{X}^2 C^2_x,
\]
\[
A_3 = \bar{X}(\lambda - \lambda')(\theta - 2k)C^2_x,
\]
\[
A_4 = \left[1 + \frac{\theta(\lambda - \lambda')C^2_x}{8} (\theta - 4k + 2)\right]
\]

The MSE of \( t_s \) is minimum when
\[
\begin{bmatrix}
A_1 & A_3 \\
A_3 & A_2
\end{bmatrix}
\begin{bmatrix}
\delta_1 \\
\delta_2
\end{bmatrix} = \begin{bmatrix}
A_4 \\
(A_3/2)
\end{bmatrix}
\] (2.7)

Solving (2.7) we get the optimum values of \( \delta_1 \) and \( \delta_2 \) as
\[
\delta_1 = \frac{2A_2 A_4 - A_3^2}{2(A_1 A_3 - A_3^2)} = \delta_{10}
\] (2.8)
\[
\delta_2 = \frac{A_3(A_1 - 2A_4)}{2(A_1 A_2 - A_3^2)} = \delta_{20}
\] (2.9)

Putting (2.8) and (2.9) in (2.6) we get the minimum MSE of \( t_s \) as
\[
\min \text{MSE}(t_s) = \bar{Y}^2 \left[ 1 - \frac{4A_2 A_4^2 + A_1 A_3^2 - 4A_2 A_4}{4(A_1 A_2 - A_3^2)} \right]
\] (2.10)

For \( \delta_1 = 1 \), the proposed class of estimators \( t_s \) reduces to:
\[
t_{s(1)} = \bar{Y} \left[ 1 + \delta_2 (\bar{X}' - \bar{X}) \left( \frac{\bar{X}'}{\bar{X}} \right)^\alpha e^{\delta (\bar{X}' - \bar{X}) / (\bar{X}' - \bar{X})} \right]
\] (2.11)

Putting \( \delta_1 = 1 \) in (2.4) and (2.6) we get the bias and MSE of \( t_{s(1)} \) to the first degree of approximation are respectively given by
\[ B(t_{s(1)}) = \sum \frac{(\lambda - \lambda')C_x^2}{8} \left[ \theta(\theta - 4k + 2) + 4\delta_2 X(\theta - 2k) \right] \]  

(2.12)

\[ \text{MSE}(t_{s(1)}) = \sum^2 \left[ 1 + A_1 - 2A_4 + \delta_2 A_2 + \delta_2 A_3 \right] \]  

(2.13)

The MSE \( t_{s(1)} \) is minimized for

\[ \delta_2 = -\frac{A_3}{2A_2} = d_{2(0pt)} \]  

(2.14)

Thus resulting minimum MSE of \( t_{s(1)} \) is given by

\[
\min \text{MSE}(t_{s(1)}) = \sum^2 \left[ 1 + A_1 - 2A_4 - \frac{A^2_1}{4A_2} \right]
\]

\[ = S_y^2 \left[ \lambda - (\lambda - \lambda')p^2 \right] \]  

(2.15)

which is equal to the approximate variance of the usual regression estimator \( \bar{y}_{ld} \) in two phase sampling.

3. Efficiency comparisons

For \( (\delta_1, \delta_2, \alpha, \delta) = (1,0,0,0) \) in (2.2.1), the class of estimators \( t_s \) reduces to the usual unbiased estimator

\[ t_s^{(1)} = \bar{y} \]  

(3.1)

The variance of \( \bar{y} \) is given by

\[ \text{var}(\bar{y}) = \text{MSE}(\bar{y}) = \lambda \bar{y}^2 C_y^2 = \lambda S_y^2 \]  

(3.2)

From (2.1.14), (2.1.15), (2.1.6) and (2.3.2) we have

\[ \text{MSE}(\bar{y}) - \text{MSE}(t_{dlr}) = (\lambda - \lambda')\bar{y}^2 C_x^2 k^2 \geq 0 \]  

(3.3)

\[ \text{MSE}(t_{dlr}) - \text{MSE}(t_{dlr}) = (\lambda - \lambda')\bar{y}^2 C_x^2 (1 - k)^2 \geq 0 \]  

(3.4)

\[ \text{MSE}(t_{dp}) - \text{MSE}(t_{dlr}) = (\lambda - \lambda')\bar{y}^2 C_x^2 (1 + k) \geq 0 \]  

(3.5)
\[ \text{MSE}(t_{\text{der}}) - \text{MSE}(t_{\text{dlr}}) = \frac{(\lambda - \lambda') Y^2}{4} C_x (1 - 2k)^2 \geq 0 \] (3.6)

\[ \text{MSE}(t_{\text{dep}}) - \text{MSE}(t_{\text{dlr}}) = \frac{(\lambda - \lambda') Y^2}{4} C_x (1 + 2k)^2 \geq 0 \] (3.7)

It is clear from (3.3), to (3.7) that the usual two-phase regression estimator \( \bar{y}_{\text{dlr}} \) is better than sample mean \( \bar{y} \), ratio estimator \( t_{\text{dr}} \), the product estimator \( t_{\text{dp}} \) in two phase sampling, two phase exponential ratio estimator \( t_{\text{der}} \) and the two phase exponential procedure estimator \( t_{\text{dep}} \).

Further from (2.10) and (2.15) we have

\[ \text{MSE}(t_{\text{dlr}}) - \min \text{MSE}(t_s) = Y^2 \left[ A_1 - 2A_4 - \frac{A_2^2}{4A_2} + \frac{4A_2^2A_4^2 + A_1A_3^2 - 4A_2^2A_4}{4(A_1A_2 - A_3^2)} \right] \]

\[ = \frac{Y^2 (2A_1A_2 - A_3^2 - 2A_2A_4)}{4A_2(A_1A_2 - A_3^2)} \geq 0 \] (3.8)

which shows that the proposed class of estimator’s \( t_s \) is more efficient than the regression estimator \( t_{\text{dlr}} \). Thus the proposed class of estimator’s \( t_s \) is better than usual unbiased estimator \( \bar{y} \), conventional two phase ratio estimator \( t_{\text{dr}} \) two – phase product estimator \( t_{\text{dp}} \) and the two phase regression estimator \( t_{\text{dlr}} \).

4. Empirical Study

To have tangible idea about the performance of the members of the proposed class of estimators \( t_s \) over usual unbiased estimator \( \bar{y} \) we have computed the percent relative efficiency (PRE) of the member of the suggested class of estimator’s \( t_s \) with respect to \( \bar{y} \) by using the formula:

\[ \text{PRE}(t_s, \bar{y}) = \frac{\lambda C_y^2}{\left[ 1 + \delta_1 A_1 + \delta_2 A_2 + 2\delta_1 \delta_2 A_3 - 2\delta_1 A_4 - \delta_2 A_3 \right]} \times 100 \] (4.1)

For better eye view, we have considered two natural population data set with positive correlation between \( y \) and \( x \) and two population data sets with negative correlation between \( y \) and \( x \) earlier considered by Khatua and Mishra (2013). The descriptions are given below:

**Population I:** Murthy (1967), P. 228

\( y \): Output, \( x \): Fixed Capital, \( N = 80, n' = 70, n = 30 \)
\[ \lambda = 0.02083, \lambda' = 0.00179, \quad \bar{Y} = 5182.64, \quad C_y^2 = 0.1255, \quad C_x^2 = 0.5635, \quad C_{yx} = 0.2503, \]
\[ \rho = 0.9413, \quad k = 0.44419 \]

**Population II:** Das (1988)

y: No. of agricultural labors for 1971, \( x \): No of agricultural labors for 1961, \( N = 278, \quad n' = 70, \quad n = 30 \)
\[ \lambda = 0.0297, \quad \lambda' = 0.01069, \quad \bar{Y} = 39.0, \quad 6. \quad C_y^2 = 2.0883, \quad C_x^2 = 2.6237, \quad C_{yx} = 1.6883, \]
\[ \rho = 0.7213, \quad k = 0.64348 \]

**Population III:** Steel and Torrie (1960), P. 282

y: Log of leaf burn in sacs, \( x \): Chlorine percentage, \( N = 30, \quad n' = 12, \quad n = 4 \)
\[ \lambda = 0.21667, \quad \lambda' = 0.05, \quad \bar{Y} = 0.6, \quad 8. \quad C_y^2 = 0.2306, \quad C_x^2 = 0.5614, \quad C_{yx} = -0.1798, \]
\[ \rho = -0.4996, \quad k = -0.32027 \]

**Population IV:** Gujurati (1999), P. 259

y: Year to year percentage change in the index of hourly earnings, \( x \): The unemployment rate (%), \( N = 12, \quad n' = 8, \quad n = 5, \)
\[ \lambda = 0.11667, \quad \lambda' = 0.041667, \quad \bar{Y} = 4.066, \quad C_y^2 = 0.0977, \quad C_x^2 = 0.0535, \quad C_{yx} = -0.0519, \]
\[ \rho = -0.718, \quad k = -0.970 \]

We have computed the percent relative efficiencies of the different members of the proposed class of estimator's \( t_{dge} \) with respect to usual unbiased estimator \( \bar{y} \) and findings are shown in Tables 2.4.1 and 2.4.2.
Tables 2.4.1 and 2.4.2 clearly show that the modified two-phase estimator is more efficient than mean per unit estimator, two-phase ratio estimator, two-phase ratio type exponential estimator, two-phase product estimator, two-phase product type exponential and two-phase regression estimator. Tables 2.4.1 and 2.4.2 also exhibits that the new generated estimators are more efficient than the existing once. Thus the proposed estimators are to be preferred in practice.

References


