Efficient Ratio-type Exponential Estimator for Population Variance

K.B. Panda1, M. Sen2 and P. Das3

1,2,3 Department of Statistics, Utkal University, Bhubaneswar, Odisha-751004, India

Corresponding Author: kunja.st@utkaluniversity.ac.in

Available online at: www.isroset.org

Received: 02/Jun/2018, Revised: 09/May/2018, Accepted: 23/Jul/2018, Online: 30/Jul/2018

Abstract—In this paper, a new exponential ratio type estimator has been proposed for estimating the population variance using auxiliary information. To the first order of approximation, i.e., to o(n−1), the expressions for the bias and the mean square error of the proposed exponential ratio-type estimator have been derived. The optimum value of the characterizing scalar, which minimizes the MSE of proposed estimator, has been obtained. With this optimum value, the expression for minimum MSE of the proposed estimator has been arrived at. The proposed estimator has been compared theoretically with sample variance, traditional ratio estimator due to Isaki [1], and exponential ratio-type estimator due to Singh et.al.[3] and it is found that, under practical conditions, the proposed estimator fares better than its competing estimators. An empirical investigation has been carried out to demonstrate the efficiency of the proposed estimator.

Keywords—Auxiliary variable, single-phase sampling, mean square error, bias

I. INTRODUCTION

In survey sampling, the utilization of auxiliary information is frequently acknowledged to enhance the accuracy of the estimation of population characteristics. Estimation of the finite population variance has great significance in various fields such as in matters of health, variation in body temperature, pulse beat and blood pressure etc. Using auxiliary information, we, in this paper, introduce one new estimator which fares better than competing estimators.

Consider a finite population of size N, arbitrarily labelled 1, 2,...,N. Let \(y_i\) and \(x_i\) be, respectively, the values of the study variable \(y\) and the auxiliary variable \(x\), in respect of the \(i^{th}\) unit (\(i=1, 2,… N\)) of the population. When the auxiliary variable \(x\) is positively correlated with the study variable \(y\) and \(S_x^2\), the population variance of \(x\) is known, ratio method of estimation is usually invoked to estimate the population variance \(S_y^2\) of the study variable.

II. NOTATIONS AND SOME EXISTING ESTIMATORS

In simple random sampling without replacement, we know that the sample variance \(s_y^2\) provides an unbiased estimator of the population variance \(S_y^2\).

\[
S_y^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2
\]

\[
s_y^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2.
\]

Accordingly, we define

\[
S_x^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2
\]

\[
s_x^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2.
\]

study variable \(y\) and the auxiliary variable \(x\), in respect of the \(i^{th}\) unit (\(i=1, 2,… N\)) of the population. When the auxiliary variable \(x\) is positively correlated with the study variable \(y\) and \(S_x^2\), the population variance of \(x\) is known, ratio method of estimation is usually invoked to estimate the population variance \(S_y^2\) of the study variable.

\[
\text{and } s_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2.
\]

as the population and sample variances, respectively, for the auxiliary variable \(x\).

Let \(e_0 = \frac{s_y^2 - s_x^2}{s_x^2}\), i.e., \(s_y^2 = S_x^2 (1 + e_0)\),

\[
e_0 = e_0.
\]

such that \(E(e_0) = E(e_1) = 0\), \(E(e_0^2) = \frac{1}{n}(\lambda_{40} - 1)\), \(E(e_1^2) = \frac{1}{n}(\lambda_{41} - 1)\) and \(E(e_0 e_1) = \frac{1}{n}(\lambda_{22} - 1)\).
where \( \lambda_{pq} = \frac{\mu_{pq}}{\mu_{20}/\mu_{02}^{1/2}} \)

and \( \mu_{pq} = \frac{1}{N} \sum_{i=1}^{N} (Y_i - \bar{Y})(x_i - \bar{x})^{p+q} \); (p, q) being non-negative integers and \( \mu_{02} \), \( \mu_{20} \) are the second order moments and \( C_x = \frac{\sigma_x}{\mu_x} \) is the coefficient of variation for auxiliary variable \( X \). With the above notations, the variance of the estimator \( s_y^2 \) is expressed as

\[
V(s_y^2) = \frac{1}{n} S_y^4(\lambda_{40} - 1).
\]

Isaki (1983) proposed the ratio type estimator for estimating the population variance of the study variable as

\[
s_y^2 = \frac{\sum_{i=1}^{N} Y_i x_i}{\sum_{i=1}^{N} x_i} - \frac{1}{n} S_y^2(\lambda_{04} - \lambda_{22})
\]

whose bias and mean square error, up to the first order of approximation, i.e., to \( o(n^{-1}) \) are respectively,

\[
B(s_y^2) = \frac{1}{n} S_y^2(\lambda_{04} - \lambda_{22})
\]

and

\[
MSE(s_y^2) = \frac{S_y^4}{n}[\lambda_{40} + \lambda_{04} - 2\lambda_{22}]
\]

Singh et. al. (2011) suggested ratio-type exponential estimator for population variance in single phase sampling as

\[
s_y^{2}_{Re} = s_y^2 \exp \left[ \frac{S_y^2 - S_x^2}{S_x^2 + S_x^2} \right]
\]

where, \( \alpha \) is a suitably chosen pre-assigned constant. It may be noted here that if \( \alpha = 1 \), the new estimator reduces to usual ratio-type estimator due to Singh et al. (2011)

Substituting the value of \( e_0, e_1 \) in the expression (8), we get

\[
s_y^{2}_{Re} = s_y^2 \exp \left[ \frac{\alpha(S_y^2 - S_x^2 - S_x^2 e_1)}{S_x^2 + S_x^2 + S_x^2 e_1} \right]
\]

\[
= s_y^2 \exp \left[ \frac{\alpha(-e_1 S_y^2)}{S_x^2(2 + e_1)} \right]
\]

\[
=S_y^2(1 + e_0) \exp [\alpha(-e_1)(2 + e_1)^{-1}]
\]

Retaining only up to \( n^{-1} \) term, we get

\[
s_y^{2}_{Re} = S_y^2(1 + e_0)\left[1 - \alpha e_1 + \alpha^2 \frac{e_1^2}{4} + \alpha^2 \frac{e_1^2}{8} \right]
\]

\[
= S_y^2\left[1 - \alpha e_1 + \alpha^2 \frac{e_1^2}{4} + \alpha^2 \frac{e_1^2}{8} + e_0 - \alpha \frac{e_0 e_1}{2} \right]
\]

The bias of the proposed exponential ratio estimator, to the first degree of approximation, i.e., to \( o(n^{-1}) \), has been derived as follows

\[
Bias(s_y^{2}_{Re}) = \frac{S_y^4}{n}\left[\frac{\alpha}{n}(\lambda_{04} - 1) + \frac{\alpha^2}{8}(\lambda_{04} - 1) - \frac{\alpha}{2}(\lambda_{22} - 1) \right]
\]

The mean square error of the proposed exponential ratio estimator, to the first degree of approximation, i.e., to \( o(n^{-1}) \), has been derived as follows

\[
MSE(s_y^{2}_{Re}) = E\left[s_y^{2}_{Re} - s_y^2 \right]^2
\]

\[
= E\left[s_y^2\left(1 - \alpha e_1 + \alpha^2 \frac{e_1^2}{4} + \alpha^2 \frac{e_1^2}{8} + e_0 - \alpha \frac{e_0 e_1}{2} \right) - s_y^2 \right]^2
\]

\[
= E\left[\frac{S_y^4}{n}\left(\alpha e_1 + \alpha^2 \frac{e_1^2}{4} + \alpha^2 \frac{e_1^2}{8} + e_0 - \alpha \frac{e_0 e_1}{2} \right)^2
\]

\[
= S_y^4\left[\frac{\alpha}{n}(\lambda_{04} - 1) + \frac{\alpha^2}{8}(\lambda_{04} - 1) - \frac{\alpha}{2}(\lambda_{22} - 1) \right]^2
\]

\[
= S_y^4\left[\frac{\alpha}{n}(\lambda_{04} - 1) + \frac{\alpha^2}{8}(\lambda_{04} - 1) - \frac{\alpha}{2}(\lambda_{22} - 1) \right]^2
\]

\[
MSE(s_y^{2}_{Re}) = \frac{S_y^4}{n}[(\lambda_{04} - 1) + \frac{\alpha^2}{4}(\lambda_{04} - 1) - \alpha(\lambda_{22} - 1)]
\]
With a view to determining the most suitable value of $\alpha$, to be called $\alpha_{opt}$, we proceed to minimize the mean square error subject to variation in $\alpha$, implying thereby that

$$\frac{\partial MSE (S_{YRe})}{\partial \alpha} = 0$$

$$\Rightarrow \frac{s^4}{n} \left[ \frac{1}{4} \lambda_{04} - 1 \right] - (\lambda_{22} -1) = 0$$

$$\Rightarrow \alpha = \frac{2(\lambda_{22} -1)}{(\lambda_{04} -1)}.$$  \hspace{1cm} (12)

Thus, $\alpha_{opt} = \frac{2(\lambda_{22} -1)}{(\lambda_{04} -1)}.$

Substituting this value of $\alpha$ in the expression for MSE $(s^2_{YRe})$, i.e in (11), we arrive at the minimum value of MSE $(s^2_{YRe})$, which is expressed as

$$MSE_{opt}(s^2_{YRe}) = \frac{s^4}{n}(\lambda_{40} -1)$$

$$\Rightarrow \frac{(\lambda_{22} -1)^2}{(\lambda_{04} -1)} - \frac{2(\lambda_{22} -1)^2}{(\lambda_{04} -1)} - \frac{1}{n} S^4_Y(\lambda_{40} -1) < 0$$

$$\Rightarrow 2(\lambda_{40} -1) + \frac{2(\lambda_{22} -1)^2}{(\lambda_{04} -1)} - \frac{1}{n} S^4_Y(\lambda_{40} -1) < 0$$

$$\Rightarrow \lambda_{22} - \frac{1}{n} S^4_Y(\lambda_{40} -1) < 0.$$ \hspace{1cm} (13)

On comparison of (13) with (1), the following results can be arrived at

$$MSE_{opt}(s^2_{YRe}) - V(s^2_Y) < 0$$

$$\Rightarrow \frac{s^4}{n}(\lambda_{40} -1) + \frac{(\lambda_{22} -1)^2}{(\lambda_{04} -1)} - \frac{2(\lambda_{22} -1)^2}{(\lambda_{04} -1)} - \frac{1}{n} S^4_Y(\lambda_{40} -1) < 0$$

$$\Rightarrow \lambda_{22} - \frac{1}{n} S^4_Y(\lambda_{40} -1) < 0.$$ \hspace{1cm} (14)

On comparison of (13) with (4), the following results can be arrived at

$$MSE_{opt}(s^2_{YRe}) - MSE(s^2_Y) < 0$$

$$\Rightarrow \frac{s^4}{n}(\lambda_{40} -1) + \frac{(\lambda_{22} -1)^2}{(\lambda_{04} -1)} - \frac{2(\lambda_{22} -1)^2}{(\lambda_{04} -1)} - \frac{1}{n} S^4_Y(\lambda_{40} -1) < 0$$

$$\Rightarrow \lambda_{22} - \frac{1}{n} S^4_Y(\lambda_{40} -1) < 0.$$ \hspace{1cm} (15)

On comparison of (13) with (7), the following results can be arrived at

$$MSE_{opt}(s^2_{YRe}) - MSE(s^2_Y) < 0$$

$$\Rightarrow \frac{s^4}{n}(\lambda_{40} -1) + \frac{(\lambda_{22} -1)^2}{(\lambda_{04} -1)} - \frac{2(\lambda_{22} -1)^2}{(\lambda_{04} -1)} - \frac{1}{n} S^4_Y(\lambda_{40} -1) < 0$$

$$\Rightarrow \lambda_{22} - \frac{1}{n} S^4_Y(\lambda_{40} -1) < 0.$$ \hspace{1cm} (16)

(1) The newly proposed estimator $s^2_{YRe}$ performs better than the simple variance estimator of population variance $S^2_Y$ if

$$\lambda_{22} - \frac{1}{n} S^4_Y(\lambda_{40} -1) < 0.$$ \hspace{1cm} (14)

(2) The newly proposed estimator $s^2_{YRe}$ performs better than the ratio-type estimator due to Isaki (1983) for variance $S^2_Y$ if

$$\lambda_{22} - \frac{1}{n} S^4_Y(\lambda_{40} -1) < 0.$$ \hspace{1cm} (15)

(3) The newly proposed estimator $s^2_{YRe}$ performs better than the ratio-type exponential estimator due to Singh et. al. (2011) for variance $S^2_{Re}$, i.e. $S^2_{YRe}$ if

$$\lambda_{22} - \frac{1}{n} S^4_Y(\lambda_{40} -1) < 0.$$ \hspace{1cm} (16)

IV. EMPIRICAL FINDINGS

With a view to establishing the supremacy of the proposed estimator over the competing estimator, we consider the following example for ratio method of estimation which is taken from Sukhatme and Sukhatme (1970, p. 185), where in the variables of interest are;

Y: Area under wheat in 1937 (in acres)

X: Cultivated area in 1931

Table 4.1: Parameters of population

<table>
<thead>
<tr>
<th>Sl. No</th>
<th>Parameters</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>N</td>
<td>80</td>
</tr>
<tr>
<td>2</td>
<td>n</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>$\lambda_{40}$</td>
<td>3.5469</td>
</tr>
<tr>
<td>4</td>
<td>$\lambda_{04}$</td>
<td>32816</td>
</tr>
<tr>
<td>5</td>
<td>$\lambda_{22}$</td>
<td>2.6601</td>
</tr>
</tbody>
</table>

Making use of the corresponding values of $\lambda_{40}$, $\lambda_{04}$ and $\lambda_{22}$ given in the table 4.1 in the expression (16), we get
The percentage relative efficiency (PRE) of different estimators with respect to usual unbiased estimator $\bar{Y}$ is computed by the formula

$$\text{PRE} (\hat{Y}) = \frac{\sigma_Y^2}{MSE} \times 100.$$  

It is clear from the above table that the newly proposed estimator $s^2_{YRe}$ performs better than the competing estimators.

V. CONCLUSION

We have proposed a ratio-type exponential estimator for estimating the population variance and demonstrated both theoretically and numerically that the proposed estimator fares better than its competing estimators under conditions that hold good in practice. This work can be extended to two-phase sampling also.

REFERENCES


