Almost Unbiased Ratio and Product Type Estimator of Finite Population Variance Using the Knowledge of Kurtosis of an Auxiliary Variable in Sample Surveys
Abstract

It is well recognized that the use of auxiliary information in sample survey design results in efficient estimators of population parameters under some realistic conditions. Out of many ratio, product and regression methods of estimation are good examples in this context. Using the knowledge of kurtosis of an auxiliary variable Upadhyaya and Singh (1999) have suggested an estimator for population variance. In this paper, following the approach of Singh and Singh (1993), we have suggested almost unbiased ratio and product-type estimators for population variance.
1. Introduction

Let \( U = \{U_1, U_2, \ldots, U_N\} \) denote a population of \( N \) units from which a simple random sample without replacement (SRSWOR) of size \( n \) is to be drawn. Further let \( y \) and \( x \) denote the study and the auxiliary variables respectively. The problem is to estimate the parameter

\[
S_y^2 = \frac{N}{N-1} \sigma_y^2
\]

(1.1)

with \( \sigma_y^2 = \frac{1}{N} \sum_{i=1}^{N} (y_i - \bar{Y})^2 \) of the study variate \( y \) when the parameter

\[
S_x^2 = \frac{N}{N-1} \sigma_x^2
\]

(1.2)

with \( \sigma_x^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{X})^2 \) of the auxiliary variate \( x \) is known,

where \( \bar{Y} = \frac{1}{N} \sum_{i=1}^{N} y_i \) and \( \bar{X} = \frac{1}{N} \sum_{i=1}^{N} x_i \); are the population means of \( y \) and \( x \) respectively.

The conventional unbiased estimator of \( S_y^2 \) is defined by

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

(1.3)

where \( \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \) is the sample mean of \( y \).

Using information on \( S_x^2 \), Isaki (1983) proposed a ratio estimator for \( S_y^2 \) as

\[
t_i = s_y^2 \frac{S_x^2}{S_{S}^2}
\]

(1.4)

where \( s_x^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (x_i - \bar{x})^2 \) is unbiased estimator of \( S_x^2 \).
In many survey situations the values of the auxiliary variable \( x \) may be available for each unit in the population. Thus the value of the kurtosis \( \beta_2(x) \) of the auxiliary variable \( x \) is known. Using information on both \( S^2_x \) and \( \beta_2(x) \) Upadhyay and Singh (1999) suggested a ratio type estimator for \( S^2_y \) as

\[
t_2 = S^2_y \left[ \frac{S^2_x + \beta_2(x)}{S^2_x + \beta_2(x)} \right]
\]

(1.5)

For simplicity suppose that the population size \( N \) is large enough relative to the sample size \( n \) and assume that the finite population correction (fpc) term can be ignored. Up to the first order of approximation, the variance of \( S^2_y \), and \( t_1 \) and bias and variances of \( t_2 \) (ignoring fpc term) are respectively given by

\[
\text{var}(S^2_y) = \frac{S^4_y}{n} \{\beta_2(y) - 1\}
\]

(1.6)

\[
\text{var}(t_1) = \frac{S^4_y}{n} \left[ \{\beta_2(y) - 1\} + \{\beta_2(x) - 1\}(1 - 2C) \right]
\]

(1.7)

\[
B(t_2) = \frac{S^2_y}{n} \left[ \{\beta_2(x) - 1\}\theta(\theta - C) \right]
\]

(1.8)

\[
\text{var}(t_2) = \frac{S^4_y}{n} \left[ \{\beta_2(y) - 1\} + \theta\{\beta_2(x) - 1\}(\theta - 2C) \right]
\]

(1.9)

where \( \theta = \frac{S^2_x}{S^2_x + \beta_2(x)} \); \( \beta_2(y) = \frac{\mu_{40}}{\mu^2_{20}} \); \( \beta_2(x) = \frac{\mu_{40}}{\mu^2_{20}} \); \( h = \frac{\mu_{22}}{(\mu_{20}, \mu_{20})} \); \( C = \frac{(h - 1)}{\beta_2(x) - 1} \) and

\[
\mu_{rs} = \frac{1}{N} \sum_{i=1}^{N} (y_i - \bar{Y})(x_i - \bar{X})^r
\]

(56)
From (1.8), we see that the estimator $t_2$ suggested by Upadhyay and Singh (1999) is a biased estimator. In some application bias is disadvantageous. This led authors to suggest almost unbiased estimators of $S_y^2$.

2. A class of ratio-type estimators

Consider $t_{Ri} = s_y^2 \left( \frac{s_x^2 + \beta_2(x)}{s_x^2 + \beta_2(x)} \right)^i$ such that $t_{Ri} \in R$, for $i = 1, 2, 3$; where $R$ denotes the set of all possible ratio-type estimators for estimating the population variance $S_y^2$. We define a class of ratio-type estimators for $S_y^2$ as –

$$t_r = \sum_{i=1}^{3} w_i t_{Ri} \in R,$$  \hspace{1cm} (2.1)

where $\sum_{i=1}^{3} w_i = 1$ and $w_i$ are real numbers. \hspace{1cm} (2.2)

For simplicity we assume that the population size $N$ is large enough so that the fpc terms are ignored. We write

$$s_y^2 = S_y^2 (1 + e_0), s_x^2 = S_x^2 (1 + e_1)$$

such that $E(e_0) = E(e_1) = 0$.

Noting that for large $N$, $\frac{1}{N} \equiv 0$ and $\frac{n}{N} \equiv 0$, and thus to the first degree of approximation,

$$E(e_0^2) = \frac{\beta_2(y) - 1}{n}, \ E(e_1^2) = \frac{\beta_2(x) - 1}{n}, \ E(e_0 e_1) = \frac{(h - 1)}{n} = \frac{[\beta_2(x) - 1]C}{n}.$$  

Expressing (2.1) in terms of $e$’s we have

$$t_r = s_y^2 (1 + e_0) \sum_{i=1}^{3} a_i \left( 1 + \theta e_1 \right)^i$$  \hspace{1cm} (2.3)
Assume that $|\phi e_i| < 1$ so that $(1 + \phi e_i)^i$ is expandable. Thus expanding the right hand side of the above expression (2.3) and retaining terms up to second power of e’s, we have

$$t_r = S^2_y \left[ 1 + e_o - \sum_{i=1}^{3} a_i \left( \phi e_i + \phi e_o e_i - \left( \frac{i+1}{2} \right) \theta^2 e_i^2 \right) \right]$$

or

$$t_r - S^2_y = S^2_y \left[ e_o - \sum_{i=1}^{3} a_i \left( \phi e_i + \phi e_o e_i - \left( \frac{i+1}{2} \right) \theta^2 e_i^2 \right) \right]$$

(2.4)

Taking expectation of both sides of (2.3) we get the bias of $t_r$, to the first degree of approximation, as

$$B(t_r) = \frac{S^2_y}{2n} \left[ \beta_2(x) - 1 \sum_{i=1}^{3} a_i \theta (\theta - 2C) \right]$$

(2.5)

Squaring both sides of (2.4), neglecting terms involving power of e’s greater than two and then taking expectation of both sides, we get the mean-squared error of $t_r$ to the first degree of approximation, as

$$MSE(t_r) = \frac{S^4_y}{n} \left[ \beta_2(y) - 1 \right] + R \left[ \theta^2 \beta_2(x) - 1 \right] \left[ \theta R_1 - 2C \right]$$

(2.6)

where $R_1 = \sum_{i=1}^{3} i w_i$

(2.7)

Minimizing the MSE of $t_r$ in (2.7) with respect to $R_1$ we get the optimum value of $R_1$ as

$$R_1 = \frac{C}{\theta}$$

(2.8)

Thus the minimum MSE of $t_r$ is given by

$$\text{min.MSE}(t_r) = \frac{S^4_y}{n} \left[ \beta_2(y) - 1 \right] - \left( \beta_2(x) - 1 \right) C^2$$
\[ S_3 = \frac{1}{n} \left[ \beta_2(y) - 1 \right] (1 - \rho^3_1) \]  \hspace{1cm} (2.9)

where \( \rho_1 = \frac{(h - 1)}{\sqrt{\beta_2(x) - 1} \beta_2(y) - 1} \) is the correlation coefficient between \((y - \bar{y})^2\) and \((x - \bar{x})^2\).

From (2.2), (2.7) and (2.8) we have

\[ \sum_{i=1}^{3} w_i = 1 \]  \hspace{1cm} (2.10)

and

\[ \sum_{i=1}^{3} iw_i = \frac{C}{\theta} = \rho_1 \left[ \frac{\beta_2(y) - 1}{\beta_2(x) - 1} \right]^{1/2} \]  \hspace{1cm} (2.11)

From (2.10) and (2.11) we have three unknowns to be determined from two equations only. It is therefore, not possible to find a unique value of the constants \( w_i ' s(i = 1,2,3) \). Thus in order to get the unique values of the constants \( w_i ' s(i = 1,2,3) \), we shall impose a linear constraint as

\[ B(t_r) = 0 \]  \hspace{1cm} (2.12)

which follows from (2.5) that

\[ (\theta - C)a_1 + (3\theta - 2C)a_2 + (6\theta - 3C)a_3 = 0 \]  \hspace{1cm} (2.13)

Equation (2.10), (2.11) and (2.13) can be written in the matrix form as

\[
\begin{bmatrix}
1 & 1 & 1 \\
1 & 2 & 3 \\
(\theta - C) & (3\theta - 2C) & (6\theta - 3C)
\end{bmatrix}
\begin{bmatrix}
w_1 \\
w_2 \\
w_3
\end{bmatrix}
= \begin{bmatrix}
1 \\
C / \theta \\
0
\end{bmatrix}
\]  \hspace{1cm} (2.14)

Using (2.14) we get the unique values of \( w_i ' s(i = 1,2,3) \) as
\[ w_1 = \frac{1}{\theta^2} \left[ 3\theta^2 - 3\theta C + C^2 \right] \]
\[ w_2 = \frac{1}{\theta^2} \left[ -3\theta^2 + 5\theta C - 2C^2 \right] \]
\[ w_3 = \frac{1}{\theta^2} \left[ \theta^2 - 2\theta C + C^2 \right] \]  

Use of these \( w_i \)'s (\( i = 1, 2, 3 \)) remove the bias up to terms of order \( o(n^{-1}) \) at (2.1).

Substitution of (2.14) in (2.1) yields the almost unbiased optimum ratio-type estimator of the population variance \( S_y^2 \).

3. A class of product-type estimators

Consider \( t_{pi} = S_y^2 \left[ \frac{S_i^2 + \beta_2(x)}{S_y^2 + \beta_2(x)} \right] \) such that \( t_{pi} \in P \), for \( i = 1, 2, 3 \); where \( P \) denotes the set of all possible product-type estimators for estimating the population variance \( S_y^2 \).

We define a class of product-type estimators for \( S_y^2 \) as –

\[ t_p = \sum_{i=1}^{3} k_i t_{pi} \in P, \]  

where \( k_i \)'s (\( i = 1, 2, 3 \)) are suitably chosen scalars such that

\[ \sum_{i=1}^{3} k_i = 1 \text{ and } k_i \text{ are real numbers.} \]

Proceeding as in previous section, we get

\[ B(t_p) = \frac{S_y^2}{2n} \left[ \left\{ \beta_2(x) - 1 \right\} \sum_{i=1}^{3} i \alpha_i \theta (\theta i + 2C - \theta) \right] \]  

(3.2)
\[ \text{MSE}(t_p) = \frac{S_y^4}{n} \left[ \{ \beta_2(y) - 1 \} + R_2 \theta \left( \{ \beta_2(x) - 1 \} (\theta R_2 + 2C) \right) \right] \] (3.3)

where, \[ R_2 = \sum_{i=1}^{3} i k_i \] (3.4)

Minimizing the MSE of \( t_p \) in (3.4) with respect to \( R_2 \), we get the optimum value of \( R_2 \) as

\[ R_2 = -\frac{C}{\theta} \] (3.5)

Thus the minimum MSE of \( t_p \) is given by

\[ \text{min } \text{MSE}(t_p) = \frac{S_y^4}{n} \{ \beta_2(y) - 1 \} (1 - \rho_y^2) \] (3.7)

which is same as that of minimum MSE of \( t \), at (2.9).

Following the approach of previous section, we get

\[
\begin{align*}
k_1 &= \frac{1}{\theta^2} \left[ 3\theta^2 + 2\theta C + C^2 \right] \\
k_2 &= \frac{1}{\theta^2} \left[ 3\theta^2 + 3\theta C + 2C^2 \right] \\
k_3 &= \frac{1}{\theta^2} \left[ \theta^2 + \theta C + C^2 \right]
\end{align*}
\] (3.8)

Use of these \( k_i \)'s (i=1,2,3) removes the bias up to terms of order \( O(n^{-1}) \) at (3.1).

4. **Empirical Study**

The data for the empirical study are taken from two natural population data sets considered by Das (1988) and Ahmed et. al. (2003).

**Population I** – Das (1988)

The variables and the required parameters are:

X: number of agricultural laborers for 1961.
Y: number of agricultural laborers for 1971.

\[ \beta_1(x) = 38.8898, \beta_2(y) = 25.8969, h=26.8142, S_x^2 = 1654.44. \]

**Population II** – Ahmed et. al. (2003)

The variables and the required parameters are:

X: number of households

Y: number of literate persons

\[ \beta_1(x) = 8.05448, \beta_2(y) = 10.90334, S_x^2 = 11838.85, h=7.31399. \]

In table 4.1 the values of scalars \( w_i \)'s (i=1,2,3) and \( k_i \)'s (i=1,2,3) are listed.

**Table 4.1: Values of scalars \( w_i \)'s and \( k_i \)'s (i=1,2,3)**

<table>
<thead>
<tr>
<th>Scalars</th>
<th>Population</th>
<th>Scalars</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
<td>II</td>
<td>I</td>
</tr>
<tr>
<td>( w_1 )</td>
<td>1.3942</td>
<td>1.1154</td>
<td>( k_1 )</td>
</tr>
<tr>
<td>( w_2 )</td>
<td>-0.4858</td>
<td>-0.1261</td>
<td>( k_2 )</td>
</tr>
<tr>
<td>( w_3 )</td>
<td>0.0916</td>
<td>0.0109</td>
<td>( k_3 )</td>
</tr>
</tbody>
</table>

Using these values of \( w_i \)'s and \( k_i \)'s (i=1,2,3) given in table 4.1, one can reduce the bias to the order \( O(n^{-1}) \) respectively, in the estimators \( t_r \) and \( t_p \) at (2.1) and (3.1).

In table 4.2 percent relative efficiency (PRE) of \( s_{t_1,t_2,t_3}^2 \) (in optimum case) and \( t_p \) (in optimum case) are computed with respect to \( s_{t_1}^2 \).
Table 4.2: PRE of different estimators of $S_y^2$ with respect to $s_y^2$

<table>
<thead>
<tr>
<th>Estimators</th>
<th>PRE ($\cdot S_y^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Population I</td>
</tr>
<tr>
<td>$S_y^2$</td>
<td>100</td>
</tr>
<tr>
<td>$t_1$</td>
<td>223.14</td>
</tr>
<tr>
<td>$t_2$</td>
<td>235.19</td>
</tr>
<tr>
<td>$t_r$ (optimum)</td>
<td>305.66</td>
</tr>
<tr>
<td>$t_p$ (optimum)</td>
<td>305.66</td>
</tr>
</tbody>
</table>

Table 4.2 clearly shows that the suggested estimators $t_r$ and $t_p$ in their optimum case are better than the usual unbiased estimator $s_y^2$, Isaki’s (1983) estimator $t_1$ and Upadhayaya and Singh (1999) estimator $t_2$.

References


