Efficient Estimators for Population Variance using Auxiliary Information

Subhash Kumar Yadav

Dept. of Maths & Stats, Dr. RML Avadh University,
Faizabad, U.P., India
E-mail: drskystats@gmail.com

Abstract

In this paper a family of efficient estimators, estimating population variance of the variable under study using auxiliary information has been proposed. The expressions for its bias and mean squared error (MSE) have been obtained up to \( O(n^{-1}) \). A comparison has been made with the general family of estimators for population variance of R. Singh et.al. (2007), which contains some well known estimators of population variance as a particular member such as Isaki (1983), Upadhyaya and Singh (1999) and Kadilar and Cingi (2006) etc. An improvement has been shown over above family of estimators through an empirical study.

Keywords: Auxiliary information, variance estimator, bias, mean squared error, efficiency.

Mathematics Subject Classification 2000: Primary 62D05 Secondary 62F10

Introduction and Notation

The use of auxiliary information may increase the precision of the estimators. When the variable under study \( Y \) is highly correlated with the auxiliary variable \( X \), ratio and product type estimators are used for improved estimation of population parameters. Many authors have proposed ratio type estimators for the estimation of population variance using different known parameters of the auxiliary variable. Isaki (1983) was the first who used auxiliary information to estimate the variance of the variable under study. He has shown that his estimator \( t_1 \) is better than the usual estimator \( t_0 \), which does not utilize the auxiliary information in the sense of having lesser mean squared error. In the series of improvement Upadhyaya and Singh (1999) has given an
estimator using the known population coefficient of kurtosis of auxiliary variable and he showed that his estimator $t_a$ is better than $t_0$ and $t_1$. Kadilar and Cingi (2006) proposed an estimator which utilizes the known population coefficient of variation and showed that his estimator is better than the all above estimators. In the present study, we suggest a new family of estimators for estimating population variance of the variable under study.

**Material and Methods**

Let the population consists of $N$ units and a sample of size $n$ is drawn from this population using simple random sampling without replacement. Let $Y_i$ and $X_i$ be the values for the $i$th unit ($i = 1, 2, \ldots, N$) of the population for the study variable and auxiliary variable respectively. Further, let $\bar{Y}$ and $\bar{X}$ be the sample means of the study and auxiliary variable respectively.

In order to study the large sample properties of the proposed family of estimators, we define

\[
s_y^2 = S_y^2(1 + e_0) \quad \text{and} \quad s_x^2 = S_x^2(1 + e_1) \quad \text{with} \quad E(e_i) = 0, \quad i = (0,1)
\]

in case of simple random sampling without replacement, ignoring finite population correction term, the following expectations could be obtained either directly or by the method due to Kendall and Stuart (1977) as

\[
E(e_0^2) = \frac{1}{n} (\lambda_{40} - 1), \quad E(e_1^2) = \frac{1}{n} (\lambda_{44} - 1) \quad \text{and} \quad E(e_0 e_1) = \frac{1}{n} (\lambda_{22} - 1)
\]

Where $\lambda_s = \frac{\mu_{rs}}{\mu_{00}}$ and $\mu_{rs} = \frac{1}{N} \sum_{i=1}^{N} (y_i - \bar{Y})^r (x_i - \bar{X})^s$, $(r, s) = 0, 1, 2, 3, 4$

**Suggested family of estimators**

R. Singh et al. (2007) suggested a family of estimators for population variance as

\[
t = s_y^2 \left[ \frac{(aS_y^2 - b)}{\alpha(aS_y^2 - b) + (1 - \alpha)(aS_y^2 - b)} \right]
\]  

where $(a \neq 0)$, $b$ are either real numbers or the function of the known parameters of the auxiliary variable $x$ such as coefficient of variation $C_x$ and coefficient of kurtosis $\beta_2(x) = \lambda_{44}$.

The MSE of this family of estimators is given by

\[
MSE(t) = \frac{S_y^2}{n} \left[ (\lambda_{40} - 1) - 2\alpha \nu (\lambda_{22} - 1) + \alpha^2 \nu^2 (\lambda_{44} - 1) \right]
\]  

(3.2)
Efficient Estimators for Population Variance

where

$$v = \frac{aS_x^2}{(as_x^2 - b)}$$

The minimum MSE for $\alpha_{op} = \frac{C}{v}$ where $C = \frac{\lambda_{a2} - 1}{\lambda_{b4} - 1}$ is

$$MSE_{\min}(t) = \frac{S_x^2}{n} [(\lambda_{a0} - 1) - (\lambda_{b4} - 1)]$$  \hspace{1cm} (3.3)

The ratio type estimators, given in table 1 are in t-family with the MSE as

$$MSE(t_i) = \frac{S_x^2}{n} \left[ (\lambda_{a0} - 1) - 2\alpha v_i (\lambda_{a2} - 1) + \alpha^2 v_i^2 (\lambda_{b4} - 1) \right]$$  \hspace{1cm} (3.4)

$$i = 0, 1, \ldots, 6$$

With

$$v_0 = 0, \hspace{0.2cm} v_1 = 1, \hspace{0.2cm} v_2 = \frac{S_x^2}{(S_x^2 - C_x)}, \hspace{0.2cm} v_3 = \frac{S_x^2}{(S_x^2 - \lambda_{b4})}, \hspace{0.2cm} v_4 = \frac{S_x^2 \lambda_{b4}}{(S_x^2 \lambda_{b4} - C_x)},$$

$$v_5 = \frac{S_x^2 C_x}{(S_x^2 C_x - \lambda_{b4})} \hspace{0.2cm} \text{and} \hspace{0.2cm} v_6 = \frac{S_x^2}{(S_x^2 + \lambda_{b4})}.$$

Table 1: Some members of t-family of estimators.

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Values of $\alpha$</th>
<th>$a$</th>
<th>$b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_0 = s_x^2$</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$t_1 = \frac{s_x^2}{s_x^2} S_x^2$</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
| Isaki (1983) estimator
| $t_2 = \frac{s_x^2}{s_x^2 - C_x} [S_x^2 - C_x]$ | 1     | 1   | $C_x$ |
| Kadilar & Cingi (2006) estimator
| $t_3 = \frac{s_x^2}{s_x^2 - \beta_2(x)} [S_x^2 - \beta_2(x)]$ | 1     | 1   | $\beta_2(x)$ |
| $t_4 = \frac{s_x^2}{\beta_2(x) s_x^2 - C_x} [\beta_2(x) S_x^2 - C_x]$ | 1     | $\beta_2(x)$ | $C_x$ |
| $t_5 = \frac{s_x^2}{C_x s_x^2 - \beta_2(x)} [C_x S_x^2 - \beta_2(x)]$ | 1     | $C_x$ | $\beta_2(x)$ |
Motivated by Nursel Koyuncu and Cem Kadilar (2009), we propose a new family of estimators for population variance as

$$\xi = kS_y^2 \frac{(aS_y^2 - b)}{[\alpha(as_y^2 - b) + (1 - \alpha)(aS_y^2 - b)]}$$  \hspace{1cm} (3.5)

where $k$ is suitably chosen constant to be determined.

Now expressing the estimator $\xi$ \hspace{0.5cm} in terms of $e_i \hspace{0.5cm} (i = 0,1)$, (2.5) can be written as

$$\xi = kS_y^2 (1 + e_o)(1 + \alpha ve_i)^{-1} \hspace{1cm} (3.6)$$

Expanding the right hand side of (2.6) to the first order of approximation and subtracting $S_y^2$ \hspace{0.5cm} from both the sides, we get

$$\xi = kS_y^2 (1 + e_o)(1 - \alpha ve_i + \alpha^2 v^2 e_i^2)$$

Now

$$\xi - S_y^2 = kS_y^2 (1 + e_o - \alpha ve_i + \alpha^2 v^2 e_i^2 - \alpha ve_o e_i) - S_y^2$$ \hspace{1cm} (3.7)

Taking expectation on both sides of (2.7), we get the bias of the estimator $\xi$ as

$$B(\xi) = \frac{kS_y^2}{n} [\alpha^2 v^2 (\lambda_{04} - 1) - \alpha v (\lambda_{22} - 1)] + S_y^2 (k - 1)$$ \hspace{1cm} (3.8)

Squaring both sides to equation (2.7), it gives

$$(\xi - S_y^2)^2 = k^2 S_y^4 (1 + e_o - \alpha ve_i + \alpha^2 v^2 e_i^2 - \alpha ve_o e_i)^2 + S_y^4 - 2kS_y^4 (1 + e_o - \alpha ve_i + \alpha^2 v^2 e_i^2 - \alpha ve_o e_i)$$

Now taking expectation both sides, we get MSE of $\xi$ upto $O(n^{-1})$ as

$$MSE(\xi) = S_y^4 \frac{1}{n} [k^2 (\lambda_{40} - 1) + (3k^2 - 2k)\alpha^2 v^2 (\lambda_{04} - 1)$$

$$- 2\alpha(2k^2 - k)(\lambda_{22} - 1)] + (k - 1)^2 \hspace{1cm} (3.9)$$

The minimum of $MSE(\xi)$ is obtained for optimum value of $k$ which is $k_{opt} = \frac{A}{B}$.

Where \hspace{0.5cm} $A = \frac{1}{n} [\alpha^2 v^2 (\lambda_{04} - 1) - \alpha v (\lambda_{22} - 1)] + 1$
Efficient Estimators for Population Variance

And $B = \frac{1}{n}[(\lambda_{40} - 1) + 3\alpha^2 v^2 (\lambda_{04} - 1) - 4\alpha v(\lambda_{22} - 1)] + 1$

Thus the minimum MSE of the family $\xi$ of estimators is

$$MSE_{\text{min}}(\xi) = S_1^2 \left[1 - \frac{A^2}{B}\right]$$

(3.10)

The ratio type estimators, given in table 2 are in $\xi$-family with the MSE as

$$MSE(\xi_i) = S_1^2 \left[\frac{1}{n} (k^2 (\lambda_{40} - 1) + (3k^2 - 2k)\alpha^2 v_i^2 (\lambda_{04} - 1) - 2\alpha v_i (2k^2 - k)(\lambda_{22} - 1)) + (k - 1)^2 \right]$$

$i = 1, \ldots, 6$

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Values of $\alpha$</th>
<th>$A$</th>
<th>$b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\xi_1 = k \frac{s^2_y}{s^2_x} S^2_x$</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>$\xi_2 = k \frac{s^2_y}{s^2_x - C_x} [S^2_x - C_x]$</td>
<td>1</td>
<td>1</td>
<td>$C_x$</td>
</tr>
<tr>
<td>$\xi_3 = k \frac{s^2_y}{s^2_x - \beta_2(x)} [S^2_x - \beta_2(x)]$</td>
<td>1</td>
<td>1</td>
<td>$\beta_2(x)$</td>
</tr>
<tr>
<td>$\xi_4 = k \frac{s^2_y}{s^2_x \beta_2(x) - C_x} [S^2_x \beta_2(x) - C_x]$</td>
<td>1</td>
<td>$\beta_2(x)$</td>
<td>$C_x$</td>
</tr>
<tr>
<td>$\xi_5 = k \frac{s^2_y}{s^2_x C_x - \beta_2(x)} [S^2_x C_x - \beta_2(x)]$</td>
<td>1</td>
<td>$C_x$</td>
<td>$\beta_2(x)$</td>
</tr>
<tr>
<td>$\xi_6 = k \frac{s^2_y}{s^2_x + \beta_2(x)} [S^2_x + \beta_2(x)]$</td>
<td>1</td>
<td>1</td>
<td>$-\beta_2(x)$</td>
</tr>
</tbody>
</table>

Many more estimators can be formed just by putting different values of the parameters $a$ and $b$ of auxiliary variable.

**Efficiency comparison**

To the first order of approximation, after ignoring finite population correction, the variance and MSE of estimators $t_0$ and $t_1$ respectively are given as
\[ V(t_0) = \frac{S^4}{n} [\hat{\lambda}_{40} - 1] \]

\[ MSE(t_1) = \frac{S^4}{n} [(\hat{\lambda}_{40} - 1) - 2(\hat{\lambda}_{22} - 1) + (\hat{\lambda}_{04} - 1)] \]

and

\[ V(t_0) - MSE(t_1) = \frac{S^4}{n} [2(\hat{\lambda}_{22} - 1) - (\hat{\lambda}_{04} - 1)] > 0 \text{ if,} \]

\[ 2(\hat{\lambda}_{22} - 1) > (\hat{\lambda}_{04} - 1) \]

When this condition is attended, \( t_1 \) is more efficient than \( t_0 \).

Similarly \( t \)-family is more efficient than \( t_1 \) if,

\[ MSE(t_1) < MSE(t_i), \quad i = 2, 3, 4, 5, 6. \]

that is

\[ 2(\hat{\lambda}_{22} - 1) < (1 + v_i)(\hat{\lambda}_{04} - 1) \]

Now suggested estimators \( \xi_i \) are more efficient than \( t_i \), (\( i = 1, \ldots, 6 \)) estimators if,

\[ MSE_{\min}(\xi_i) < MSE(t_i), \quad i = 2, 3, 4, 5, 6. \]

that is

\[ [1 - \frac{A^2}{B}] < \frac{1}{n} [(\hat{\lambda}_{40} - 1) + v_i^2 (\hat{\lambda}_{04} - 1) - 2v_i (\hat{\lambda}_{22} - 1)] \]

and the minimum MSE of \( \xi \)-family is less than the minimum MSE of \( t \)-family if,

\[ MSE_{\min}(\xi_i) < MSE_{\min}(t_i), \quad i = 2, 3, 4, 5, 6. \]

that is

\[ [1 - \frac{A^2}{B}] < \frac{1}{n} [(\hat{\lambda}_{40} - 1) - (\hat{\lambda}_{22} - 1)] \]

**Empirical study**

We used the data in R. Singh et.al (2007), given in table3, which was early used by Kadilar and Cingi (2004) for the comparison of efficiencies of the \( \xi \)-family to \( t \)-family of estimators of R. Singh et.al (2007).
Table 3: Data Statistics.

\[
\begin{align*}
N &= 106, \ n = 20, \ C_r = 4.18, \ C_s = 2.02, \ \bar{Y} = 15.37, \ \bar{X} = 243.76 \\
S_y &= 64.25, \ S_x = 491.89, \ \lambda_{04} = 25.71, \ \lambda_{40} = 80.13, \ \lambda_{22} = 33.30
\end{align*}
\]

The MSE of \( \xi_i \) and \( t_i, \) (i = 1, ..., 6) and percentage relative efficiency (PRE) each \( \xi_i \) to its corresponding \( t_i \) is given in the following table4.

Result

From the table4, we see that the proposed \( \xi \)-family of estimators is better than the \( t \)-family of estimators of R. Singh et.al (2007) in the sense of having lesser mean squared error. Table4 also shows that proposed family is much better than the estimator \( t_0 \) not utilizing auxiliary information.

Table 4: MSE and Efficiencies comparison.

<table>
<thead>
<tr>
<th>( t )-family</th>
<th>( \xi )-family</th>
<th>PRE of estimator ( \xi_i ) over ( t_i )</th>
<th>PRE of ( t_i ) over ( t_0 )</th>
<th>PRE of ( \xi_i ) over ( t_0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimator</td>
<td>MSE</td>
<td>Estimator</td>
<td>MSE</td>
<td>( S_y^4 ) (1.962)</td>
</tr>
<tr>
<td>( t_1 )</td>
<td>( S_y^4 ) (1.9620203)</td>
<td>( \xi_2 )</td>
<td>( S_y^4 ) (.8252257)</td>
<td>237.7556</td>
</tr>
<tr>
<td>( t_2 )</td>
<td>( S_y^4 ) (1.9619192)</td>
<td>( \xi_3 )</td>
<td>( S_y^4 ) (.8251859)</td>
<td>237.7548</td>
</tr>
<tr>
<td>( t_3 )</td>
<td>( S_y^4 ) (1.962)</td>
<td>( \xi_4 )</td>
<td>( S_y^4 ) (.825229)</td>
<td>237.7520</td>
</tr>
<tr>
<td>( t_4 )</td>
<td>( S_y^4 ) (1.9619601)</td>
<td>( \xi_5 )</td>
<td>( S_y^4 ) (.8252077)</td>
<td>237.7535</td>
</tr>
<tr>
<td>( t_5 )</td>
<td>( S_y^4 ) (1.9620805)</td>
<td>( \xi_6 )</td>
<td>( S_y^4 ) (.8252723)</td>
<td>237.7595</td>
</tr>
</tbody>
</table>

Conclusion

From the results of the empirical study and theoretical discussions, it is concluded that the proposed \( \xi \)-family of estimators for estimating population variance under optimum condition perform much better than the usual estimator \( s_y^2 \) and also better than \( t \)-family of estimators proposed by R. Singh et.al. (2007), which contains some well known estimators of Issaki (1983), Upadhyaya and Singh (1999) and Kadilar and Cingi (2006) etc.
References


