A DUAL TO VARIANCE RATIO-TYPE ESTIMATOR IN SIMPLE RANDOM SAMPLING

Javid Shabbir

Department of Statistics, Quaid-i-Azam University, Islamabad 45320, Pakistan

Received February 2006, accepted July 2006

Communicated by Prof. Dr. Q. K. Ghori

Abstract: We propose a dual to variance ratio-type estimator to estimate the finite population variance. The optimum mean square error of this estimator is equal to variance of a linear regression estimator and is better than the usual unbiased variance estimator and Isaki (Jour Amer. Statis Assoc. 78:117-123, 1983) estimator. We use the jackknife technique to make the proposed estimator unbiased. The validity of the proposed estimator is examined by using the various data sets.

Keywords: Ratio estimator, auxiliary variable, bias, mean square error (MSE), efficiency, Jacknife technique

Introduction

Let \( U \) be a finite population consisting of \( N \) units \( U_1, U_2, \ldots, U_N \) from which a sample of size \( n \) is to be drawn by simple random sampling without replacement (SRSWOR). Let \( y \) and \( x \) denote the study and auxiliary variables respectively and \( y \) is positively correlated with \( x \). Let \( S_y^2 = \sum_{i=1}^{N} (y_i - \bar{Y})^2 / (N-1) \) and 
\[
S_x^2 = \sum_{i=1}^{N} (x_i - \bar{X})^2 / (N-1)
\]
derive the population variances of \( y \) and \( x \) respectively. Similarly, one can obtain the sample variances \( s_y^2 = \sum_{i=1}^{n} (y_i - \bar{Y})^2 / (n-1) \) and \( s_x^2 = \sum_{i=1}^{n} (x_i - \bar{X})^2 / (n-1) \) of \( y \) and \( x \) respectively. Let \( C_y = S_y^2 / \bar{Y} \) and \( C_x = S_x^2 / \bar{X} \) denote the coefficient of variations of \( y \) and \( x \), respectively. To estimate \( S_y^2 \), it is assumed that \( S_y^2 \) is known.

Definition:

Let \( \Delta_0 = (s_y^2 - S_y^2) / S_y^2 \) and \( \Delta_i = (s_i^2 - S_i^2) / S_i^2 \), therefore \( E(\Delta_0) = E(\Delta_i) = 0 \), \( E(\Delta_i^2) = \left( \frac{1}{n} - \frac{1}{N} \right) (\lambda_{xx} - 1) \),
\[
E(\Delta_0) = \left( \frac{1}{n} - \frac{1}{N} \right) (\lambda_{xx} - 1), \quad \text{where} \quad \lambda_{pq} = \mu_{pq} / (\mu_{02} \mu_{02}^{1/2}),
\]
and \( \mu_{pq} = \sum_{i=1}^{N} (y_i - \bar{Y}) (x_i - \bar{X}) / (N-1) \).

The conventional unbiased variance estimator is defined as
\[
\hat{S}_0^2 = s_y^2
\]
The variance of \( \hat{S}_0^2 \) is given by
\[
\text{Var}(\hat{S}_0^2) = \left( \frac{1}{n} - \frac{1}{N} \right) S_y^4 (\lambda_{40} - 1).
\]
Isaki [2] suggested the following variance ratio estimator
\[
\hat{S}_1^2 = s_y^2 \left( \frac{S_x^2}{s_x^2} \right).
\]
The bias and MSE to first order of approximation are given by
Dual to variance ratio-type estimator

\[
Bias(\hat{S}^2_1) = \left(1 - \frac{1}{n} \right) S^2_i \left[ (\lambda_{04} - 1) - (\lambda_{22} - 1) \right]
\]  
(4)

and

\[
MSE(\hat{S}^2_1) = \left(1 - \frac{1}{n} \right) S^2_i \left[ (\lambda_{04} - 1)^2 + (\lambda_{04} - 1) - 2(\lambda_{22} - 1) \right].
\]  
(5)

Now we propose the following dual to variance ratio-type estimator.

**Proposed Estimator**

The following proposed estimator is the combination of usual unbiased variance estimator given in (1) and dual to variance ratio estimator with weights \(w_1\) and \(w_2\) such that \(w_1 + w_2 = 1\) as

\[
\hat{S}^2_p = w_1 S^2_y + w_2 S^2_y \left( \frac{s^2}{S^2_y} \right),
\]  
(6)

where \(s^2 = \frac{N s^2 - n \hat{S}^2}{N - n}\) is due to Srivenkataramana [6].

From (6), we have

\[
\hat{S}^2_p = w_1 S^2_y (1 + \Delta_0) + w_2 S^2_y (1 + \Delta_0) (1 + \eta \Delta_1),
\]  
(7)

where \(\eta = n/(N - n)\).

From (7), we have

\[
\hat{S}^2_p - \hat{S}^2_y = S^2_y [\Delta_0 - w_2 \eta (\Delta_1 + \Delta_0 \Delta_1)].
\]  
(8)

Solving (8), we get the bias and MSE of \(\hat{S}^2_p\) which are given by

\[
Bias(\hat{S}^2_p) = E(\hat{S}^2_p - S^2_y) = -\left(1 - \frac{1}{n} \right) w_2 S^2_y \eta (\lambda_{22} - 1)
\]  
(9)

and

\[
MSE(\hat{S}^2_p) = E(\hat{S}^2_p - S^2_y)^2 = S^2_y E[\Delta_0 - w_2 \eta \Delta_1]^2.
\]  
(10)

From (2) and (10), it follows that

(i) \(MSE(\hat{S}^2_p) < Var(\hat{S}^2_i)\) if either \(\frac{(\lambda_{04} - 1)}{(\lambda_{04} - 1)} > \frac{1}{2} w \eta\).  
(11)

From (5) and (10), it follows that

(ii) \(MSE(\hat{S}^2_p) < MSE(\hat{S}^2_i)\) if either \(\frac{(\lambda_{04} - 1)}{(\lambda_{04} - 1)} < \frac{1}{2} (w \eta + 1)\).  
(12)

Thus, from (12), the estimator \(\hat{S}^2_p\) is more efficient than \(\hat{S}^2_i\) and \(\hat{S}^2_i\) if

\[
\frac{1}{2} w \eta < \frac{(\lambda_{22} - 1)}{(\lambda_{04} - 1)} \leq \frac{1}{2} (w \eta + 1)
\]  
(13)

The above inequality is obviously true.

The proposed estimator \(\hat{S}^2_p\) is efficient if the following conditions are satisfied.

Cond. (i) \(w \eta \leq (\lambda_{22} - 1) \leq (\lambda_{04} - 1)\).

Cond. (ii) \(w \eta \leq (\lambda_{22} - 1) \leq (\lambda_{04} - 1)\).

Cond. (iii) \(\frac{MSE(\hat{S}^2_p)}{MSE(\hat{S}^2_i)} \leq 1\).

Cond. (iv) \(\frac{MSE(\hat{S}^2_p)}{MSE(\hat{S}^2_i)} \leq 1\).

The optimum choice of \(w_2\) minimizing (10) is given by
Substitution of optimum value of $w^*_2$ in (10), we get the minimum MSE of $\hat{S}_p^2$ which is given by

$$MSE(\hat{S}_p^2)_{\text{min}} = \left(\frac{1}{n} - \frac{1}{N}\right)S^2_{\eta}[(\lambda_{40} - 1) - \frac{(\lambda_{22} - 1)^2}{(\lambda_{04} - 1)}]. \ (14)$$

The MSE of $\hat{S}_p^2$ in (14) is equal to variance of the linear regression estimator, $\hat{S}_p^2 = s^2_\eta + b(s^2_x - s^2_\eta)$

where $b = \frac{s^2_\eta(\lambda_{22} - 1)}{s^2_\eta(\lambda_{04} - 1)}$ is sample regression coefficient. It is noted that Equation (14) is independent of $\eta$.

**Efficiency comparison based on optimum MSE**

We compare the proposed estimator with the usual variance estimator and Isaki [2] estimator.

(i) By (2) and (12),

$$\text{Var}(\hat{S}_p^2) - MSE(\hat{S}_p^2)_{\text{min}} = \frac{(\lambda_{22} - 1)^2}{(\lambda_{04} - 1)} > 0; \ (\lambda_{04} - 1) > 0.$$  

(ii) By (5) and (12),

$$\text{Var}(\hat{S}_p^2)_{\text{max}} - MSE(\hat{S}_p^2)_{\text{max}} > 0 \quad \text{if} \quad \left(\frac{1}{n} - \frac{1}{N}\right)S^2_{\eta} \left(\sqrt{\frac{\lambda_{04} - 1}{\lambda_{40}} - 1}\right)^2 > 0.$$

Both of the above conditions are obviously true.

**Data and Results**

For comparison, we consider the following seven data sets from various sources.

**Population 1:** (Cochran [1], p. 325)

$\lambda_40 = 2.2387, \lambda_{04} = 2.2523, \lambda_{22} = 1.5432, \lambda_{21} = 0.4536, C_x = 0.1281, C_y = 0.1450, X = 58.8, \rho = 0.6515$.

**Population 2:** (Cochran [1], p. 152)

$\lambda_40 = 8.5362, \lambda_{04} = 7.3617, \lambda_{22} = 7.8780, \lambda_{21} = 0.2295, C_x = 1.0126, C_y = 0.9634, X = 103.1, \rho = 0.9820$.

**Population 3:** (Cochran [1], p. 203)

$\lambda_40 = 1.9249, \lambda_{04} = 2.5932, \lambda_{22} = 2.1149, \lambda_{21} = 0.1875, C_x = 0.1621, C_y = 0.1840, X = 56.9, \rho = 0.9937$.

**Population 4:** (Sukhatme and Sukhatme [7], p. 185)

$\lambda_40 = 3.1842, \lambda_{04} = 2.2030, \lambda_{22} = 2.5597, \lambda_{21} = 0.6665, C_x = 0.5625, C_y = 0.6163, X = 265.8, \rho = 0.977$.

**Population 5:** (Upadhyaya and Singh [8])

$\lambda_40 = 40.8536, \lambda_{04} = 48.1567, \lambda_{22} = 43.7615, \lambda_{21} = 5.9786, C_x = 2.1971, C_y = 2.1118, X = 2900.4, \rho = 0.977$.

**Population 6:** (Singh et al. [4])

$\lambda_40 = 40.608000.69, \lambda_{04} = 48.1567, \lambda_{22} = 43.7615, \lambda_{21} = 5.9786, C_x = 2.1971, C_y = 2.1118, X = 2900.4, \rho = 0.977$.

**Population 7:** (Singh et al. [4])

$\lambda_40 = 40.608000.69, \lambda_{04} = 48.1567, \lambda_{22} = 43.7615, \lambda_{21} = 5.9786, C_x = 2.1971, C_y = 2.1118, X = 2900.4, \rho = 0.977$.
$\lambda_{40} = 24.8969, \lambda_{04} = 37.8898, \lambda_{22} = 25.8142, $  
$\lambda_{21} = 3.4347, C_x = 1.61198, C_y = 104451, X = 250111, \rho = 0.7273.$

**Population 7:** (Singh [5])

$y$: Amount (in $1000) of real estate farm loans in different states during 1997,

$x$: Amount (in $1000) of non real estate farm loans in different states during 1997.

$N=50, n=8, S_y^2=7342021.5, S_x^2=1176526, \lambda_{40} = 3.5822, \lambda_{04} = 4.5247, \lambda_{22} = 2.8411, \lambda_{21} = 0.9387, C_x = 1.2352, C_y = 1.0529, X = 878.16, \rho = 0.8038.$

The relative efficiency ($RB$) and relative bias ($RE$) can be obtained by using the following expressions respectively

\[
RB = \frac{\text{Bias}(\hat{S}_j)}{\sqrt{\text{MSE}(\hat{S}_j)}}; j = I, P.
\]

and

\[
RE = \frac{\text{Var}(\hat{S}_i)}{\text{MSE}(\hat{S}_i)} \times 100; i = 0, I, P.
\]

The results are given in Tables 1, 2 and 3.

In Table 3, conditions (i) and (ii) are true only if $(\lambda_{22} - 1) \leq (\lambda_{04} - 1)$. Conditions (iii) and (iv) will always remain true.

### Table 1. $RB$ of different estimators.

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Pop. 1</th>
<th>Pop. 2</th>
<th>Pop. 3</th>
<th>Pop. 4</th>
<th>Pop. 5</th>
<th>Pop. 6</th>
<th>Pop. 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{S}_j^2$</td>
<td>0.1979</td>
<td>-0.1696</td>
<td>0.2746</td>
<td>-0.2115</td>
<td>0.3022</td>
<td>0.6234</td>
<td>0.3503</td>
</tr>
<tr>
<td>$\hat{S}_p^2$</td>
<td>-0.0648</td>
<td>-0.9093</td>
<td>-0.6321</td>
<td>-1.5412</td>
<td>-3.1324</td>
<td>-1.0730</td>
<td>-0.2448</td>
</tr>
</tbody>
</table>

### Table 2. $RE$ of different estimators w. r. t $\hat{S}_0^2$ in percentage.

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Pop. 1</th>
<th>Pop. 2</th>
<th>Pop. 3</th>
<th>Pop. 4</th>
<th>Pop. 5</th>
<th>Pop. 6</th>
<th>Pop. 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{S}_0^2$</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>$\hat{S}_j^2$</td>
<td>82.33</td>
<td>5310.92</td>
<td>320.81</td>
<td>815.61</td>
<td>2679.59</td>
<td>214.32</td>
<td>106.50</td>
</tr>
<tr>
<td>$\hat{S}_p^2$</td>
<td>121.38</td>
<td>7536.32</td>
<td>639.15</td>
<td>1347.98</td>
<td>3698.20</td>
<td>331.90</td>
<td>159.34</td>
</tr>
</tbody>
</table>

### Table 3. Conditional values.

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Pop. 1</th>
<th>Pop. 2</th>
<th>Pop. 3</th>
<th>Pop. 4</th>
<th>Pop. 5</th>
<th>Pop. 6</th>
<th>Pop. 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cond. (i)</td>
<td>0.1613</td>
<td>1.1689</td>
<td>0.4897</td>
<td>1.6809</td>
<td>0.8223</td>
<td>0.4526</td>
<td>0.2728</td>
</tr>
<tr>
<td>Cond. (ii)</td>
<td>0.4017</td>
<td>1.0812</td>
<td>0.6998</td>
<td>1.2965</td>
<td>0.9068</td>
<td>0.6728</td>
<td>0.5223</td>
</tr>
<tr>
<td>Cond. (iii)</td>
<td>0.8238</td>
<td>0.0133</td>
<td>0.1564</td>
<td>0.0742</td>
<td>0.0270</td>
<td>0.3013</td>
<td>0.6276</td>
</tr>
<tr>
<td>Cond. (iv)</td>
<td>0.6783</td>
<td>0.7047</td>
<td>0.5019</td>
<td>0.6051</td>
<td>0.7246</td>
<td>0.6457</td>
<td>0.6683</td>
</tr>
</tbody>
</table>
Unbiased Version of the Proposed Estimator

Since our proposed estimator \( \hat{S}_p^2 \) is biased, so we use the jacknifie technique to make the estimator unbiased. Following Sukhatme and Sukhatme [7], take \( n = 2m \) and split the sample at random into two sub-samples of \( m \) units each. Let \( s_{y_1}^2, s_{x_1}^2, (i=1, 2) \) be unbiased estimator of the population variance \( S_y^2 \) and \( S_x^2 \) and \( s_y^2 \) and \( s_x^2 \) be the sample variance based on entire sample. So the unbiased version of the proposed estimator is given by

\[
\hat{S}_{p(J)}^{2(U)} = \frac{(2N-n)}{N} \hat{S}_p^2 - \frac{(N-n)}{2n} \left( \hat{S}_{p1}^2 + \hat{S}_{p2}^2 \right),
\]

where \( \hat{S}_{p1}^2 = w_1s_{y_1}^2 + w_2s_{x_1}^2 \left( \frac{s_{y_2}^2}{S_y^2} \right), (i=1, 2) \) and \( w_1 + w_2 = 1 \).

The variance expression of the unbiased estimator \( \hat{S}_{p(J)}^{2(U)} \) can be derived easily. It is to be noted that the variance expression of \( \hat{S}_{p(J)}^{2(U)} \) and \( \text{MSE} \) expression of \( \hat{S}_p^2 \) are both equal. Hence we can prefer estimator \( \hat{S}_{p(J)}^{2(U)} \) as compared to \( \hat{S}_p^2 \) because of unbiasedness.

In conclusion, the unbiased version of the proposed dual to variance ratio-type estimator is as efficient as the linear regression estimator and is more efficient than the usual variance estimator \( \hat{S}_0^2 \) and Isaki [2] estimator \( \hat{S}_i^2 \). In Table 2 (Pop. 2), the efficiency of the proposed estimator is much higher as compared to the estimators \( \hat{S}_0^2 \) and \( \hat{S}_i^2 \). Also in Table 2 (Pop.1), the Isaki [2] estimator \( \hat{S}_0^2 \) is inferior and even worse than the usual variance estimator \( \hat{S}_0^2 \). The unbiased estimator \( \hat{S}_{p(J)}^{2(U)} \) is preferable as compared to the biased estimator \( \hat{S}_p^2 \).

References