Improved Exponential Estimator for Population Variance Using Two Auxiliary Variables

Published in:
Rajesh Singh, Jayant Singh, F. Smarandache (Editors)
STUDIES IN STATISTICAL INFERENACE, SAMPLING TECHNIQUES AND DEMOGRAPHY
InfoLearnQuest, Ann Arbor, USA, 2009
pp. 36 - 44
Abstract

In this paper exponential ratio and exponential product type estimators using two auxiliary variables are proposed for estimating unknown population variance $S^2_y$. Problem is extended to the case of two-phase sampling. Theoretical results are supported by an empirical study.

Key words: Auxiliary information, exponential estimator, mean squared error.

1. Introduction

It is common practice to use the auxiliary variable for improving the precision of the estimate of a parameter. Out of many ratio and product methods of estimation are good examples in this context. When the correlation between the study variate and the auxiliary variate is positive (high) ratio method of estimation is quite effective. On the other hand, when this correlation is negative (high) product method of estimation can be employed effectively. Let $y$ and $(x, z)$ denotes the study variate and auxiliary variates taking the values $y_i$ and $(x_i, z_i)$ respectively, on the unit $U_i (i=1, 2, \ldots, N)$, where $x$ is positively correlated with
y and z is negatively correlated with y. To estimate \( S^2_y = \frac{1}{(N-1)} \sum_{i=1}^{N} (y_i - \bar{y})^2 \), it is assumed that \( S^2_x = \frac{1}{(N-1)} \sum_{i=1}^{N} (x_i - \bar{X})^2 \) and \( S^2_z = \frac{1}{(N-1)} \sum_{i=1}^{N} (z_i - \bar{Z})^2 \) are known. Assume that population size \( N \) is large so that the finite population correction terms are ignored.

Assume that a simple random sample of size \( n \) is drawn without replacement (SRSWOR) from \( U \). The usual unbiased estimator of \( S^2_y \) is

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

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

When the population variance \( S^2_x = \frac{1}{(N-1)} \sum_{i=1}^{N} (x_i - \bar{X})^2 \) is known, Isaki (1983) proposed a ratio estimator for \( S^2_y \) as

\[
t_k = s^2_y \frac{S^2_x}{s^2_x}
\]

where \( s^2_x = \frac{1}{(n-1)} \sum_{i=1}^{n} (x_i - \bar{X})^2 \) is an unbiased estimator of \( S^2_x \).

Up to the first order of approximation, the variance of \( S^2_y \) and MSE of \( t_k \) (ignoring the finite population correction (fpc) term) are respectively given by

\[
\text{var}(s^2_y) = \left( \frac{S^4}{n} \right) \left[ \hat{\sigma}_{400} - 1 \right]
\]

(1.3)

\[
\text{MSE}(t_k) = \left( \frac{S^4}{n} \right) \left[ \hat{\sigma}_{400} + \hat{\sigma}_{040} - 2\hat{\sigma}_{220} \right]
\]

(1.4)
where \( \delta_{pq} = \frac{\mu_{pq}}{\mu_{q^2}} \),

\[
\mu_{pq} = \frac{1}{N} \sum_{i=1}^{N} \left( y_i - \bar{Y} \right)^p \left( x_i - \bar{X} \right)^q \left( z_i - \bar{Z} \right)^r ; p, q, r \text{ being the non-negative integers.}
\]

Following Bahl and Tuteja (1991), we propose exponential ratio type and exponential product type estimators for estimating population variance \( S_x^2 \) as –

\[
t_1 = s_y^2 \exp \left[ \frac{S_x^2 - S_y^2}{S_x^2 + s_y^2} \right]
\]

\[
t_2 = s_y^2 \exp \left[ \frac{s_x^2 - S_y^2}{s_x^2 + S_y^2} \right]
\]

2. Bias and MSE of proposed estimators

To obtain the bias and MSE of \( t_1 \), we write

\[
s_y^2 = S_y^2 (1 + e_0), \quad s_x^2 = S_x^2 (1 + e_1)
\]

Such that \( E(e_0) = E(e_1) = 0 \)

and \( E(e_0^2) = \frac{1}{n} (\partial_{040} - 1), \quad E(e_1^2) = \frac{1}{n} (\partial_{040} - 1), \quad E(e_0 e_1) = \frac{1}{n} (\partial_{220} - 1) \).

After simplification we get the bias and MSE of \( t_1 \) as

\[
B(t_1) \equiv \frac{S_y^2}{n} \left[ \frac{\partial_{040}}{8} - \frac{\partial_{220}}{2} + \frac{3}{8} \right]
\]

\[
\text{MSE}(t_1) \equiv \frac{S_y^2}{n} \left[ \frac{\partial_{040} + \partial_{040}}{4} - \partial_{220} + \frac{1}{4} \right]
\]

To obtain the bias and MSE of \( t_2 \), we write

\[
s_y^2 = S_y^2 (1 + e_0), \quad s_x^2 = S_x^2 (1 + e_2)
\]

Such that \( E(e_0) = E(e_2) = 0 \)
E(e^2) = \frac{1}{n}(\bar{\theta}_{004} - 1), \ E(e_0e_2) = \frac{1}{n}(\bar{\theta}_{202} - 1)

After simplification we get the bias and MSE of t_2 as

\[ B(t_2) \approx \frac{S_y^2}{n} \left[ \bar{\theta}_{004} + \frac{\bar{\theta}_{202}}{2} - \frac{5}{8} \right] \]  

(2.3)

\[ \text{MSE}(t_2) \approx \frac{S_y^2}{n} \left[ \bar{\theta}_{400} + \frac{\bar{\theta}_{004}}{4} + \bar{\theta}_{202} - \frac{9}{4} \right] \]  

(2.4)

3. Improved Estimator

Following Kadilar and Cingi (2006) and Singh et. al. (2007), we propose an improved estimator for estimating population variance S_y^2 as-

\[ t = s_y^2 \left[ \alpha \exp \left( \frac{S_x^2 - s_x^2}{S_x^2 + s_x^2} \right) + (1 - \alpha) \exp \left( \frac{s_x^2 - S_y^2}{s_x^2 + S_y^2} \right) \right] \]  

(3.1)

where \( \alpha \) is a real constant to be determined such that the MSE of t is minimum.

Expressing t in terms of e’s, we have

\[ t = S_y^2 \left[ 1 + e_0 \left( \frac{\bar{\theta}_{004}}{2} \right) + e_0 \right] \left[ \frac{\exp \left( \frac{e_i^2}{2} \right)}{1 + e_i^2} \right] + (1 - \alpha) \exp \left( \frac{e_i^2}{2} \right) \left[ \frac{\exp \left( \frac{e_i^2}{2} \right)}{1 + e_i^2} \right] \]  

(3.2)

Expanding the right hand side of (3.2) and retaining terms up to second power of e’s, we have

\[ t \approx S_y^2 \left[ 1 + e_0 + \frac{e_2^2}{2} + \frac{e_0e_2}{8} + \alpha \left( \frac{e_1^2}{2} + \frac{e_1^2}{8} \right) \right] - \alpha e_0 \left( \frac{e_2^2}{2} + \frac{e_2^2}{8} \right) \]  

(3.3)

Taking expectations of both sides of (3.3) and then subtracting S_y^2 from both sides, we get the bias of the estimator t, up to the first order of approximation, as

39
\[ B(t) = \frac{S_y^2}{n} \left[ \frac{\alpha}{8} \left( \bar{\varphi}_{004} - 1 \right) + \frac{(1-\alpha)}{8} \left( \bar{\varphi}_{004} - 1 \right) + \frac{(1-\alpha)}{2} \left( \bar{\varphi}_{202} - 1 \right) - \frac{\alpha}{2} \left( \bar{\varphi}_{220} - 1 \right) \right] \]  

(3.4)

From (3.4), we have

\[ (t - S_y^2) \equiv S_y^2 \left[ e_0 - \frac{\alpha e_1}{2} + \frac{(1-\alpha)}{2} e_2 \right] \]  

(3.5)

Squaring both the sides of (3.5) and then taking expectation, we get MSE of the estimator \( t \), up to the first order of approximation, as

\[ \text{MSE}(t) \equiv \frac{S_y^4}{n} \left[ \left( \bar{\varphi}_{004} - 1 \right) + \frac{\alpha^2}{4} \left( \bar{\varphi}_{004} - 1 \right) + \frac{(1-\alpha^2)}{4} \left( \bar{\varphi}_{004} - 1 \right) - \alpha \left( \bar{\varphi}_{220} - 1 \right) + (1-\alpha) \left( \bar{\varphi}_{202} - 1 \right) - \frac{\alpha(1-\alpha)}{2} \left( \bar{\varphi}_{022} - 1 \right) \right] \]  

(3.6)

Minimization of (3.6) with respect to \( \alpha \) yields its optimum value as

\[ \alpha = \frac{\left( \bar{\varphi}_{004} + 2(\bar{\varphi}_{220} + \bar{\varphi}_{202}) + \bar{\varphi}_{022} - 6 \right)}{\left( \bar{\varphi}_{004} + 2\bar{\varphi}_{004} + 2\bar{\varphi}_{022} - 4 \right)} = \alpha_0 \text{ (say)} \]  

(3.7)

Substitution of \( \alpha_0 \) from (3.7) into (3.6) gives minimum value of MSE of \( t \).

4. Proposed estimators in two-phase sampling

In certain practical situations when \( S^2_x \) is not known a priori, the technique of two-phase or double sampling is used. This scheme requires collection of information on \( x \) and \( z \) the first phase sample \( s' \) of size \( n' < N \) and on \( y \) for the second phase sample \( s \) of size \( n \) \( (n<n') \) from the first phase sample.

The estimators \( t_1, t_2 \) and \( t \) in two-phase sampling will take the following form, respectively

\[ t_{id} = s_y^2 \exp \left[ \frac{s^2_x - s^2_z}{s^2_x + s^2_z} \right] \]  

(4.1)
\[ t_{2d} = s_z^2 \exp \left[ \frac{s_x^2 - s_z^2}{s_x^2 + s_z^2} \right] \]  
\[ t_d = s_z^2 \left[ k \exp \left( \frac{s_x^2 - s_z^2}{s_x^2 + s_z^2} \right) + (1 - k) \exp \left( \frac{s_x^2 - s_z^2}{s_x^2 + s_z^2} \right) \right] \]  

To obtain the bias and MSE of \( t_{1d}, t_{2d}, t_d \), we write

\[ s_y^2 = S_y^2(1 + e_0), \quad s_x^2 = S_x^2(1 + e_1), \quad s_x^2 = S_x^2(1 + e_1') \]

\[ s_z^2 = S_z^2(1 + e_2), \quad s_z^2 = S_z^2(1 + e_2') \]

where \( s_x^2 = \frac{1}{(n-1)} \sum_{i=1}^{n'} (x_i - \bar{x}')^2, \quad s_z^2 = \frac{1}{(n'-1)} \sum_{i=1}^{n'} (z_i - \bar{z}')^2 \)

\[ \bar{x}' = \frac{1}{n'} \sum_{i=1}^{n'} x_i, \quad \bar{z}' = \frac{1}{n'} \sum_{i=1}^{n'} z_i \]

Also,

\[ \text{E}(e_1) = \text{E}(e_2) = 0, \]

\[ \text{E}(e_1^2) = \frac{1}{n'} (\partial_{o_0} - 1), \quad \text{E}(e_2^2) = \frac{1}{n'} (\partial_{o_0} - 1), \]

\[ \text{E}(e_1' e_2') = \frac{1}{n'} (\partial_{z_0} - 1) \]

Expressing \( t_{1d}, t_{2d}, \) and \( t_d \) in terms of \( e \)'s and following the procedure explained in section 2 and section 3 we get the MSE of these estimators, respectively as-

\[ \text{MSE}(t_{1d}) \approx S_y^4 \left[ \frac{1}{n} (\partial_{400} - 1) + \frac{1}{4} \left( \frac{1}{n} - \frac{1}{n'} \right) (\partial_{400} - 1) \right] \]

\[ + \left( \frac{1}{n} - \frac{1}{n'} \right) (\partial_{z_0} - 1) \]  
\[ \text{(4.4)} \]
Minimization of (4.6) with respect to $k$ yields its optimum value as

$$ k = \frac{\hat{\sigma}^{202} + 2(\hat{\sigma}^{220} - 1) + \hat{\sigma}^{022} - 6}{(\hat{\sigma}^{040} + \hat{\sigma}^{004} + 2\hat{\sigma}^{022} - 4)} = k_0 \text{ (say)} \quad (4.7) $$

Substitution of $k_0$ from (4.7) to (4.6) gives minimum value of MSE of $t_d$.

5. Empirical Study

To illustrate the performance of various estimators of $S_y^2$, we consider the data given in Murthy (1967, p.-226). The variates are:

$y$: output, $x$: number of workers, $z$: fixed capital,

$N=80$, $n'=25$, $n=10$.

$\hat{\sigma}^{400} = 2.2667$, $\hat{\sigma}^{040} = 3.65$, $\hat{\sigma}^{004} = 2.8664$, $\hat{\sigma}^{220} = 2.3377$, $\hat{\sigma}^{202} = 2.2208$, $\hat{\sigma}^{400} = 3.14$

The percent relative efficiency (PRE) of various estimators of $S_y^2$ with respect to conventional estimator $s_y^2$ has been computed and displayed in table 5.1.
Table 5.1: PRE of $s_y^2$, t1, t2 and min. MSE (t) with respect to $s_y^2$

<table>
<thead>
<tr>
<th>Estimator</th>
<th>PRE ($s_y^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_y^2$</td>
<td>100</td>
</tr>
<tr>
<td>t1</td>
<td>214.35</td>
</tr>
<tr>
<td>t2</td>
<td>42.90</td>
</tr>
<tr>
<td>t</td>
<td>215.47</td>
</tr>
</tbody>
</table>

In table 5.2 PRE of various estimators of $s_y^2$ in two-phase sampling with respect to $S_y^2$ are displayed.

Table 5.2: PRE of $s_y^2$, t1d, t2d and min.MSE (td) with respect to $s_y^2$

<table>
<thead>
<tr>
<th>Estimator</th>
<th>PRE ($s_y^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_y^2$</td>
<td>100</td>
</tr>
<tr>
<td>t1d</td>
<td>1470.76</td>
</tr>
<tr>
<td>t2d</td>
<td>513.86</td>
</tr>
<tr>
<td>td</td>
<td>1472.77</td>
</tr>
</tbody>
</table>

6. Conclusion

From table 5.1 and 5.2, we infer that the proposed estimators $t$ performs better than conventional estimator $s_y^2$ and other mentioned estimators.
References


