Some Improved Multivariate-Ratio-Type Estimators Using Geometric and Harmonic Means in Stratified Random Sampling

Sachin Malik and Rajesh Singh

Department of Statistics, Banaras Hindu University, Varanasi 221005, India

Correspondence should be addressed to Rajesh Singh, rsinghstat@gmail.com

Received 20 April 2012; Accepted 13 June 2012

Academic Editors: S. Lototsky, V. Makis, A. Pascucci, and O. Pons

Auxiliary variable is commonly used in survey sampling to improve the precision of estimates. Whenever there is auxiliary information available, we want to utilize it in the method of estimation to obtain the most efficient estimator. In this paper using multiauxiliary information we have proposed estimators based on geometric and harmonic mean. It was also shown that estimators based on harmonic mean and geometric mean are less biased than Olkin (1958) and Singh (1967) estimators under certain conditions. However, the MSE of Olkin (1958) estimator and geometric and harmonic estimators are same up to the first order of approximations.

1. Introduction

The problem of estimating the population mean in the presence of an auxiliary variable has been widely discussed in finite population sampling literature. Kadilar and Cingi [1], Singh et al. [2], Singh and Vishwakarma [3], Singh et al. [4], and Koyuncu and Kadilar [5] proposed estimators in stratified random sampling. Ghosh [6] and Rao [7] have suggested estimators in stratified random sampling with multiple characteristics.

Olkin [8] has considered the use of multiauxiliary variables positively correlated with the variable under study to build up a multivariate ratio estimator of population mean \( \bar{Y} \).

In this paper, we have considered the multiauxiliary variables. Olkin’s [8] estimator is based on the weighted arithmetic mean of \( r_i \bar{X}_i \)’s and is given as

\[
\bar{Y}_{ap} = \sum_{i=1}^{k} \alpha_i r_i \bar{X}_i
\] (1.1)
where (i) $\alpha_i$’s are weights such that $\sum_{i=1}^{k} \alpha_i = 1$, (ii) $\bar{X}_i$’s are the population means of the auxiliary variables and assumed to be known, and (iii) $r_i = \bar{y}_{st}/\bar{x}_{ist}$, $\bar{y}_{st}$ is the sample mean of the study variable $Y$ and $\bar{x}_{ist}$ are the sample means of the auxiliary variables $\bar{X}_i$ based on a stratified random sample of size $n$ drawn from a population of size $N$. Let the population of size, $N$, be stratified into $L$ strata with $h$th stratum containing $N_h$ units, where $h = 1, 2, 3, \ldots, L$ such that $\sum_{h=1}^{L} N_h = N$.

Following Olkin’s [8] estimator, several other estimators using multiauxiliary variables have been proposed in recent years. Singh [9] has extended Olkin’s [8] estimator to the case where auxiliary variables are negatively correlated with the variable under study. Srivastava [10] and Rao and Mudholkar [11] have given estimators, where some of the characters are positively and others are negatively correlated with the character under study. The main objective of presenting these estimators was to reduce the bias and mean square errors.

Motivated by Singh [9, 12] and Singh et al. [13], we propose an estimator in stratified sampling as

$$\bar{y}_s = \prod_{i=1}^{k} r_i \bar{X}_i.$$  \hspace{1cm} (1.2)

We also propose two alternative estimators based on geometric mean and harmonic mean, as

$$\bar{y}_{gp} = \prod_{i=1}^{k} \left( r_i \bar{X}_i \right)^{\alpha_i},$$  \hspace{1cm} (1.3)

$$\bar{y}_{hp} = \left( \sum_{i=1}^{k} \frac{\alpha_i}{r_i \bar{X}_i} \right)^{-1}$$  \hspace{1cm} (1.4)

such that $\sum_{i=1}^{k} \alpha_i = 1$.

These estimators are based on the assumptions that the auxiliary characters are positively correlated with $Y$. Let $\rho_{ij}$ ($i = 1, 2, \ldots, k; j = 1, 2, \ldots, k$) be the correlation coefficient between $X_i$ and $X_j$ and $\rho_{0i}$ the correlation coefficient between $Y$ and $X_i$.

2. BIAS and MSE of the Estimators

To obtain the bias and MSE’s of the estimators up to first order of approximation, we write

$$\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h = \bar{Y}(1 + e_0), \quad \bar{x}_{ist} = \sum_{h=1}^{L} W_h \bar{x}_{ih} = \bar{X}_i(1 + e_i),$$  \hspace{1cm} (2.1)

such that $E(e_i) = 0$, where,

$$\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h, \quad \bar{x}_{ist} = \sum_{h=1}^{L} W_h \bar{x}_{ih},$$

$$\bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi}, \quad \bar{Y}_h = \frac{1}{N_h} \sum_{i=1}^{n_h} y_{hi},$$
\[
Y = \bar{Y}_{st} = \sum_{h=1}^{L} W_h \bar{Y}_h,
\]
where, \(W_h = \frac{N_h}{N}\),

\[
E(e_0^2) = \frac{\sum_{h=1}^{L} W_h^2 \gamma_h S_{yh}^2}{\bar{Y}^2}, \quad E(e_1^2) = \frac{\sum_{h=1}^{L} W_h^2 \gamma_h S_{ixy}^2}{\bar{X}_i^2},
\]

\[
E(e_0 e_1) = \frac{\sum_{h=1}^{L} W_h^2 \gamma_h S_{ixy}^2}{\bar{X}_i^2}.
\]

(2.2)

Also,

\[
V(\bar{Y}_{st}) = \bar{Y}^2 E(e_0^2),
\]

\[
S_{yh}^2 = \frac{\sum_{i=1}^{N_h} (\bar{y}_h - \bar{Y}_h)^2}{N_h - 1},
\]

\[
S_{ixy}^2 = \frac{\sum_{i=1}^{N_h} (\bar{x}_ih - \bar{X}_ih)^2}{N_h - 1},
\]

(2.3)

\[
V(\bar{X}_i) = \bar{X}_i^2 E(e_0^2).
\]

In the same way \(C_{00}\) and \(C_{ij}\) are defined.

Further, let \(a' = (a_1, a_2, \ldots, a_k)\) and \(C = [C_{ij}]_{k \times k}, (i = 1, 2, \ldots, k; j = 1, 2, \ldots, k)\).

Using Taylor’s series expansion, under the usual assumptions, we obtain

\[
\bar{y}_{ap} = \sum_{i=1}^{k} a_i \bar{Y} (1 + e_0)(1 + e_i)^{-1}
\]

\[
= \bar{Y} \sum_{i=1}^{k} a_i \left[ (1 + e_0) \left( 1 - e_i + e_i^2 - e_i^3 \right) \right]
\]

\[
= \bar{Y} \sum_{i=1}^{k} a_i \left[ 1 + e_0 - e_i + e_i^2 - e_0 e_i + e_0 e_i^2 - e_i^3 + e_i^4 - e_0 e_i^2 \right].
\]

(2.4)

Subtracting \(\bar{Y}\) from both sides of (2.4) and then taking expectation of both sides, we get the bias of the estimator \(\bar{y}_{ap}\) up to the first order of approximation as

\[
B(\bar{y}_{ap}) = \bar{Y} \left[ \sum_{i=1}^{k} a_i C_i^2 - \sum_{i=1}^{k} a_i C_{0i} \right].
\]

(2.5)

Subtracting \(\bar{Y}\) from both sides of (2.4), taking square, and then taking expectation of both sides, we get the MSE of the estimator \(\bar{y}_{ap}\) up to the first order of approximation as

\[
\text{MSE}(\bar{y}_{ap}) = \bar{Y}^2 \left[ C_0^2 + \sum_{i=1}^{k} a_i^2 C_i^2 - 2 \sum_{i=1}^{k} a_i C_{0i} C_i + 2 \sum \sum a_i a_j C_{ij} \right].
\]

(2.6)
In the same way using the Taylor series expansion under the usual assumptions, we obtain

\[
\bar{y}_{gp} = \bar{Y} \prod_{i=1}^{k} \left[ 1 + e_0 - \alpha_i(e_i + e_0e_i) + \frac{\alpha_i(1 + \alpha_i)}{2} (e_i^2 + e_0e_i^2) - \frac{\alpha_i(1 + \alpha_i)(2 + \alpha_i)}{6} (e_i^3 + e_0e_i^3) + \cdots \right],
\]

\[
\bar{y}_{hp} = \bar{Y} \left[ 1 + e_0 - \sum_{i=1}^{k} \alpha_i e_i - \sum_{i=1}^{k} \alpha_i e_i e_i + \left( \sum_{i=1}^{k} \alpha_i e_i \right)^2 + \left( \sum_{i=1}^{k} \alpha_i e_i \right) e_0 - \left( \sum_{i=1}^{k} \alpha_i e_i \right)^3 - \left( \sum_{i=1}^{k} \alpha_i e_i \right) e_0 + \cdots \right].
\]

To calculate the bias and mean square error, we considered the terms having powers up to second degree only as the calculations become more complicated when the higher-order terms are included.

So, from (2.7), the bias and mean square error of the estimates up to 0(1/n) are obtained as

\[
B(\bar{y}_{gp}) = \bar{Y} \left[ \sum_{i} \frac{\alpha_i(\alpha_i + 1)C_i^2}{2} + \sum \alpha_i \alpha_j C_{ij} - \sum \alpha_i C_{0i} \right],
\]

\[
\text{MSE}(\bar{y}_{gp}) = \bar{Y}^2 \left[ C_0^2 + \sum_{i=1}^{k} p_i^2 C_i^2 - 2 \sum_{i=1}^{k} p_i C_0 C_i + 2 \sum \sum \alpha_i \alpha_j C_{ij} \right],
\]

\[
B(\bar{y}_{hp}) = \bar{Y} \left[ \left( \sum_{i=1}^{k} \alpha_i C_i \right)^2 - \sum \alpha_i C_{0i} \right],
\]

\[
\text{MSE}(\bar{y}_{hp}) = \bar{Y}^2 \left[ C_0^2 + \sum_{i=1}^{k} \alpha_i^2 C_i^2 - 2 \sum \alpha_i C_{0i} + 2 \sum \alpha_i \alpha_j C_{ij} \right].
\]

We see that MSE’s of these estimators are same and the biases are different. In general

\[
\text{MSE}(\bar{y}_{ap}) = \text{MSE}(\bar{y}_{gp}) = \text{MSE}(\bar{y}_{hp}).
\]

We know that in case of univariate the usual ratio-type \( \bar{y}_R \) estimator for the \( i \)th auxiliary variable is superior to the mean per unit estimator \( \bar{y} \), when

\[
\frac{C_0}{C_i \rho_{0i}} > \frac{1}{2}.
\]

Comparing the variance of \( \bar{y} = C_0^2 \bar{Y}_a^2 \) with the mean square error of all the three estimators, we note that the ratio estimators given in (1.1), (1.2), and (1.3) are more efficient than \( \bar{y} \).
3. Comparison of Biases

The biases may be either positive or negative. So, for comparison, we have compared the absolute biases of the estimates when these are more efficient than the sample mean. The bias of the estimator of geometric mean is smaller than that of arithmetic mean:

\[ |B(\bar{y}_{ap})| > |B(\bar{y}_{gp})|. \] (3.1)

Squaring and simplifying (3.1), we observe that

\[ \left[ \frac{1}{2} \sum_{i=1}^{k} \alpha_i^2 C_i^2 - 2 \sum_{i=1}^{k} \alpha_i C_{ij} + 2 \sum_{i=1}^{k} \alpha_i C_i^2 + \frac{3}{2} \sum_{i=1}^{k} \alpha_i C_i^2 \right] \times \left[ \frac{1}{2} \sum_{i=1}^{k} \alpha_i C_i^2 - 1/2 \sum_{i=1}^{k} \alpha_i C_i^2 - \sum \alpha_i C_{ij} \right] > 0. \] (3.2)

Thus the above inequality is true when both factors are either positive or negative. The first factor of (3.2)

\[ \left[ \frac{1}{2} \sum_{i=1}^{k} \alpha_i^2 C_i^2 - 2 \sum_{i=1}^{k} \alpha_i C_{ij} + 2 \sum_{i=1}^{k} \alpha_i C_i^2 + \frac{3}{2} \sum_{i=1}^{k} \alpha_i C_i^2 \right] \] (3.3)

is positive, when

\[ \frac{\sum_{i=1}^{k} \alpha_i^2 C_i^2}{\alpha C_g} > \frac{1}{3}. \] (3.4)

In the same way, it can be shown that the second factor of (3.2) is also positive when

\[ \frac{\sum_{i=1}^{k} \alpha_i^2 C_i^2}{\alpha C_g} > 1. \] (3.5)

When both factors are of (3.2) is negative, the sign of inequalities of (3.4) and (3.5) reversed.

Also comparing the square of the biases of geometric and harmonic estimator, we find that geometric estimator is more biased than harmonic estimator.

Hence we may conclude that under the situations where arithmetic, geometric and harmonic estimator are more efficient than sample mean and the relation (3.5) or

\[ \frac{\sum_{i=1}^{k} \alpha_i^2 C_i^2}{\alpha C_g} < \frac{1}{3} \] (3.6)

is satisfied, the biases of the estimates satisfy the relation

\[ |B(\bar{y}_{ap})| > |B(\bar{y}_{gp})| > |B(\bar{y}_{hp})|. \] (3.7)
In this section, we use the data set earlier used in Koyuncu and Kadilar [4].

**4. Empirical Study**

In this section, we use the data set earlier used in Koyuncu and Kadilar [5].

To illustrate the efficiency of suggested estimators, we consider the data concerning the number of teachers as the study variable \((y)\), number of students \((x)\), and number of classes \((z)\) in both primary and secondary schools as auxiliary variables for 923 districts at 6 regions (as 1: Marmara, 2: Agean, 3: Mediterranean, 4: Central Anatolia, 5: Black Sea, and 6: East and Southeast Anatolia) in Turkey in 2007 (source: The Turkish Republic Ministry of Education). The summary statistics of the data are given in Table 1. We used the Neyman allocation for allocating the samples to different strata [14].
Table 2: Bias and MSE of different estimators.

<table>
<thead>
<tr>
<th>Estimators</th>
<th>Auxiliary variables used</th>
<th>Bias</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{y}_{st} )</td>
<td>none</td>
<td>0</td>
<td>2234.649</td>
</tr>
<tr>
<td>( \bar{y}_{st}(\bar{X}<em>1/\bar{X}</em>{1st}) )</td>
<td>( \bar{X}_1 )</td>
<td>0.538</td>
<td>216.786</td>
</tr>
<tr>
<td>( \bar{y}_{st}(\bar{X}<em>2/\bar{X}</em>{2st}) )</td>
<td>( \bar{X}_2 )</td>
<td>0.642</td>
<td>334.443</td>
</tr>
<tr>
<td>( \bar{y}_{ap} )</td>
<td>( X_1 ) and ( X_2 )</td>
<td>0.610</td>
<td>170.927</td>
</tr>
<tr>
<td>( \bar{y}_{gp} )</td>
<td>( X_1 ) and ( X_2 )</td>
<td>0.573</td>
<td>170.927</td>
</tr>
<tr>
<td>( \bar{y}_{hp} )</td>
<td>( X_1 ) and ( X_2 )</td>
<td>0.115</td>
<td>170.927</td>
</tr>
<tr>
<td>( \bar{y}<em>a = \bar{y}</em>{st}(\bar{X}<em>1/\bar{X}</em>{1st})(\bar{X}<em>2/\bar{X}</em>{2st}) )</td>
<td>( X_1 ) and ( X_2 )</td>
<td>3.835</td>
<td>1759.889</td>
</tr>
</tbody>
</table>

5. Conclusion

From Table 2, we observe that the ratio estimator based on harmonic mean is less biased. However, the mean square errors of the estimators \( \bar{y}_{hp} \), \( \bar{y}_{ap} \), and \( \bar{y}_{gp} \) are same. Hence for this data set, we conclude that when more than one auxiliary variables are used for estimating the population parameters, it is better to use harmonic mean as an estimator in case of stratified random sampling.

References
