International Scholarly Research Network
ISRN Probability and Statistics
Volume 2012, Article ID 509186, 7 pages
doi:10.5402/2012/509186

Research Article

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 $\overline{Y}$.

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

$$\overline{y}_{ap} = \sum_{i=1}^{k} \alpha_i r_i \overline{X}_i, \tag{1.1}$$

where (i) α_i's are weights such that $\sum_{i=1}^{k} \alpha_i = 1$, (ii) $\overline{X}_i$'s are the population means of the auxiliary variables and assumed to be known, and (iii) $r_i = \overline{y}_{st}/\overline{x}_{ist}$, $\overline{y}_{st}$ is the sample mean of the study variable Y and $\overline{x}_{ist}$ are the sample means of the auxiliary variables $\overline{X}_i$ 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 hth 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

$$\overline{y}_s = \prod_{i=1}^{k} r_i \overline{X}_i. \tag{1.2}$$

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

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

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

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 ρ_{ij} $(i = 1, 2, \ldots, k; j = 1, 2, \ldots, k)$ be the correlation coefficient between X_i and X_j and ρ_{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

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

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

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

$$\overline{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi}, \qquad \overline{Y}_h = \frac{1}{N_h} \sum_{i=1}^{n_h} Y_{hi},$$

$$Y = \overline{Y}_{st} = \sum_{h=1}^{L} W_h \overline{Y}_h, \quad \text{where, } W_h = \frac{N_h}{N},$$
$$E\left(e_0^2\right) = \frac{\sum_{h=1}^{L} W_h^2 \gamma_h S_{yh}^2}{\overline{Y}^2}, \qquad E\left(e_i^2\right) = \frac{\sum_{h=1}^{L} W_h^2 \gamma_h S_{ixh}^2}{\overline{X}_i^2},$$
$$E(e_0 e_i) = \frac{\sum_{h=1}^{L} W_h^2 \gamma_h S_{ixyh}}{\overline{X}_i^2}. \tag{2.2}$$

Also,

$$V\left(\overline{Y}_{st}\right) = \overline{Y}^2 E\left(e_0^2\right),$$
$$S_{yh}^2 = \frac{\sum_{i=1}^{N_h} \left(\overline{y}_h - \overline{Y}_h\right)^2}{N_h - 1}, \tag{2.3}$$
$$S_{ixh}^2 = \frac{\sum_{i=1}^{N_h} \left(\overline{x}_{ih} - \overline{X}_{ih}\right)^2}{N_h - 1}, \qquad S_{ixyh} = \frac{\sum_{i=1}^{N_h} \left(\overline{y}_h - \overline{Y}_h\right)\left(\overline{x}_{ih} - \overline{X}_{ih}\right)}{N_h - 1}.$$

In the same way C_{0i} and C_{ij} are defined.

Further, let $\underset{\sim}{\alpha}' = (\alpha_1, \alpha_2, \ldots, \alpha_k)$ and $C = \lfloor C_{ij} \rfloor_{k\times k}$, $(i = 1, 2, \ldots, k; j = 1, 2, \ldots, k)$.

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

$$\begin{aligned}
\overline{y}_{ap} &= \sum_{i=1}^{k} \alpha_i \overline{Y}(1 + e_0)(1 + e_i)^{-1} \\
&= \overline{Y} \sum_{i=1}^{k} \alpha_i \left[(1 + e_0)\left(1 - e_i + e_i^2 - e_i^3\right)\right] \\
&= \overline{Y} \sum_{i=1}^{k} \alpha_i \left[1 + e_0 - e_i + e_i^2 - e_0 e_1 + e_0 e_i^2 - e_i^3 + e_i^4 - e_0 e_i^3\right].
\end{aligned} \tag{2.4}$$

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

$$B\left(\overline{y}_{ap}\right) = \overline{Y}\left[\sum_{i=1}^{k} \alpha_i C_i^2 - \sum_{i=1}^{k} \alpha_i C_{0i}\right]. \tag{2.5}$$

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

$$\mathrm{MSE}\left(\overline{y}_{ap}\right) = \overline{Y}^2\left[C_0^2 + \sum_{i=1}^{k} \alpha_i^2 C_i^2 - 2\sum_{i=1}^{k} \alpha_i C_0 C_i + 2\sum\sum \alpha_i \alpha_j C_{ij}\right]. \tag{2.6}$$

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

$$\begin{aligned}
\overline{y}_{gp} &= \overline{Y}\prod_{i=1}^{k}\left[1+e_0-\alpha_i(e_i+e_0e_i)+\frac{\alpha_i(1+\alpha_i)}{2}\left(e_i^2+e_0e_i^2\right)\right.\\
&\qquad\left.-\frac{\alpha_i(1+\alpha_i)(2+\alpha_i)}{6}\left(e_i^3+e_0e_i^3\right)+\cdots\right],\\
\overline{y}_{hp} &= \overline{Y}\left[1+e_0-\sum_{i=1}^{k}\alpha_ie_i-\sum_{i=1}^{k}\alpha_ie_0e_i+\left(\sum_{i=1}^{k}\alpha_ie_i\right)^2\right.\\
&\qquad\left.+\left(\sum_{i=1}^{k}\alpha_ie_i\right)^2e_0-\left(\sum_{i=1}^{k}\alpha_ie_i\right)^3-\left(\sum_{i=1}^{k}\alpha_ie_i\right)^3e_0+\cdots\right].
\end{aligned}\tag{2.7}$$

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

$$\begin{aligned}
B\left(\overline{y}_{gp}\right) &= \overline{Y}\left[\sum_i\frac{\alpha_i(\alpha_i+1)C_i^2}{2}+\sum\sum\alpha_i\alpha_jC_{ij}-\sum_i\alpha_iC_{0i}\right],\\
\mathrm{MSE}\left(\overline{y}_{gp}\right) &= \overline{Y}^2\left[C_0^2+\sum_{i=1}^{k}p_i^2C_i^2-2\sum_{i=1}^{k}p_iC_0C_i+2\sum\sum p_ip_jC_{ij}\right],\\
B\left(\overline{y}_{hp}\right) &= \overline{Y}\left[\left(\sum_{i=1}^{k}\alpha_iC_i\right)^2-\sum_{i=1}^{k}\alpha_iC_{0i}\right],\\
\mathrm{MSE}\left(\overline{y}_{hp}\right) &= \overline{Y}^2\left[C_0^2+\sum_{i=1}^{k}\alpha_i^2C_i^2-2\sum_{i=1}^{k}\alpha_iC_{0i}+2\sum\sum\alpha_i\alpha_jC_{ij}\right].
\end{aligned}\tag{2.8}$$

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

$$\mathrm{MSE}\left(\overline{y}_{ap}\right)=\mathrm{MSE}\left(\overline{y}_{gp}\right)=\mathrm{MSE}\left(\overline{y}_{hp}\right).\tag{2.9}$$

We know that in case of univariate the usual ratio-type $\overline{y}_R$ estimator for the ith auxiliary variable is superior to the mean per unit estimator $\overline{y}$, when

$$\frac{C_0}{C_i}\rho_{0i}>\frac{1}{2}.\tag{2.10}$$

Comparing the variance of $\overline{y}=C_0^2\overline{Y}_{st}^2$ 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 $\overline{y}$.

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:

$$\left|B\left(\overline{y}_{ap}\right)\right| > \left|B\left(\overline{y}_{gp}\right)\right|. \tag{3.1}$$

Squaring and simplifying (3.1), we observe that

$$\begin{aligned}&\left[\frac{1}{2}\sum_{i=1}^{k}\alpha_i^2C_i^2 - 2\sum_{i=1}^{k}\alpha_iC_{0i} + 2\sum\sum\alpha_iw_jC_{ij} + \frac{3}{2}\sum_{i=1}^{k}\alpha_iC_i^2\right]\\&\quad\times\left[\frac{1}{2}\sum_{1=1}^{k}\alpha_iC_i^2 - \frac{1}{2}\sum_{i=1}^{k}\alpha_i^2C_i^2 - \sum\sum\alpha_i\alpha_jC_{ij}\right] > 0.\end{aligned} \tag{3.2}$$

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

$$\left[\frac{1}{2}\sum_{i=1}^{k}\alpha_i^2C_i^2 - 2\sum_{i=1}^{k}\alpha_iC_{0i} + 2\sum\sum\alpha_i\alpha_jC_{ij} + \frac{3}{2}\sum_{i=1}^{k}\alpha_iC_i^2\right] \tag{3.3}$$

is positive, when

$$\frac{\sum_{i=1}^{k}\alpha_i^2C_i^2}{\underset{\sim}{\alpha}'C\underset{\sim}{\alpha}} > \frac{1}{3}. \tag{3.4}$$

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

$$\frac{\sum_{i=1}^{k}\alpha_i^2C_i^2}{\underset{\sim}{\alpha}'C\underset{\sim}{\alpha}} > 1. \tag{3.5}$$

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

$$\frac{\sum_{i=1}^{k}\alpha_i^2C_i^2}{\underset{\sim}{\alpha}'C\underset{\sim}{\alpha}} < \frac{1}{3} \tag{3.6}$$

is satisfied, the biases of the estimates satisfy the relation

$$\left|B\left(\overline{y}_{ap}\right)\right| > \left|B\left(\overline{y}_{gp}\right)\right| > \left|B\left(\overline{y}_{hp}\right)\right|. \tag{3.7}$$

Table 1: Data statistics.

$N_1 = 127$	$N_2 = 117$	$N_3 = 103$
$N_4 = 170$	$N_5 = 205$	$N_6 = 201$
$n_1 = 31$	$n_2 = 21$	$n_3 = 29$
$n_4 = 38$	$n_5 = 22$	$n_6 = 39$
$Sy1 = 883.835$	$Sy2 = 644.922$	$Sy3 = 1033.467$
$Sy4 = 810.585$	$Sy5 = 403.654$	$Sy6 = 711.723$
$\overline{Y}_1 = 703.74$	$\overline{Y}_2 = 413$	$\overline{Y}_3 = 573.17$
$\overline{Y}_4 = 424.66$	$\overline{Y}_5 = 267.03$	$\overline{Y}_6 = 393.84$
$Sx1 = 30486.751$	$Sx2 = 15180.769$	$Sx3 = 27549.697$
$Sx4 = 18218.931$	$Sx5 = 8497.776$	$Sx6 = 23094.141$
$\overline{X}_1 = 20804.59$	$\overline{X}_2 = 9211.79$	$\overline{X}_3 = 14309.30$
$\overline{X}_4 = 9478.85$	$\overline{X}_5 = 5569.95$	$\overline{X}_6 = 12997.59$
$Sxy1 = 25237153.52$	$Sxy2 = 9747942.85$	$Sxy3 = 28294397.04$
$Sxy4 = 14523885.53$	$Sxy5 = 3393591.75$	$Sxy6 = 15864573.97$
$\rho_{yx1} = 0.936$	$\rho_{yx2} = 0.996$	$\rho_{yx3} = 0.994$
$\rho_{yx4} = 0.983$	$\rho_{yx5} = 0.989$	$\rho_{yx6} = 0.965$
$w1 = 0.138$	$w2 = 0.127$	$w3 = 0.112$
$w4 = 0.184$	$w5 = 0.222$	$w6 = 0.218$
$Sz1 = 555.5816$	$Sz2 = 365.4576$	$Sz3 = 612.9509281$
$Sz4 = 458.0282$	$Sz5 = 260.8511$	$Sz6 = 397.0481$
$\overline{Z}_1 = 498.28$	$\overline{Z}_2 = 318.33$	$\overline{Z}_3 = 431.36$
$\overline{Z}_4 = 311.32$	$\overline{Z}_5 = 227.22$	$\overline{Z}_6 = 313.71$
$Syz1 = 480688.2$	$Syz2 = 230092.8$	$Syz3 = 623019.3$
$Syz4 = 364943.4$	$Syz5 = 101539.1$	$Syz6 = 277696.1$
$Sxz1 = 15914648$	$Sxz2 = 5379190$	$Sxz3 = 16490674.56$
$Sxz4 = 8041254$	$Sxz5 = 2144057$	$Sxz6 = 8857729$

Biases and MSEs of different estimators under comparison based on the above data are given in Table 2.

Usually the weights are so chosen so as to minimize the MSE of an estimator subject to the condition

$$\sum_{i=1}^{k} \alpha_i = 1. \tag{3.8}$$

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.

Estimators	Auxiliary variables used	Bias	MSE
$\overline{y}_{st}$	none	0	2234.649
$\overline{y}_{st}(\overline{X}_1/\overline{x}_{1st})$	$\overline{X}_1$	0.538	216.786
$\overline{y}_{st}(\overline{X}_2/\overline{x}_{2st})$	$\overline{X}_2$	0.642	334.443
$\overline{y}_{ap}$	X_1 and X_2	0.610	170.927
$\overline{y}_{gp}$	X_1 and X_2	0.573	170.927
$\overline{y}_{hp}$	X_1 and X_2	0.115	170.927
$\overline{y}_s = \overline{y}_{st}(\overline{X}_1/\overline{x}_{1st})(\overline{X}_2/\overline{x}_{2st})$	X_1 and X_2	3.835	1759.889

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 $\overline{y}_{hp}, \overline{y}_{ap}$, and $\overline{y}_{gp}$ 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

[1] C. Kadilar and H. Cingi, "Ratio estimators in stratified random sampling," *Biometrical Journal*, vol. 45, no. 2, pp. 218–225, 2003.

[2] H. P. Singh, R. Tailor, S. Singh, and J.-M. Kim, "A modified estimator of population mean using power transformation," *Statistical Papers*, vol. 49, no. 1, pp. 37–58, 2008.

[3] H. P. Singh and G. K. Vishwakarma, "A family of estimators of population mean using auxiliary information in stratified sampling," *Communications in Statistics Theory and Methods*, vol. 37, no. 7, pp. 1038–1050, 2008.

[4] R. Singh, M. Kumar, M. K. Chaudhary, and C. Kadilar, "Improved exponential estimator in stratified random sampling," *Pakistan Journal of Statistics and Operation Research*, vol. 5, no. 2, pp. 67–82, 2009.

[5] N. Koyuncu and C. Kadilar, "Family of estimators of population mean using two auxiliary variables in stratified random sampling," *Communications in Statistics Theory and Methods*, vol. 38, no. 13–15, pp. 2398–2417, 2009.

[6] S. P. Ghosh, "A note on stratified random sampling with multiple characters," *Calcutta Statistical Association Bulletin*, vol. 8, no. 30-31, pp. 81–90, 1958.

[7] T. J. Rao, "Optimum allocation of sample size and prior distributions: a review," *International Statistical Review*, vol. 45, no. 2, pp. 173–179, 1977.

[8] I. Olkin, "Multivariate ratio estimation for finite populations," *Biometrika*, vol. 45, pp. 154–165, 1958.

[9] M. P. Singh, "Multivariate product method of estimation for finite populations," *Journal of the Indian Society of Agricultural Statistics*, vol. 19, pp. 1–10, 1967.

[10] S. K. Srivastava, "An estimate of the mean of a finite population using several auxiliary variables," *Journal of the Indian Statistical Association*, vol. 3, pp. 189–194, 1965.

[11] P. S. R. S. Rao and G. S. Mudholkar, "Generalized multivariate estimator for the mean of finite populations," *Journal of the American Statistical Association*, vol. 62, pp. 1009–1012, 1967.

[12] M. P. Singh, "On the estimation of ratio and product of the population parameters," *Sankhya*, vol. 27, pp. 321–328, 1965.

[13] R. Singh, P. Chauhan, N. Sawan, and F. Smarandache, *Auxiliary Information and a Priori Values in Construction of Improved Estimators*, Renaissance High Press, Ann Arbor, Mich, USA, 2007.

[14] W. G. Cochran, *Sampling Techniques*, John Wiley & Sons, New York, NY, USA, 1977.