Modified Ratio Estimators in Stratified Random Sampling

Prayad Sangngam[1*], Sasiprapa Hiriote[2]

Received: 18 February 2013 Accepted: 15 June 2013

Abstract

This paper considers two modified ratio estimators of population mean in stratified random sampling. The approximated mean squared error and bias of the proposed estimators are derived and theoretically compared with those of the existing estimators. The results show that the modified estimators produce smaller mean squared error and bias than the existing estimators in some conditions. Moreover, the theoretical result is confirmed by using a census data set.

Keywords: Ratio estimator, Mean squared error, Stratified random sampling.

Introduction

The problem of improving an unbiased estimator by using ratio estimators has received considerable attention in sampling theory. When an auxiliary variable (X), available for all units in the population, is highly correlated with a study variable, a ratio estimator can be used to improve the unbiased estimator. The efficiency of a ratio estimator depends on the coefficients of variation of auxiliary variable (C_x) and coefficients of variation of study variable Y (C_y). Murthy[1] has suggested that if $\rho > \frac{C_x}{2C_y}$, the ratio estimator performs better than the unbiased estimator under simple random sampling where ρ is the correlation coefficient between X and Y. When the C_x is known, Sissodia and Dwivedi[2] has proposed a modified ratio estimator for the population mean $(\bar{Y})$ as

$$\bar{y}_{SD} = \bar{y}\frac{\bar{X}+C_x}{\bar{x}+C_x},$$

where $\bar{X}$ is the population mean of auxiliary variable. In addition, there are several authors, such as Upadhyaya and Singh[3], Singh and Tailor[4], who have developed various ratio estimators under simple random sampling. When the population is heterogeneous and can be divided into homogenous subpopulations, it is advantageous to draw a sample by stratified random sampling. An unbiased estimator under stratified random sampling is given by

$$\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h , \quad (1)$$

where L is the number of stratum, $W_h = \frac{N_h}{N}$ is stratum weight, N is the population size, N_h is the number of units in stratum h, and $\bar{y}_h$ is the sample mean of the study variable in stratum h. The variance of the unbiased estimator is

$$V(\bar{y}_{st}) = \sum_{h=1}^{L} W_h^2 \gamma_h S_{yh}^2 , \quad (2)$$

where $\gamma_h = (1-f_h)/n_h$, $f_h = \frac{n_h}{N_h}$ is sampling fraction in stratum h, n_h is sample size in stratum h and S_{yh}^2 is the variance of the study variable in stratum h. According to[5], there are two types of ratio estimators in stratified random sampling, namely combined and separate ratio estimators. The combined ratio estimator is given by

$$\bar{y}_{RC} = \frac{\bar{y}_{st}}{\bar{x}_{st}}\bar{X} , \quad (3)$$

where $\bar{x}_{st} = \sum_{h=1}^{L} W_h \bar{x}_h$ is an unbiased estimator of $\bar{X}$ and $\bar{x}_h$ is the sample mean of auxiliary variable in stratum h. An approximated mean squared error (MSE) of the combined ratio estimator is

$$MSE(\bar{y}_{RC}) \approx \sum_{h=1}^{L} W_h^2 \gamma_h \left(S_{yh}^2 + R^2 S_{xh}^2 - 2RS_{xyh}\right), \quad (4)$$

[1] *Department of Statistics, Faculty of Science, Silpakorn University.*

[2] *Department of Statistics, Faculty of Science, Silpakorn University. E-mail: ssiprapa@su.ac.th*

* *Author for correspondence; E-mail: prayad@su.ac.th*

where $R = \frac{\bar{Y}}{\bar{X}}$ is the population ratio, S_{xh}^2 is the variance of auxiliary variable in stratum h and S_{xyh} is the covariance variance between auxiliary and study variables in stratum h. An approximated bias of the combined ratio estimator is

$$B(\bar{y}_{RC}) \approx \sum_{h=1}^{L} W_h^2 \gamma_h \left(\frac{R}{\bar{X}} S_{xh}^2 - \frac{1}{\bar{X}} S_{xyh} \right). \quad (5)$$

The separate ratio estimator is given by

$$\bar{y}_{RS} = \sum_{h=1}^{L} W_h \frac{\bar{y}_h \bar{X}_h}{\bar{x}_h}. \quad (6)$$

An approximated MSE of the separate ratio estimator can be given as,

$$MSE(\bar{y}_{RS}) \approx \sum_{h=1}^{L} W_h^2 \gamma_h \left(S_{yh}^2 + R_h^2 S_{xh}^2 - 2R_h S_{xyh} \right) \quad (7)$$

where $R_h = \frac{\bar{Y}_h}{\bar{X}_h}$ is the population ratio in stratum h. We can derive an approximated bias of the separate ratio estimator as,

$$E(\bar{y}_{RS} - \bar{Y}) = E\left[\sum_{h=1}^{L} W_h \frac{\bar{y}_h \bar{X}_h}{\bar{x}_h} \right] - \bar{Y}$$
$$= \sum_{h=1}^{L} W_h \left[E\left(\frac{\bar{y}_h \bar{X}_h}{\bar{x}_h} \right) - \bar{Y}_h \right].$$

Applying the bias of $\bar{y}_{Rh} = \frac{\bar{y}_h \bar{X}_h}{\bar{x}_h}$ for estimating $\bar{Y}_h$ under simple random sampling to draw in stratum h, we get

$$B(\bar{y}_{RS}) \approx \sum_{h=1}^{L} W_h \gamma_h \left(\frac{R_h}{\bar{X}_h} S_{xh}^2 - \frac{1}{\bar{X}_h} S_{xyh} \right). \quad (8)$$

For combined ratio estimation in stratified sampling, Kadilar and Cingi[6] have proposed several modified ratio estimators. The simplest estimator based on[2] is defined as

$$\bar{y}_{RC\_KC} = \bar{y}_{st} \frac{\sum_{h=1}^{L} W_h \left(\bar{X}_h + C_{xh} \right)}{\sum_{h=1}^{L} W_h \left(\bar{x}_h + C_{xh} \right)}, \quad (9)$$

where C_{xh} is coefficient of variation of auxiliary variable in stratum h. The MSE and bias of this estimator are approximated as follow:

$$MSE(\bar{y}_{RC\_KC}) \approx \sum_{h=1}^{L} W_h^2 \gamma_h \left(S_{yh}^2 + R_{KC}^2 S_{xh}^2 - 2R_{KC} S_{xyh} \right), \quad (10)$$

$$B(\bar{y}_{RC\_KC}) \approx \sum_{h=1}^{L} W_h^2 \gamma_h \left(\frac{R_{KC}}{\bar{X}_{KC}} S_{xh}^2 - \frac{1}{\bar{X}_{KC}} S_{xyh} \right), \quad (11)$$

where $R_{KC} = \frac{\bar{Y}_{st}}{\bar{X}_{KC}} = \frac{\bar{Y}_{st}}{\sum_{h=1}^{L} W_h \left(\bar{X}_h + C_{xh} \right)}$.

Kadilar and Cingi[7] have improved the combined ratio estimator in stratified random sampling based on[8] estimator. However, this estimator depends on such several unknown parameters that it is very difficult for application. Therefore, in the next section, a new combined ratio estimator in stratified random sampling based on[2] will be proposed. We also develop a new modified separate ratio estimator for stratified random sampling. The approximated MSE and bias of the two modified estimators will be derived. In Section 3, the comparison of efficiency between the modified estimators and the existing estimators will theoretically be provided. A numerical example will be used to confirm the result in Section 4.

Modified Ratio Estimators

In stratified random sampling, when the coefficient of variation C_x is known, a combined ratio estimator can be modified based on[2] as follows:

$$\bar{y}_{RC\_SD} = \frac{\bar{y}_{st}}{\bar{x}_{st} + C_x} \left(\bar{X} + C_x \right). \quad (12)$$

To obtain the MSE and bias of this estimator, let $e_1 = \frac{\bar{y}_{st} - \bar{Y}}{\bar{Y}}$ and $e_2 = \frac{\bar{x}_{st} - \bar{X}}{\bar{X} + C_x}$. It may be noted that

$E(e_1) = 0$, $E(e_2) = 0$, $E(e_1^2) = \frac{V(\bar{y}_{st})}{\bar{Y}^2}$,

$E(e_2^2) = \frac{V(\bar{x}_{st})}{\left(\bar{X} + C_x \right)^2}$ and $E(e_1 e_2) = \frac{Cov(\bar{x}_{st}, \bar{y}_{st})}{\bar{Y}\left(\bar{X} + C_x \right)}$.

The estimator $\bar{y}_{RC\_SD}$ can be written as

$$\bar{y}_{RC\_SD} = \bar{Y}(1 + e_1)(1 + e_2)^{-1}$$

Using Taylor series, we obtain

$$\bar{y}_{RC\_SD} = \bar{Y}\left(1+e_1\right)\left(1-e_2+e_2^2-...\right)$$
$$= \bar{Y}\left(1+e_1-e_2+e_2^2-e_1e_2+...\right).$$

When the terms of degree greater than two are ignored, we get

$$B\left(\bar{y}_{RC\_SD}\right) = E\left(\bar{y}_{RC\_SD}-\bar{Y}\right)$$
$$\approx \bar{Y}E\left(e_2^2-e_1e_2\right)$$
$$= \bar{Y}\left[\frac{V\left(\bar{x}_{st}\right)}{\left(\bar{X}+C_x\right)^2}-\frac{Cov\left(\bar{x}_{st},\bar{y}_{st}\right)}{\bar{Y}\left(\bar{X}+C_x\right)}\right].$$

By substituting $Cov\left(\bar{x}_{st},\bar{y}_{st}\right)=\sum_{h=1}^{L}W_h^2\gamma_hS_{xyh}$ and $V\left(\bar{x}_{st}\right)=\sum_{h=1}^{L}W_h^2\gamma_hS_{xh}^2$, an approximated bias of the modified combined ratio estimator is

$$B\left(\bar{y}_{RC\_SD}\right)\approx\sum_{h=1}^{L}W_h^2\gamma_h\left(\frac{R_x}{\bar{X}+C_x}S_{xh}^2-\frac{S_{xyh}}{\bar{X}+C_x}\right), \quad (13)$$

where $R_x=\dfrac{\bar{Y}}{\bar{X}+C_x}$ When the terms of degree greater than two are ignored, an approximated MSE of the estimator is equal to

$$MSE\left(\bar{y}_{RC\_SD}\right)=E\left(\bar{y}_{RC\_SD}-\bar{Y}\right)^2$$
$$\approx\bar{Y}^2E\left(e_1^2+e_2^2-e_1e_2\right)$$
$$MSE\left(\bar{y}_{RC\_SD}\right)\approx\sum_{h=1}^{L}W_h^2\gamma_h\left(S_{yh}^2+R_x^2S_{xh}^2-2R_xS_{xyh}\right). \quad (14)$$

For sample estimate of the $MSE\left(\bar{y}_{RC\_SD}\right)$, we substitute the sample estimates of R_x, S_{yh}^2, S_{xh}^2 and S_{xyh}. The estimate of the $MSE\left(\bar{y}_{RC\_SD}\right)$ is given by

$$\widehat{MSE}\left(\bar{y}_{RC\_SD}\right)=\sum_{h=1}^{L}W_h^2\gamma_h\left(s_{yh}^2+\hat{R}_x^2s_{xh}^2-2\hat{R}_xs_{xyh}\right),$$

where $s_{xyh}^2=\dfrac{\sum_{i=1}^{n_h}\left(x_{hi}-\bar{x}_h\right)\left(y_{hi}-\bar{y}_h\right)}{n_h-1}$, $s_{yh}^2=\dfrac{\sum_{i=1}^{n_h}\left(y_{hi}-\bar{y}_h\right)^2}{n_h-1}$

$s_{xh}^2=\dfrac{\sum_{i=1}^{n_h}\left(x_{hi}-\bar{x}_h\right)^2}{n_h-1}$ and $\hat{R}_x=\dfrac{\bar{y}_{st}}{\bar{x}_{st}+C_x}$.

When the coefficient of variation C_{xh} in stratum h is known, the separate ratio estimator can be modified as

$$\bar{y}_{RS\_SD}=\sum_{h=1}^{L}W_h\frac{\bar{y}_h\left(\bar{X}_h+C_{xh}\right)}{\bar{x}_h+C_{xh}}. \quad (15)$$

To obtain the MSE and bias of the modified separated ratio estimator, applying the MSE and bias of

$\bar{y}_{SDh}=\dfrac{\bar{y}_h\left(\bar{X}_h+C_{xh}\right)}{\bar{x}_h+C_{xh}}$ under simple random sampling to draw in stratum h, the results are as follows:

$$B\left(\bar{y}_{RS\_SD}\right)\approx\sum_{h=1}^{L}W_h\gamma_h\left(\frac{R_{xh}}{\bar{X}_h+C_{xh}}S_{xh}^2-\frac{1}{\bar{X}_h+C_{xh}}S_{xyh}\right) \quad (16)$$

$$MSE\left(\bar{y}_{RS\_SD}\right)\approx\sum_{h=1}^{L}W_h^2\gamma_h\left(S_{yh}^2+R_{xh}^2S_{xh}^2-2R_{xh}S_{xyh}\right), \quad (17)$$

where $R_{xh}=\dfrac{\bar{Y}_h}{\bar{X}_h+C_{xh}}$. For estimating $MSE\left(\bar{y}_{RS\_SD}\right)$, we
substitute the sample estimates as

$$\widehat{MSE}\left(\bar{y}_{RS\_SD}\right)=\sum_{h=1}^{L}W_h^2\gamma_h\left(s_{yh}^2+\hat{R}_{xh}^2s_{xh}^2-2\hat{R}_{xh}s_{xyh}\right), \quad (18)$$

where $\hat{R}_{xh}=\dfrac{\bar{y}_h}{\bar{x}_h+C_{xh}}$. Note that the bias of the modified separate ratio estimator is the cumulative bias of ratio estimates in each stratum. The bias of this estimator may not be negligible when the biases have the same signs in all strata. However, if the sample size in each stratum is large enough, the bias is negligible.

Comparison of Efficiency

We compare the modified combined ratio estimator with the unbiased estimator. The condition is as follows:

$$MSE\left(\bar{y}_{RC\_SD}\right)<MSE\left(\bar{y}_{st}\right)$$
$$\sum_{h=1}^{L}W_h^2\gamma_h\left(S_{yh}^2+R_x^2S_{xh}^2-2R_xS_{xyh}\right)<\sum_{h=1}^{L}W_h^2\gamma_hS_{yh}^2$$
$$\sum_{h=1}^{L}W_h^2\gamma_hS_{xh}^2<2\frac{1}{R_x}\sum_{h=1}^{L}W_h^2\gamma_hS_{xyh} \quad (19)$$
$$\frac{\sum_{h=1}^{L}W_h^2\gamma_hS_{xh}^2}{2\sum_{h=1}^{L}W_h^2\gamma_hS_{xyh}}<\frac{\bar{X}+C_x}{\bar{Y}}$$

When the condition (19) is satisfied, the modified combined ratio estimator will be more efficient than the unbiased estimator. The condition to compare the modified combined ratio estimator with the combined ratio estimator is as follows:

$$\mathrm{MSE}\left(\bar{y}_{RC\_SD}\right) < \mathrm{MSE}\left(\bar{y}_{RC}\right),$$

$$\sum_{h=1}^{L} W_h^2 \gamma_h \left(S_{yh}^2 + R_x^2 S_{xh}^2 - 2R_x S_{xyh}\right) < \sum_{h=1}^{L} W_h^2 \gamma_h \left(S_{yh}^2 + R^2 S_{xh}^2 - 2RS_{xyh}\right),$$

$$\sum_{h=1}^{L} W_h^2 \gamma_h \left(R_x^2 S_{xh}^2 - 2R_x S_{xyh}\right) < \sum_{h=1}^{L} W_h^2 \gamma_h \left(R^2 S_{xh}^2 - 2RS_{xyh}\right),$$

$$R_x^2 A - 2R_x B < R^2 A - 2RB$$

$$\text{where } A = \sum_{h=1}^{L} W_h^2 \gamma_h S_{xh}^2 \text{ and } B = \sum_{h=1}^{L} W_h^2 \gamma_h S_{xyh}^2 . \tag{20}$$

When the condition (20) is satisfied, the modified combined ratio estimator will be more efficient than the combined ratio estimator.

Next, we compare the modified separated ratio estimator with the unbiased estimator. The condition is as follows:

$$\mathrm{MSE}\left(\bar{y}_{RS\_SD}\right) < \mathrm{MSE}\left(\bar{y}_{st}\right)$$

$$\sum_{h=1}^{L} W_h^2 \gamma_h \left(S_{yh}^2 + R_{xh}^2 S_{xh}^2 - 2R_{xh} S_{xyh}\right) < \sum_{h=1}^{L} W_h^2 \gamma_h S_{yh}^2$$

$$\frac{\sum_{h=1}^{L} W_h^2 \gamma_h S_{xh}^2 R_{xh}}{2\sum_{h=1}^{L} W_h^2 \gamma_h S_{xyh} R_{xh}} < 1 \tag{21}$$

When the condition (21) is satisfied, the modified separated ratio estimator is more efficient than the unbiased estimator. The condition to compare the modified separated ratio estimator with the separated ratio estimator is given by

$$\mathrm{MSE}\left(\bar{y}_{RS\_SD}\right) < \mathrm{MSE}\left(\bar{y}_{RS}\right),$$

$$\sum_{h=1}^{L} W_h^2 \gamma_h \left(S_{yh}^2 + R_{xh}^2 S_{xh}^2 - 2R_{xh} S_{xyh}\right) < \sum_{h=1}^{L} W_h^2 \gamma_h \left(S_{yh}^2 + R_h^2 S_{xh}^2 - 2R_h S_{xyh}\right),$$

$$\sum_{h=1}^{L} W_h^2 \gamma_h \left(R_{xh}^2 S_{xh}^2 - 2R_{xh} S_{xyh}\right) < \sum_{h=1}^{L} W_h^2 \gamma_h \left(R_h^2 S_{xh}^2 - 2R_h S_{xyh}\right). \tag{22}$$

When the condition (22) is satisfied, the modified separated ratio estimator will be more efficient than the separated ratio estimator. Note that a difference in efficiency between these latter MSEs of the estimators is the ratio R_{xh} and R_h

Application

We use the dataset from a census of all farms in Jefferson County, Iowa in[5] to demonstrate the relative efficiency of the modified estimators compared with the existing estimators. In this population y_{hi} represents acres in corn and x_{hi} acres in the farm. The population consists of two strata with stratum size 1,580 and 430. Here the sample sizes of the two strata are $n_1 = 70$ and $n_2 = 30$, respectively. Table 1 shows the population characteristics.

Table 1 Populations Characteristics

Stratum	N_h	$\bar{Y}_h$	$\bar{X}_h$	S_{yh}^2	S_{xh}^2	S_{xyh}
1	1,580	82.56	19.4	312	2,055	494
2	430	244.85	51.63	922	7,357	858
Population	**2,010**	**117.28**	**26.30**	**620**	**7,615**	**1,453**

Table 2 The Approximated MSE and Bias of the Estimators

Estimators	Bias	MSE
$\bar{y}_{st}$	0	3.9405
$\bar{y}_{RS}$	-0.1184	3.0599
$\bar{y}_{RC}$	-0.0384	2.9221
$\bar{y}_{RC\_KC}$	-0.0383	2.9214
$\bar{y}_{RS\_SD}$	-0.1181	2.8898
$\bar{y}_{RC\_SD}$	-0.0382	2.9202

From Table 2, the proposed combined ratio estimator gives the smallest absolute bias among the combined and separate ratio estimators. Whereas the modified separate ratio estimator gives smaller absolute bias than the original separate ratio estimator. In addition, the results show that the modified separate ratio estimator produces the smallest MSE and the proposed combined ratio estimator provides smaller MSE than the existing combined ratio estimators. It can be examined that all of the four conditions are satisfied as follows:

$$\frac{\sum_{h=1}^{L} W_h^2 \gamma_h S_{xh}^2}{2\sum_{h=1}^{L} W_h^2 \gamma_h S_{xyh}} = 2.9078 < \frac{\bar{X} + C_x}{\bar{Y}} = 4.4828 ,$$

$$R_x^2 A - 2R_x B = -1.0203 \quad < \quad R^2 A - 2RB = -1.0184 ,$$

$$\frac{\sum_{h=1}^{L} W_h^2 \gamma_h S_{xh}^2 R_{xh}}{2\sum_{h=1}^{L} W_h^2 \gamma_h S_{xyh} R_{xh}} = 0.4158 < 1 \text{ and}$$

$$\sum_{h=1}^{L} W_h^2 \gamma_h \left(R_{xh}^2 S_{xh}^2 - 2R_{xh} S_{xyh} \right) = -1.0507$$

$$< \quad \sum_{h=1}^{L} W_h^2 \gamma_h \left(R_h^2 S_{xh}^2 - 2R_h S_{xyh} \right) = -1.0506 .$$

Therefore, the modified combined and separate ratio estimators are more efficient than the traditional ratio estimators for this data.

Discussion

Using the modified combined ratio estimator, the coefficient of variation and the mean of the auxiliary variable in the whole population must be known. To use the modified separate ratio estimator, the coefficients of variation and the means of the auxiliary variable in all strata are required. The bias of the modified separate ratio estimator is larger than that of the modified combined ratio estimator. Because the formula of the MSE and bias were derived by using the first two terms of Taylor series, the simulation study should be used to compare the accuracy and the efficiency of the estimators in the future. Since the conditions of the efficiency comparison among ratio estimators depend on some unknown parameters, sample estimates of these parameters may be used in practice.

References

1. Murthy, M.N., "Product method of estimation", Sankhya. 26, 1964, 69-74.
2. Sisodia, B. V. S., Dwivedi, V. K., "A modified ratio estimator using coefficient of variation of auxiliary variable", J. Indian Soc. Agricul. Statist. 33: 1981, 13–18.
3. Upadhyaya, L. N., Singh, H. P., Use of transformed auxiliary variable in estimating the finite population mean. Biometrical Journal.41(5), 1999, pp. 627–636.
4. Singh, H.P., Tailor, R., "Use of known correlation coefficient in estimating the finite population mean", Statistics in Transition. 6: 2003, 555–560.
5. Cochran, W.G., Sampling Techniques.Newyork: Jonh Wiley and Sons, 1997.
6. Kadilar, C., Cingi, H., "Ratio estimators in stratified random sampling", Biometrical J.45(2): 2003, 218–225.
7. Kadilar, C., Cingi, H., "A new ratio estimator in stratified random sampling". Communication in Statistics Theory and Method. 34: 2005, 597-620.
8. Prasad, B., Some improved ratio type estimators of population mean and ratio infinite populationsample surveys. Commun. Statist. Theor. Meth.18(1), 1898, 379–392.