Hacettepe Journal of Mathematics and Statistics
Volume 35 (1) (2006), 103 – 109

AN IMPROVEMENT IN ESTIMATING THE POPULATION MEAN BY USING THE CORRELATION COEFFICIENT

Cem Kadılar* and Hülya Çıngı*

Received 10 : 03 : 2005 : Accepted 02 : 06 : 2006

Abstract

We propose a class of ratio estimators for the estimation of population mean by adapting the estimators in Upadhyaya and Singh [7] to the estimator in Singh and Tailor [6]. We obtain mean square error (MSE) equations for all proposed estimators and find theoretical conditions that make each proposed estimator more efficient than the traditional estimators and ratio estimator in Singh and Tailor [6]. In addition, these conditions are satisfied with an application with original data.

Keywords: Ratio estimator, Auxiliary variable, Simple random sampling, Efficiency.

2000 AMS Classification: 62 D 05

1. Introduction

The classical ratio estimator for the population mean $\overline{Y}$ of the study variable y is defined by

$$(1) \qquad \bar{y}_r = \frac{\overline{y}}{\overline{x}}\bar{X},$$

where $\overline{y}$ and $\overline{x}$ are the sample means of the study and auxiliary variables, respectively, and it is assumed that the population mean $\overline{X}$ of the auxiliary variable x is known [2]. The MSE of this estimator is as follows:

$$(2) \qquad \mathrm{MSE}\,(\bar{y}_r) \cong \frac{1-f}{n}\overline{Y}^2\left[C_y^2 + C_x^2\,(1-2\theta)\right].$$

Here $f = \frac{n}{N}$; n is the sample size; N is the number of units in the population; $\theta = \rho\frac{C_y}{C_x}$; C_x and C_y are the population coefficients of variation of auxiliary and study variables, respectively [1].

*Hacettepe University, Faculty of Science, Department of Statistics, 06800 Beytepe, Ankara, Turkey. E-mail : (C. Kadılar) kadilar@hacettepe.edu.tr (H. Çıngı) hcingi@hacettepe.edu.tr

Singh and Tailor [6] suggested the following ratio estimator:

$$\overline{y}_{ST} = \frac{\overline{y}}{\overline{x}+\rho}\left(\overline{X}+\rho\right), \tag{3}$$

where ρ is the correlation coefficient between auxiliary and study variables. The MSE of this ratio estimator is as follows:

$$\text{MSE}\left(\bar{y}_{ST}\right) \cong \frac{1-f}{n}\overline{Y}^2\left[C_y^2 + C_x^2\omega\left(\omega - 2\theta\right)\right], \tag{4}$$

where $\omega = \frac{\overline{X}}{\overline{X}+\rho}$.

2. The Suggested Estimators

Motivated by the estimators in Upadhyaya and Singh [7], we propose ratio estimators using the correlation coefficient as follows:

$$\overline{y}_{pr1} = \frac{\overline{y}}{\overline{x}C_x+\rho}\left(\overline{X}C_x+\rho\right) \tag{5}$$

$$\overline{y}_{pr2} = \frac{\overline{y}}{\overline{x}\rho+C_x}\left(\overline{X}\rho+C_x\right) \tag{6}$$

$$\overline{y}_{pr3} = \frac{\overline{y}}{\overline{x}\beta_2\left(x\right)+\rho}\left(\overline{X}\beta_2\left(x\right)+\rho\right) \tag{7}$$

$$\overline{y}_{pr4} = \frac{\overline{y}}{\overline{x}\rho+\beta_2\left(x\right)}\left(\overline{X}\rho+\beta_2\left(x\right)\right), \tag{8}$$

where $\beta_2\left(x\right)$ is the population coefficient of the kurtosis of the auxiliary variable. We assume that C_x, ρ, and $\beta_2\left(x\right)$ are known.

We obtain the MSE and bias equations for these proposed estimators as

$$\begin{aligned} \text{MSE}\left(\bar{y}_{pri}\right) &\cong \frac{1-f}{n}\overline{Y}^2\left[C_y^2 + C_x^2\lambda_i\left(\lambda_i - 2\theta\right)\right],\ i = 1,2,3,4, \\ B\left(\bar{y}_{pri}\right) &\cong \frac{1-f}{n}\overline{Y}\ C_x^2\lambda_i\left(\lambda_i - \theta\right),\ i = 1,2,3,4, \end{aligned} \tag{9}$$

respectively, where $\lambda_1 = \frac{\overline{X}C_x}{\overline{X}C_x+\rho}$; $\lambda_2 = \frac{\overline{X}\rho}{\overline{X}\rho+C_x}$; $\lambda_3 = \frac{\overline{X}\beta_2(x)}{\overline{X}\beta_2(x)+\rho}$ and $\lambda_4 = \frac{\overline{X}\rho}{\overline{X}\rho+\beta_2(x)}$. (for details, please see the Appendix)

3. Efficiency Comparisons

In this section, we try to obtain the efficiency conditions for the proposed estimators by comparing the MSE of the proposed estimators with the MSE of the sample mean, traditional ratio estimator and the ratio estimator suggested by Singh and Tailor [6].

It is well known that under simple random sampling without replacement (SRSWOR) the variance of the sample mean is

$$\begin{aligned} V\left(\overline{y}\right) &= \frac{1-f}{n}S_y^2 \\ &= \frac{1-f}{n}\overline{Y}^2C_y^2. \end{aligned} \tag{10}$$

Comparing the MSE of the proposed estimators, given in (9), with the variance of this sample mean, we have the following conditions:

$$\mathrm{MSE}(\overline{y}_{pri}) < V(\overline{y}),\ i = 1,2,3,4.,$$

$$C_x^2\lambda_i(\lambda_i - 2\theta) < 0$$

$$\lambda_i < 2\theta \text{ if } \lambda_i > 0 \tag{11}$$

$$\lambda_i > 2\theta \text{ if } \lambda_i < 0. \tag{12}$$

When either of the restrictions (11) or (12) is satisfied, the proposed estimators are more efficient than the sample mean.

Comparing the MSE of the proposed estimators with the MSE of the classical ratio estimator, given in (2), we have the following conditions:

$$\mathrm{MSE}\left(\overline{y}_{pri}\right) < \mathrm{MSE}\left(\overline{y}_r\right)\ i = 1,2,3,4.,$$

$$\lambda_i\left(\lambda_i - 2\theta\right) < 1 - 2\theta,$$

$$\lambda_i^2 - 2\lambda_i\theta + 2\theta < 1$$

$$\lambda_i^2 + 2\theta\left(1 - \lambda_i\right) < 1$$

$$\left(\lambda_i + 1\right) > 2\theta \text{ if } \lambda_i < 1, \tag{13}$$

$$\left(\lambda_i + 1\right) < 2\theta \text{ if } \lambda_i > 1. \tag{14}$$

When either of the conditions (13) or (14) is satisfied, the proposed estimators are more efficient than the traditional ratio estimator.

Comparing the MSE of the proposed estimators with the MSE of the estimator in Singh and Tailor [6], given in (4), we have the following conditions:

$$\mathrm{MSE}\left(\overline{y}_{pri}\right) < \mathrm{MSE}\left(\overline{y}_{ST}\right),\ i = 1,2,3,4,$$

$$\lambda_i\left(\lambda_i - 2\theta\right) < \omega\left(\omega - 2\theta\right),$$

$$\lambda_i^2 - 2\lambda_i\theta < \omega^2 - 2\omega\theta,$$

$$\lambda_i^2 - \omega^2 < 2\lambda_i\theta - 2\omega\theta,$$

$$\left(\lambda_i - \omega\right)\left(\lambda_i + \omega\right) < 2\theta\left(\lambda_i - \omega\right),$$

$$\left(\lambda_i + \omega\right) < 2\theta \text{ if } \lambda_i > \omega, \tag{15}$$

$$\left(\lambda_i + \omega\right) > 2\theta \text{ if } \lambda_i < \omega. \tag{16}$$

When either of the conditions (15) or (16) is satisfied, the proposed estimators are more efficient than the ratio estimator suggested by Singh and Tailor [6].

Comparing the MSE of one proposed estimator with another, we have the following conditions:

$$\mathrm{MSE}\left(\overline{y}_{pri}\right) < \mathrm{MSE}\left(\overline{y}_{prj}\right),\ i = 1,2,3,4;\ j = 1,2,3,4 \text{ and } i \neq j,$$

$$\lambda_i\left(\lambda_i - 2\theta\right) < \lambda_j\left(\lambda_j - 2\theta\right),$$

$$\lambda_i^2 - \lambda_j^2 < 2\theta\left(\lambda_i - \lambda_j\right),$$

$$\left(\lambda_i - \lambda_j\right)\left(\lambda_i + \lambda_j\right) < 2\theta\left(\lambda_i - \lambda_j\right),$$

$$\left(\lambda_i + \lambda_j\right) < 2\theta \text{ if } \lambda_i > \lambda_j, \tag{17}$$

$$\left(\lambda_i + \lambda_j\right) > 2\theta \text{ if } \lambda_i < \lambda_j. \tag{18}$$

When either of the conditions (17) or (18) is satisfied, the ith proposed estimator is more efficient than the jth proposed estimator.

4. An Application

In this section we use the data set in Kadilar and Cingi [4]. We apply the traditional ratio estimator, given in (1), the Singh-Tailor ratio estimator, given in (3), and the proposed estimators, given in (5)-(8), to data concerning the level of apple production (as the study variable) and the number of apple trees (as the auxiliary variable) (1 unit=1000 trees) in 104 villages in the East Anatolia region of Turkey in 1999 (Source: Institute of Statistics, Republic of Turkey). We take the sample size as $n = 20$ and use simple random sampling [1]. The MSE of these estimators are computed as given in (2), (4) and (9) and these estimators are compared to each other with respect to their MSE values.

Table 1. Data Statistics

$N = 104$	$C_y = 1.866$	$\lambda_1 = 0.964$
$n = 20$	$C_x = 1.653$	$\lambda_2 = 0.879$
$\rho = 0.865$	$\beta_2(x) = 17.52$	$\lambda_3 = 0.996$
$\overline{X} = 13.93$	$\theta = 0.977$	$\lambda_4 = 0.408$
$\overline{Y} = 625.37$	$\omega = 0.942$	

We observe in Table 1 statistics about the population. Note that the correlation between the variables is 87%. When we examine the conditions determined in Section 3 for this data set, they are satisfied for the first and third proposed estimators as follows:

$$\left.\begin{array}{lll}\lambda_1 = 0.964 & : & 2\theta = 1.95 \\ \lambda_3 = 0.996 & : & \lambda_1, \lambda_3 > 0\end{array}\right. \Longrightarrow \text{Condition (11) is satisfied.}$$

$$\left.\begin{array}{lll}\lambda_1 + 1 = 1.96 & : & 2\theta = 1.95 \\ \lambda_3 + 1 = 2.00 & : & \lambda_1, \lambda_3 < 1\end{array}\right. \Longrightarrow \text{Condition (13) is satisfied.}$$

$$\left.\begin{array}{lll}\lambda_1 + \omega = 1.91 & : & 2\theta = 1.95 \\ \lambda_3 + \omega = 1.94 & : & \lambda_1, \lambda_3 > \omega\end{array}\right. \Longrightarrow \text{Condition (15) is satisfied.}$$

For the MSE comparison between the first and third proposed estimators, the condition (18) is satisfied as follows:

$$\lambda_1 + \lambda_3 = 1.96 \quad : \quad 2\theta = 1.95 \quad : \quad \lambda_1 < \lambda_3.$$

Thus, the first proposed estimator is more efficient than the third.

As a result, we suggest that we should apply the first and third estimators to this data set. In Table 2, the values of the MSE are given. As expected, it is seen that the first and third proposed estimators have a smaller MSE than the sample mean, the traditional ratio and the Singh-Tailor ratio estimators. It is obvious that the proposed estimators are more efficient than the other estimators. It is worth pointing out that the classical ratio estimator is more efficient than the ratio estimator, suggested by Singh and Tailor [6], for this data set.

Table 2. MSE Values of the Ratio Estimators

sample mean $(\overline{y})$	55000.09
traditional $(\overline{y}_r)$	13860.27
Singh-Tailor $(\overline{y}_{ST})$	13890.58
proposed 1 $\left(\overline{y}_{pr1}\right)$	13844.27
proposed 3 $\left(\overline{y}_{pr3}\right)$	13853.73

5. Conclusion

We develop some ratio estimators using the correlation coefficient and give a theoretical argument to show that the proposed estimators have a smaller MSE than the traditional and the Singh-Tailor ratio estimators under certain conditions. These theoretical conditions are also satisfied by the results of an application with original data. In future work, we hope to adapt the ratio estimators, presented here, to ratio estimators suggested by Kadilar and Cingi [3] and to ratio estimators in stratified random sampling as in Kadilar and Cingi [4,5].

Appendix

To the first degree of approximation, the MSE of the third proposed estimator can be found using the Taylor series method defined by

$$\text{(A1)} \qquad \text{MSE}\left(\overline{y}_{pr}\right) \cong \mathbf{d}\,\mathbf{\Sigma}\,\mathbf{d}\,'$$

where

$$\mathbf{d} = \left[\frac{\partial h(a,b)}{\partial a}\Big|_{\overline{Y},\,\overline{X}} \quad \frac{\partial h(a,b)}{\partial b}\Big|_{\overline{Y},\,\overline{X}}\right]$$

$$\mathbf{\Sigma} = \frac{1-f}{n}\begin{bmatrix} S_y^2 & S_{yx} \\ S_{xy} & S_x^2 \end{bmatrix}$$

(see Wolter [8]). Here $h(a,b) = h(\overline{y},\overline{x}) = \overline{y}_{pr3}$ in (7); S_y^2 and S_x^2 denote the population variances of the study and the auxiliary variables, respectively, and $S_{yx} = S_{xy}$ denotes the population covariance between the study and the auxiliary variables. According to this definition, we obtain $\mathbf{d}$ for the third proposed estimator as

$$\mathbf{d} = \left[1 \quad -\frac{\overline{Y}\beta_2(x)}{\overline{X}\beta_2(x)+\rho}\right]$$

We obtain the MSE equation of the third proposed estimator using (A1) as follows:

$$\begin{aligned}
\text{MSE}\left(\overline{y}_{pr3}\right) &\cong \frac{1-f}{n}\left(S_y^2 - 2\frac{\overline{Y}\beta_2(x)}{\overline{X}\beta_2(x)+\rho}S_{yx} + \frac{\overline{Y}^2\beta_2^2(x)}{\left[\overline{X}\beta_2(x)+\rho\right]^2}S_x^2\right) \\
&\cong \frac{1-f}{n}\overline{Y}^2\left(C_y^2 - 2\frac{\beta_2(x)}{\left[\overline{X}\beta_2(x)+\rho\right]\overline{Y}}S_{yx} + \frac{\beta_2^2(x)}{\left[\overline{X}\beta_2(x)+\rho\right]^2}S_x^2\right) \\
&\cong \frac{1-f}{n}\overline{Y}^2\left\{C_y^2 + \frac{\beta_2(x)\,\overline{X}}{\overline{X}\beta_2(x)+\rho}\left(\frac{\beta_2(x)\,\overline{X}\;S_x^2}{\left[\overline{X}\beta_2(x)+\rho\right]\overline{X}^2} - 2\frac{S_{yx}}{\overline{XY}}\right)\right\} \\
&\cong \frac{1-f}{n}\overline{Y}^2\left[C_y^2 + \lambda_3\left(\lambda_3 C_x^2 - 2\rho C_y C_x\right)\right] \\
&\cong \frac{1-f}{n}\overline{Y}^2\left[C_y^2 + C_x^2\lambda_3\left(\lambda_3 - 2\theta\right)\right].
\end{aligned}$$

We would like to remark that the MSE equations of the other proposed estimators can easily be obtained in the same way.

In general, Taylor series method for k variables can be given as

$$h\left(\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_k\right) = h\left(\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_k\right) + \sum_{j=1}^{k} d_j\left(\overline{x}_j - \overline{X}_j\right) + R_k\left(\overline{X}_k, a\right)$$

where

$$d_j = \frac{\partial h\left(a_1, a_2, \ldots, a_k\right)}{\partial a_j}$$

and

$$R_k\left(\overline{X}_k, a\right) = \sum_{j=1}^{k}\sum_{i=1}^{k}\frac{1}{2!}\frac{\partial^2 h\left(\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_k\right)}{\partial\overline{X}_i\partial\overline{X}_j}\left(\overline{x}_j - \overline{X}_j\right)\left(\overline{x}_i - \overline{X}_i\right) + O_k.$$

Here O_k represents the terms in the expansion of the Taylor series of more than the second degree. When we omit the term $R_k\left(\overline{X}_k, a\right)$, we obtain Taylor series method for two variables as follows:

$$\text{(A2)}\qquad h(\overline{x}, \overline{y}) - h(\overline{X}, \overline{Y}) \cong \left.\frac{\partial h(c,d)}{\partial c}\right|_{\overline{X},\overline{Y}}\left(\overline{x} - \overline{X}\right) + \left.\frac{\partial h(c,d)}{\partial d}\right|_{\overline{X},\overline{Y}}\left(\overline{y} - \overline{Y}\right).$$

Instead of using (A2), we can also obtain the MSE of the proposed estimator using (A1), an alternative method to (A2) [8]. If we do not omit the term $R_k\left(\overline{X}_k, a\right)$, we obtain the bias of the proposed estimators by the following equation:

$$\text{(A3)}\qquad B\left(\overline{y}_{pri}\right) = \frac{1}{2}\left[d_{11}V\left(\overline{y}\right) + d_{12}cov\left(\overline{y}, \overline{x}\right) + d_{21}cov\left(\overline{x}, \overline{y}\right) + d_{22}V\left(\overline{x}\right)\right],$$

where

$$d_{11} = \frac{\partial^2 h\left(\overline{y}_{pri}\right)}{\partial\overline{y}\partial\overline{y}} = 0;$$

$$d_{12} = d_{21} = \frac{\partial^2 h\left(\overline{y}_{pri}\right)}{\partial\overline{x}\partial\overline{y}} = \frac{\partial^2 h\left(\overline{y}_{pri}\right)}{\partial\overline{y}\partial\overline{x}} = -\frac{\lambda_i}{\overline{X}};\ d_{22} = \frac{\partial^2 h\left(\overline{y}_{pri}\right)}{\partial\overline{x}\partial\overline{x}} = 2\frac{R}{\overline{X}}\lambda_i^2.$$

Using these equalities, we can write (A3) as

$$\begin{aligned} B\left(\overline{y}_{pri}\right) &\cong \frac{1}{2}\left(-2\frac{\lambda_i}{\overline{X}}\,\frac{1-f}{n}S_{yx} + 2\frac{R}{\overline{X}}\lambda_i^2\frac{1-f}{n}S_x^2\right) \\ &\cong \frac{1-f}{n}\overline{Y}\lambda_i\left(\lambda_i C_x^2 - C_{yx}\right) \\ &\cong \frac{1-f}{n}\overline{Y}\lambda_i C_x^2\left(\lambda_i - \theta\right);\ i = 1, 2, 3, 4. \end{aligned}$$

where $C_{yx} = \rho C_y C_x$.

References

[1] Cingi, H. *Örnekleme Kuramı* (Hacettepe University Press, 1994).

[2] Cochran, W. G. *Sampling Techniques* (John Wiley and Sons, New-York, 1977).

[3] Kadilar, C. and Cingi, H. *Ratio estimators in simple random sampling*, Applied Mathematics and Computation **151**, 893–902, 2004.

[4] Kadilar, C. and Cingi, H. *Ratio estimators in stratified random sampling*, Biometrical Journal **45** (2), 218–225, 2003.

[5] Kadilar, C. and Cingi, H. *A new ratio estimator in stratified random sampling*, Communications in Statistics: Theory and Methods **34** (3), 2005.

[6] Singh, H. P. and Tailor, R. *Use of known correlation coefficient in estimating the finite population mean*, Statistics in Transition **6** (4), 555–560, 2003.

[7] Upadhyaya, L. N. and Singh, H. P. *Use of transformed auxiliary variable in estimating the finite population mean*, Biometrical Journal **41** (5), 627–636, 1999.
[8] Wolter, K. M. *Introduction to Variance Estimation* (Springer-Verlag, 2003).