

New Ratio Estimators Using Correlation Coefficient

Cem Kadilar and Hulya Cingi

Hacettepe University, Department of Statistics, Beytepe, 06800, Ankara, Turkey.

e-mails : kadilar@hacettepe.edu.tr ; hcingi@hacettepe.edu.tr

Abstract : We propose a class of ratio estimators for the estimation of population mean by adapting the estimators in Singh and Tailor (2003) to the estimators in Kadilar and Cingi (2004). 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 similarly for comparison to the ratio estimator in Singh and Tailor (2003), and for those in Kadilar and Cingi (2004). In addition, these conditions are supported by a numerical example.

Key words : Ratio estimator, auxiliary variable, simple random sampling, efficiency.

2000 AMS Classification : 62 D 05, 62 G 05

1. Introduction

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

$$\bar{y}_r = \frac{\bar{y}}{\bar{x}} \bar{X} \quad (1)$$

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

$$MSE(\bar{y}_r) \cong \frac{1-f}{n} (S_y^2 + R^2 S_x^2 [1 - 2\theta]) \quad (2)$$

where $f = \frac{n}{N}$; n is the sample size; N is the number of units in the population;

S_x^2 and S_y^2 are the population variances of auxiliary and study variables,

respectively; $R = \frac{\bar{Y}}{\bar{X}}$ and $\theta = \rho \frac{C_y}{C_x}$. Here C_x and C_y are the population

coefficients of variation of auxiliary and study variables, respectively (Cochran, 1977, pages 151-154).

Singh and Tailor (2003) suggested the following ratio estimator:

$$\bar{y}_{ST} = \frac{\bar{y}}{\bar{x} + \rho} (\bar{X} + \rho) \quad (3)$$

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

$$MSE(\bar{y}_{ST}) \cong \frac{1-f}{n} (S_y^2 + R^2 S_x^2 \omega [\omega - 2\theta]) \quad (4)$$

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

Kadilar and Cingi (2004) suggested the following ratio estimators:

$$\bar{y}_{KC1} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{\bar{x}} \bar{X} \quad (5)$$

$$\bar{y}_{KC2} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{\bar{x} + \beta_2(x)} [\bar{X} + \beta_2(x)] \quad (6)$$

$$\bar{y}_{KC3} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{\bar{x} + C_x} (\bar{X} + C_x) \quad (7)$$

$$\bar{y}_{KC4} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{\bar{x}\beta_2(x) + C_x} [\bar{X}\beta_2(x) + C_x] \quad (8)$$

$$\bar{y}_{KC5} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{\bar{x}C_x + \beta_2(x)} [\bar{X}C_x + \beta_2(x)] \quad (9)$$

where $\beta_2(x)$ is the population coefficient of the kurtosis of the auxiliary variable and $b = \frac{s_{yx}}{s_x^2}$ is the regression coefficient. Here s_x^2 is the sample variance of auxiliary variable and s_{yx} is the sample covariance between the study and auxiliary variables.

Kadilar and Cingi (2004) obtained the MSE equation of these ratio estimators as follows:

$$MSE(\bar{y}_{KC_i}) \cong \frac{1-f}{n} [R_{KC_i}^2 S_x^2 + S_y^2 (1-\rho^2)] \quad ; i = 1,2,\dots,5 \quad (10)$$

where $R_{KC1} = R = \frac{\bar{Y}}{\bar{X}}$; $R_{KC2} = \frac{\bar{Y}}{\bar{X} + \beta_2(x)}$; $R_{KC3} = \frac{\bar{Y}}{\bar{X} + C_x}$; $R_{KC4} = \frac{\bar{Y}\beta_2(x)}{\bar{X}\beta_2(x) + C_x}$

and $R_{KC5} = \frac{\bar{Y}C_x}{\bar{X}C_x + \beta_2(x)}$.

2. Suggested Estimators

Adapting the estimator given in (3) to the estimators given in (5)-(9), we develop new ratio estimators using the correlation coefficient as follows:

$$\bar{y}_{pr1} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{\bar{x} + \rho}(\bar{X} + \rho) \quad (11)$$

$$\bar{y}_{pr2} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{\bar{x}C_x + \rho}(\bar{X}C_x + \rho) \quad (12)$$

$$\bar{y}_{pr3} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{\bar{x}\rho + C_x}(\bar{X}\rho + C_x) \quad (13)$$

$$\bar{y}_{pr4} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{\bar{x}\beta_2(x) + \rho}[\bar{X}\beta_2(x) + \rho] \quad (14)$$

$$\bar{y}_{pr5} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{\bar{x}\rho + \beta_2(x)}[\bar{X}\rho + \beta_2(x)] \quad (15)$$

We obtain the MSE equation for these proposed estimators as

$$MSE(\bar{y}_{pri}) \cong \frac{1-f}{n} [R_{pri}^2 S_x^2 + S_y^2 (1-\rho^2)] \quad ; i = 1, 2, \dots, 5 \quad (16)$$

where $R_{pr1} = \frac{\bar{Y}}{\bar{X} + \rho}$; $R_{pr2} = \frac{\bar{Y}C_x}{\bar{X}C_x + \rho}$; $R_{pr3} = \frac{\bar{Y}\rho}{\bar{X}\rho + C_x}$; $R_{pr4} = \frac{\bar{Y}\beta_2(x)}{\bar{X}\beta_2(x) + \rho}$ and
 $R_{pr5} = \frac{\bar{Y}\rho}{\bar{X}\rho + \beta_2(x)}$. (for details see 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 estimators suggested by Singh and Tailor (2003) and Kadilar and Cingi (2004).

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

$$V(\bar{y}) = \frac{1-f}{n} S_y^2. \quad (17)$$

We first compare the MSE of the proposed estimators, given in (16), with the variance of the sample mean. We have the following condition:

$$\text{Let } \upsilon = \frac{S_x^2}{S_y^2}$$

$$MSE(\bar{y}_{pri}) < V(\bar{y}) \quad ; i = 1, 2, \dots, 5.$$

$$R_{pri}^2 S_x^2 - S_y^2 \rho^2 < 0,$$

$$\rho^2 > \upsilon R_{pri}^2 \quad (18)$$

When this condition is satisfied, proposed estimators are more efficient than the sample mean.

Secondly, we compare the MSE of the proposed estimators with the MSE of the classical ratio estimator, given in (2). We have the following condition:

$$\begin{aligned}
 &MSE(\bar{y}_{pri}) < MSE(\bar{y}_r) \quad ; i = 1, 2, \dots, 5. \\
 &R_{pri}^2 S_x^2 - S_y^2 \rho^2 < R^2 S_x^2 - 2R^2 S_x^2 \theta, \\
 &\rho^2 > \upsilon (R_{pri}^2 - R^2 + 2R^2 \theta). \tag{19}
 \end{aligned}$$

When this condition is satisfied, proposed estimators are more efficient than the traditional ratio estimator.

Thirdly, we compare the MSE of the proposed estimators with the MSE of the estimator in Singh and Tailor (2003), given in (4). We have the following condition:

$$\begin{aligned}
 &MSE(\bar{y}_{pri}) < MSE(\bar{y}_{ST}) \quad ; i = 1, 2, \dots, 5 \\
 &R_{pri}^2 S_x^2 - S_y^2 \rho^2 < R^2 S_x^2 \omega^2 - 2R^2 S_x^2 \omega \theta, \\
 &\rho^2 > \upsilon (R_{pri}^2 - R^2 \omega^2 + 2R^2 \omega \theta). \tag{20}
 \end{aligned}$$

When this condition is satisfied, proposed estimators are more efficient than the ratio estimator, suggested by Singh and Tailor (2003).

Finally, we compare the MSE of the proposed estimators with the MSE of the estimators in Kadilar and Cingi (2004), given in (10). We have the following condition:

$$\begin{aligned}
 &MSE(\bar{y}_{pri}) < MSE(\bar{y}_{KCj}) \quad ; i = 1, 2, \dots, 5 \text{ and } j = 1, 2, \dots, 5. \\
 &R_{pri}^2 < R_{KCj}^2 \tag{21}
 \end{aligned}$$

When this condition is satisfied, proposed estimators are more efficient than the ratio estimators, suggested by Kadilar and Cingi (2004). We can examine the

condition (21) for each proposed estimator. For example, when we take $i = j = 1$, we obtain the condition:

$$\begin{aligned} \bar{X}^2 &< \bar{X}^2 + 2\bar{X}\rho + \rho^2 \\ \rho(2\bar{X} + \rho) &> 0 \end{aligned}$$

As ρ is positive in ratio estimation we have the following condition:

$$\rho > -2\bar{X}$$

This condition is always satisfied if the auxiliary variable has positive data. In other words, first proposed estimator is more efficient than first ratio estimator in Kadilar and Cingi (2004) when the data are positive. Detail comparisons for the other proposed estimators can also be studied in the same way. Note that the efficiency comparisons among the proposed estimators also result the similar condition with (21) as follows:

$$R_{pri}^2 < R_{prj}^2 \quad ; i \neq j = 1, 2, \dots, 5 \quad (22)$$

When this condition is satisfied, i th proposed estimator is more efficient than j th proposed estimator.

4. Numerical Example

In this section, we apply the traditional ratio estimator, given in (1), the Singh-Tailor ratio estimator, given in (3), Kadilar-Cingi ratio estimators, given in (5)-(9) and proposed estimators, given in (11)-(15), to data whose statistics are given in Table 1. We assume to take the sample size $n=50$ from $N=200$ using SRSWOR. The MSE of these estimators are computed as given in (2), (4), (10) and (16) and these estimators are compared to each other with respect to their MSE values.

Table 1

Data Statistics

$N = 200$	$\bar{Y} = 500$	$R_{KC1} = 20.00$	$R_{pr1} = 19.31$
$n = 50$	$\bar{X} = 25$	$R_{KC2} = 6.67$	$R_{pr2} = 19.65$
$\rho = 0.90$	$\beta_2(x) = 50$	$R_{KC3} = 18.52$	$R_{pr3} = 18.37$
$C_y = 15$	$\theta = 6.75$	$R_{KC4} = 19.97$	$R_{pr4} = 19.99$
$C_x = 2$	$\omega = 0.97$	$R_{KC5} = 10.00$	$R_{pr5} = 6.00$

From Table 2, we understand that the most efficient estimator is fifth proposed estimator. When we examine the conditions, determined in Section 3, for this data set, we see that all of them are satisfied for fifth proposed estimator as follows:

$$\rho^2 = 0.81 \text{ and } v R_{pr5}^2 = 0.0166$$

\Rightarrow the condition (18) is satisfied.

$$\rho^2 = 0.81 \text{ and } v(R_{pr5}^2 - R^2 + 2R^2\theta) = 0.2388$$

\Rightarrow the condition (19) is satisfied.

$$\rho^2 = 0.81 \text{ and } v(R_{pr5}^2 - R^2\omega^2 + 2R^2\omega\theta) = 0.2317$$

\Rightarrow the condition (20) is satisfied.

$$R_{pr5}^2 < R_{KC i}^2 \quad ; i = 1, 2, \dots, 5$$

\Rightarrow the condition (21) is satisfied.

$$R_{pr5}^2 < R_{pr j}^2 \quad ; j = 1, 2, 3, 4$$

\Rightarrow the condition (22) is satisfied.

Therefore, we suggest that we should apply fifth proposed estimator to this data set. It is worth point out that the traditional ratio estimator is more efficient than the ratio estimator, suggested by Singh and Tailor (2003), for this data set.

Table 2
MSE Values of Ratio Estimators

Estimators	MSE	Estimators	MSE	Estimators	MSE
\bar{y}	843750.00	\bar{y}_{KC1}	175312.50	\bar{y}_{pr1}	174288.14
\bar{y}_r	656250.00	\bar{y}_{KC2}	161979.17	\bar{y}_{pr2}	174786.74
\bar{y}_{ST}	662262.32	\bar{y}_{KC3}	173172.58	\bar{y}_{pr3}	172963.48
		\bar{y}_{KC4}	175264.61	\bar{y}_{pr4}	175290.92
		\bar{y}_{KC5}	164062.50	\bar{y}_{pr5}	161757.21

5. Conclusion

We develop some ratio estimators using the correlation coefficient and theoretically show that the proposed estimators have a smaller MSE than the traditional, the Singh and Tailor's (2003) and Kadilar and Cingi's (2004) ratio estimators in certain conditions. These theoretical conditions are also satisfied by the results of a numerical example. In future work, we hope to adapt the ratio estimators, presented here, to ratio estimators in stratified random sampling as in Kadilar and Cingi (2003; 2005).

Appendix

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

$$MSE(\bar{y}_{pr3}) \cong \delta \Sigma \delta' \quad (\text{A.1})$$

where

$$\delta = \left[\frac{\partial h(c, d)}{\partial c} \Big|_{B, \bar{y}, \bar{x}} \quad \frac{\partial h(c, d)}{\partial d} \Big|_{B, \bar{y}, \bar{x}} \right],$$

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

(see Wolter, 2003, pages 221-228). Here $B = \frac{S_{yx}}{S_x^2}$; $h(c, d) = h(\bar{y}, \bar{x}) = \bar{y}_{pr3}$ in (13)

and $S_{yx} = S_{xy}$ denotes the population covariance between study and auxiliary variables. According to this definition, we obtain δ for the third proposed estimator as

$$\delta = [1 \quad -R_{pr3} - B].$$

We obtain the MSE equation of the third proposed estimator using (A.1) as follows:

$$\begin{aligned} MSE(\bar{y}_{pr3}) &\cong \frac{1-f}{n} (S_y^2 - 2R_{pr3}S_{yx} - 2BS_{yx} + R_{pr3}^2S_x^2 + 2R_{pr3}BS_x^2 + B^2S_x^2) \\ &= \frac{1-f}{n} \left(S_y^2 - 2R_{pr3}S_{yx} - 2\frac{S_{yx}^2}{S_x^2} + R_{pr3}^2S_x^2 + 2R_{pr3}S_{yx} + \frac{S_{yx}^2}{S_x^2} \right) \\ &= \frac{1-f}{n} (S_y^2 - 2\rho^2S_y^2 + R_{pr3}^2S_x^2 + \rho^2S_y^2) \\ &= \frac{1-f}{n} (R_{pr3}^2S_x^2 - \rho^2S_y^2 + S_y^2) \\ &= \frac{1-f}{n} [R_{pr3}^2S_x^2 + S_y^2(1-\rho^2)]. \end{aligned}$$

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

References

- Cingi, H. (1994). Sampling Theory. Hacettepe University Press.
- Cochran, W.G. (1977). Sampling Techniques. John Wiley and Sons, New-York.
- Kadilar, C. and Cingi, H. (2005). A New Ratio Estimator in Stratified Random Sampling. Communications in Statistics: Theory and Methods, 34, 597-602.
- Kadilar, C. and Cingi, H. (2004). Ratio Estimators in Simple Random Sampling. Applied Mathematics and Computation, 151, 893-902.
- Kadilar, C. and Cingi, H. (2003). Ratio Estimators in Stratified Random Sampling. Biometrical Journal, 45, 218-225.
- Singh, H. P. and Tailor, R. (2003). Use of Known Correlation Coefficient in Estimating the Finite Population Mean. Statistics in Transition, 6, 555-560.
- Upadhyaya, L. N. and Singh, H. P. (1999). Use of Transformed Auxiliary Variable in Estimating the Finite Population Mean. Biometrical Journal, 41, 627-636.
- Wolter, K. M. (2003). Introduction to Variance Estimation. Springer-Verlag.