Pak. J. Statist.
2008 Vol. 24(3), 167-172

RATIO ESTIMATORS WITH UNEQUAL PROBABILITY DESIGNS

Sevil Bacanli[1] and **Cem Kadilar**[2]
Department of Statistics, Hacettepe University,
Beytepe 06800, Ankara, Turkey.
Email: [1]sevil@hacettepe.edu.tr, [2]kadilar@hacettepe.edu.tr

ABSTRACT

We propose ratio estimators that can be used under unequal probability designs by adapting Horvitz-Thompson estimators to ratio estimators in literature. Mean square error (MSE) of all proposed ratio estimators is obtained and compared with the MSE of classical estimators. By these comparisons the conditions, which make each proposed estimator more efficient than the classical one, are found. The theoretical results are supported by a numerical example.

KEY WORDS

Auxiliary variable, Horvitz-Thompson estimators, efficiency, Ratio-type estimators,

1. INTRODUCTION

The classical ratio estimate for finite population mean $\bar{Y}$ of the study variable y and the mean square error (MSE) of this estimator are given by

$$\bar{y}_r = \frac{\bar{y}}{\bar{x}}\bar{X} = \hat{R}\bar{X} \tag{1.1}$$

$$MSE\left(\bar{y}_r\right) \cong \frac{1-f}{n}\left(R^2 S_x^2 - 2RS_{xy} + S_y^2\right), \tag{1.2}$$

respectively. Here $\bar{y}$ is the sample mean of the study variable; $\bar{x}$ is the sample mean of the auxiliary variable; $f = \frac{n}{N}$; n is the sample size and N is the population size. From equation (1.1), $\hat{R} = \frac{\bar{y}}{\bar{x}}$ so $R \doteq \frac{\bar{Y}}{\bar{X}}$ in (1.2) is the population ratio. S_{xy} is the population covariance between the auxiliary and study variables; S_x^2 and S_y^2 are population variances of auxiliary and study variables, respectively. Note that it is assumed that the population mean $\bar{X}$ of the auxiliary variable x is known (Cochran, 1977).

When the population coefficient of variation and kurtosis of the auxiliary variable C_x and $\beta_2\left(x\right)$, respectively, are known, Sisodia and Dwivedi (1981), Singh and Kakran (1993), Upadhyaya and Singh (1999) suggest ratio-type estimators for $\bar{Y}$ as

$$\bar{y}_{SD} = \bar{y}\,\frac{\bar{X}+C_x}{\bar{x}+C_x} = \frac{\bar{y}}{\bar{x}_{SD}}\,\bar{X}_{SD}\ , \tag{1.3}$$

$$\bar{y}_{SK} = \bar{y}\,\frac{\bar{X}+\beta_2(x)}{\bar{x}+\beta_2(x)} = \frac{\bar{y}}{\bar{x}_{SK}}\,\bar{X}_{SK}\ , \tag{1.4}$$

$$\bar{y}_{US1} = \bar{y}\,\frac{\bar{X}\beta_2(x)+C_x}{\bar{x}\beta_2(x)+C_x} = \frac{\bar{y}}{\bar{x}_{US1}}\,\bar{X}_{US1}\ , \tag{1.5}$$

$$\bar{y}_{US2} = \bar{y}\,\frac{\bar{X}C_x+\beta_2(x)}{\bar{x}C_x+\beta_2(x)} = \frac{\bar{y}}{\bar{x}_{US2}}\,\bar{X}_{US2}\ , \tag{1.6}$$

where

$$\bar{x}_{SD} = \bar{x}+C_x;\ \bar{X}_{SD} = \bar{X}+C_x;\ \bar{x}_{SK} = \bar{x}+\beta_2(x);\ \bar{X}_{SK} = \bar{X}+\beta_2(x);$$
$$\bar{x}_{US1} = \bar{x}\beta_2(x)+C_x;\ \bar{X}_{US1} = \bar{X}\beta_2(x)+C_x;\ \bar{x}_{US2} = \bar{x}C_x+\beta_2(x);\ \bar{X}_{US2} = \bar{X}C_x+\beta_2(x).$$

The MSE equation of these estimators is given by

$$MSE(\bar{y}_k) \cong \frac{1-f}{n}\left(R_k^2 S_x^2 - 2R_k S_{xy} + S_y^2\right)\ ;\ \text{k = SD, SK, US1, US2.} \tag{1.7}$$

where $R_{SD} \doteq \dfrac{\bar{Y}}{\bar{X}_{SD}}$; $R_{SK} \doteq \dfrac{\bar{Y}}{\bar{X}_{SK}}$; $R_{US1} \doteq \dfrac{\bar{Y}\beta_2(x)}{\bar{X}_{US1}}$; $R_{US2} \doteq \dfrac{\bar{Y}C_x}{\bar{X}_{US2}}$.

These estimators are developed assuming that the sample is selected from population with equal probability under simple random sampling, whereas Kadilar and Cingi (2003; 2005) analyze these estimators in the stratified random sampling and Rueda et al. (2006) examine them under various sampling designs. However, none of these studies consider those estimators for the unequal probability design. Therefore, this paper aims to investigate these ratio-type estimators when the unequal probability design is made.

2. THE SUGGESTED ESTIMATORS

For unequal probability design in which ith unit is selected with probability proportional to size, π_i , the generalized ratio estimator can be written as

$$\bar{y}_{rHT} = \frac{\bar{y}_{HT}}{\bar{x}_{HT}}\,\bar{X}\ , \tag{2.1}$$

where $\bar{y}_{HT} = \dfrac{1}{N}\sum_{i=1}^{n}\dfrac{y_i}{\pi_i}$ and $\bar{x}_{HT} = \dfrac{1}{N}\sum_{i=1}^{n}\dfrac{x_i}{\pi_i}$ are Horvitz-Thompson type estimators of the population means of the study and auxiliary variables, respectively (Thompson, 2002).

The MSE of the estimator in (2.1) is

$$MSE(\bar{y}_{rHT}) \cong \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=i}^{N}\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\,y_i^{\dagger}y_j^{\dagger}\ , \tag{2.2}$$

where, π_{ij} is probability that the ith and jth units are both in the sample, $y_i^{\dagger} = y_i - Rx_i$ and $\pi_{jj} = \pi_j$. When the sample is selected without replacement with selection probabilities proportional to size, we can define, $\pi_{ij} = 1 - \frac{\binom{M_T - m_i}{n} + \binom{M_T - m_j}{n} - \binom{M_T - m_i - m_j}{n}}{\binom{M_T}{n}}$ and $\pi_i = 1 - \frac{\binom{M_T - m_i}{n}}{\binom{M_T}{n}}$. Here $M_T = \sum_{i=1}^{N} m_i$ and m_i is the number of units in ith set. Note that the classical ratio estimator is a special case of the generalized ratio estimator under the simple random sampling, in which $\pi_i = \frac{n}{N}$ and $\pi_{ij} = \frac{n(n-1)}{N(N-1)}$ (Thompson, 1990; Thompson and Seber, 1996).

Therefore, we can rewrite the MSE equation of classical estimators, (1.2), as

$$MSE(\bar{y}_r) \cong \frac{1}{N^2}\left(\frac{(n-1)N}{(N-1)n} - 1\right)\left\{\sum_{i=1}^{N}\sum_{j=i}^{N} y_i^{\dagger} y_j^{\dagger}\right\} \tag{2.3}$$

and similarly, we can also re-write (1.7) as

$$MSE(\bar{y}_k) \cong \frac{1}{N^2}\left(\frac{(n-1)N}{(N-1)n} - 1\right)\left\{\sum_{i=1}^{N}\sum_{j=i}^{N} y_{ik}^{*} y_{jk}^{*}\right\} \tag{2.4}$$

where $y_{ik}^{*} = y_i - R_k x_i$; k = SD, SK, US1, US2.

Combining the estimators given in the Section 1 and Horvitz-Thompson ratio type estimator given in (2.1), we propose following estimators:

$$\bar{y}_{prSD} = \frac{\bar{y}_{HT}}{\bar{x}_{SDHT}} \bar{X}_{SD}, \tag{2.5}$$

$$\bar{y}_{prSK} = \frac{\bar{y}_{HT}}{\bar{x}_{SKHT}} \bar{X}_{SK}, \tag{2.6}$$

$$\bar{y}_{prUS1} = \frac{\bar{y}_{HT}}{\bar{x}_{US1HT}} \bar{X}_{US1}, \tag{2.7}$$

$$\bar{y}_{prUS2} = \frac{\bar{y}_{HT}}{\bar{x}_{US2HT}} \bar{X}_{US2}, \tag{2.8}$$

where $\bar{x}_{SDHT} = \bar{x}_{HT} + C_x$; $\bar{x}_{SKHT} = \bar{x}_{HT} + \beta_2(x)$; $\bar{x}_{US1HT} = \bar{x}_{HT}\beta_2(x) + C_x$ and $\bar{x}_{US2HT} = \bar{x}_{HT} C_x + \beta_2(x)$.

Using (2. 2), the MSE of the proposed estimators can be given by

$$MSE\left(\bar{y}_{prk}\right) \cong \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=i}^{N}\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}\,y_{ik}^{*}y_{jk}^{*} \,. \qquad (2.9)$$

where $y_{ik}^{*} = y_i - R_k x_i$; k = SD, SK, US1, US2.

3. EFFICIENCY COMPARISONS

In this section, we compare the MSE of proposed ratio estimators, given in (2.9), with the MSE of classical ratio estimator under unequal design, given in (2.2), as follows:

$$MSE\left(\bar{y}_{prk}\right) < MSE\left(\bar{y}_{rHT}\right); \qquad k = \text{SD, SK, US1, US2.}$$

Let $b_{ij} = \dfrac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}$.

$$\sum_{i=1}^{N}\sum_{j=1}^{N} b_{ij}\left(y_i - R_k x_i\right)\left(y_j - R_k x_j\right) < \sum_{i=1}^{N}\sum_{j=1}^{N} b_{ij}\left(y_i - Rx_i\right)\left(y_j - Rx_j\right),$$

$$\sum_{i=1}^{N}\sum_{j=1}^{N} b_{ij}\left(y_i y_j - R_k y_i x_j - R_k y_j x_i + R_k^2 x_i x_j\right) < \sum_{i=1}^{N}\sum_{j=1}^{N} b_{ij}\left(y_i y_j - Ry_i x_j - Ry_j x_i + R^2 x_i x_j\right),$$

$$\sum_{i=1}^{N}\sum_{j=1}^{N} b_{ij}\left\{\left(R_k - R\right)\left(-y_i x_j - y_j x_i\right) + \left(R_k^2 - R^2\right)x_i x_j\right\} < 0\,,$$

$$R_k + R < \tau\,; \qquad \text{for } R_k > R \qquad (3.1)$$

$$R_k + R > \tau\,; \qquad \text{for } R_k < R \qquad (3.2)$$

where $\tau = \dfrac{\sum_{i=1}^{N}\sum_{j=1}^{N} b_{ij}\left(y_i x_j + y_j x_i\right)}{\sum_{i=1}^{N}\sum_{j=1}^{N} b_{ij} x_i x_j}$. When the condition (3.1) or (3.2) is satisfied, proposed ratio estimators given in (2.5)-(2.8) are more efficient than the classical ratio estimator under unequal sampling design given in (2.1).

We also compare the MSE of proposed ratio estimators with the MSE of classical ratio estimators using the equations (2.9) and (2.4) and we obtain the following condition:

$$MSE\left(\bar{y}_{prk}\right) < MSE\left(\bar{y}_{k}\right)$$

$$\sum_{i=1}^{N}\sum_{j=1}^{N} b_{ij} y_{ik}^{*} y_{jk}^{*} < \delta\sum_{i=1}^{N}\sum_{j=1}^{N} y_{ik}^{*} y_{jk}^{*}\,; \quad k = \text{SD, SK, US1, US2,} \qquad (3.3)$$

where $\delta = \dfrac{n-N}{n(N-1)}$. When the condition (3.3) is satisfied, proposed Sisodia-Dwivedi ratio estimator given in (2.5) is more efficient than the classical Sisodia-Dwivedi ratio estimator given in (1.3). Similarly, when the condition (3.3) is satisfied, proposed Singh-Kakran ratio estimator given in (2.6) is more efficient than the classical Singh-Kakran ratio estimator given in (1.4). These results are also valid for Upadhyaya-Singh estimators.

4. NUMERICAL EXAMPLE AND MAIN RESULTS

We consider a real finite population presented in Cochran (1977) on page 34. This data set concerns food cost (as study variable), weekly income (as auxiliary variable) and size of 32 families (as mi). Note that we omitted the data of 1 family because it was an outlier. By these data, computing the MSE of classical and proposed ratio estimators using (2.2), (2.4) and (2.9), we compare these estimators with each other with respect to their MSE values.

Table-1
Data Statistics

N = 32	$\bar{Y}$ = 27.0625	R = 0.3707
n = 10	$\bar{X}$ = 73.0000	R_{SD} = 0.3700
M_t =119	S_y = 9.9846	R_{SK} = 0.3749
τ = 0.5373	S_x = 10.4140	R_{US1} = 0.3716
δ = -0.0710	$\beta_2(x)$ = -0.8141	R_{US2} = 0.4022

In Table 1, we observe the statistics about the population. We take sample sizes as n = 10, 15, 20 and in Table 2 values of MSE are given for each sample size. From these values, we see that all estimators under unequal probability design have smaller MSE than estimators under equal probability design for all sample sizes. We would like to remark that the condition (3.3) is satisfied for all sample sizes.

Table 2
MSE Values of Ratio Estimators for Various Sample Sizes

	MSE Values	
n = 10	**Equal Probability**	**Unequal Probability**
classical ratio estimator	8.1762	6.9741
Sisodia-Dwivedi	8.1754	6.9508
Singh-Kakran	8.1807	7.1117
Upadhyaya-Singh 1	8.1771	7.0030
Upadhyaya-Singh 2	8.2188	8.1445
n = 15		
classical ratio estimator	5.2096	4.1248
Sisodia-Dwivedi	5.2091	4.1105
Singh-Kakran	5.2125	4.2104
Upadhyaya-Singh 1	5.2102	4.1428
Upadhyaya-Singh 2	5.2368	4.8657
n = 20		
classical ratio estimator	3.7263	2.7264
Sisodia-Dwivedi	3.7260	2.7164
Singh-Kakran	3.7284	2.7863
Upadhyaya-Singh 1	3.7267	2.7389
Upadhyaya-Singh 2	3.7458	3.2535

It is worth to point out that the proposed Sisodia-Dwivedi estimator has a smaller MSE than the classical ratio estimator under unequal probability design for all sample sizes. It is an expected result because (3.2) is satisfied as,

$R = 0.3707$; $R_{SD} = 0.3700$

and $R + R_{SD} > \tau = 0.5373$; for $n = 10$

$R + R_{SD} > \tau = 0.5662$; for $n = 15$

$R + R_{SD} > \tau = 0.5852$; for $n = 20$

for the proposed Sisodia-Dwivedi estimator. However, the condition (3.1) is not satisfied by the other proposed estimators (proposed Singh-Kakran, proposed Upadhyaya-Singh 1 and 2) for this data set so they have bigger MSE values than the classical ratio estimator under unequal probability design.

As expected, MSE values decrease when the sample size increases for all estimators. In addition, we would like to point out that the order of MSE values of the estimators from smaller to bigger are similar in both equal and unequal probability designs.

REFERENCES

1. Cochran, W.G. (1977). *Sampling Techniques*. (3rd ed.), John Wiley and Sons, New York.
2. Kadilar, C. and Cingi, H. (2003). Ratio estimators in stratified random sampling. *Biometrical Journal.* 45, 218-225.
3. Kadilar, C. and Cingi, H. (2005). A new ratio estimator in stratified random sampling, *Communications in Statistics: Theory and Methods.* 34, 597-602.
4. Rueda, M.M., González, S. and Arcos, A. (2006). A general class of estimators with auxiliary information based on available units, *Applied Mathematics and Computation.* 175, 131-148.
5. Singh, H.P. and Kakran, M.S. (1993). *A modified ratio estimator using known coefficient of kurtosis of an auxiliary character*. (unpublished manuscript).
6. Sisodia, B.V.S. and Dwivedi, V.K. (1981). A modified ratio estimator using coefficient of variation of auxiliary variable. *J. Ind. Soc. Agri. Statist.* 33, 13-18.
7. Upadhyaya, L.N. and Singh, H.P. (1999). Use of transformed auxiliary variable in estimating the finite population mean. *Biometrical Journal* 41, 627-636.
8. Thompson, S.K. (1990). Adaptive cluster sampling, *J. Amer. Statist. Assoc.* 85, 1050-1059.
9. Thompson, S.K. and Seber, G.A.F. (1996). *Adaptive Sampling*. John Wiley and Sons, New York.
10. Thompson, S.K. (2002). *Sampling*. (2nd ed.). John Wiley and Sons, New York.