Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER Applied Mathematics and Computation xxx (2005) xxx–xxx

APPLIED MATHEMATICS AND COMPUTATION

www.elsevier.com/locate/amc

Ratio estimators for the population variance in simple and stratified random sampling

Cem Kadilar *, Hulya Cingi

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

Abstract

We propose some ratio-type variance estimators using ratio estimators for the population mean in literature. We obtain mean square error (MSE) equations of proposed estimators and show that proposed estimators are more efficient than the traditional ratio estimator, suggested by [C.T. Isaki, Variance estimation using auxiliary information, Journal of the American Statistical Association 78 (1983) 117–123], under certain conditions. We also adapt the proposed estimators in the simple random sampling to the stratified random sampling. In addition, we support the theoretical results with the aid of numerical examples.

Keywords: Variance estimator; Ratio-type estimator; Auxiliary information; Efficiency; Simple random sampling; Stratified random sampling

* Corresponding author.
E-mail address: kadilar@hacettepe.edu.tr (C. Kadilar).

0096-3003/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.amc.2005.04.032

1. Introduction

Ratio-type estimators take advantage of the correlation between the auxiliary variate, x and the variate of interest, y. When information is available on the auxiliary variate that is positively correlated with the variate of interest, the ratio estimator is a suitable estimator to estimate the population mean. For ratio estimators in sampling literature, population information of the auxiliary variate, such as the coefficient of variation or the kurtosis, is often used to increase the precision of the estimation of a population mean. In this study, we consider this population information to improve the efficiency of the estimation for the population variance in simple and also in stratified random sampling. The problem of constructing efficient estimators for the population variance has been widely discussed by various authors such as Das and Tripathi [5], Isaki [7], Singh et al. [13], Agrawal and Sthapit [1], Garcia and Cebrain [6] and Arcos et al. [2].

Assuming simple random sampling, we present traditional and proposed ratio-type estimators of the population variance and obtain their MSE equations in the next section. Efficiency comparisons, based on these MSE equations, are considered in Section 3. Assuming stratified random sampling, we develop variance estimator and derive its MSE equation in Section 4. We suggest a class of ratio-type estimators for the population variance in stratified random sampling and find their MSE equations in Section 5. Numerical results are reported in Section 6 and discussed in the last section.

2. Suggested estimators in simple random sampling

Isaki [7] presented the ratio estimator for the population variance using the auxiliary information, S_x^2, as

$$s_{\text{ratio}}^2 = \frac{s_y^2}{s_x^2} S_x^2, \tag{1}$$

where s_y^2 and s_x^2 are unbiased estimators of population variances S_y^2 and S_x^2, respectively. The MSE of this estimator is

$$\text{MSE}(s_{\text{ratio}}^2) \cong \lambda S_y^4[\beta_2(y) + \beta_2(x) - 2\theta] \quad [12], \tag{2}$$

where $\lambda = \frac{1}{n}$, $\beta_2(y) = \frac{\mu_{40}}{\mu_{20}^2}$, $\beta_2(x) = \frac{\mu_{04}}{\mu_{02}^2}$, $\theta = \frac{\mu_{22}}{\mu_{20}\mu_{02}}$ and $\mu_{rs} = \frac{1}{N}\sum_{j=1}^{N}(y_j - \overline{Y})^r(x_j - \overline{X})^s$ [11]. Here N is number of units in population, n is the sample size, $\overline{X}$ and $\overline{Y}$ are population means of the auxiliary variate x_j and the variate of interest y_j, respectively.

Motivated by Sisodia and Dwivedi [15], Singh and Kakran [14] and Upadhyaya and Singh [16], we propose following ratio-type estimators for the population variance as

$$s^2_{\mathrm{pr1}} = \frac{s^2_y}{s^2_x - C_x}[S^2_x - C_x], \tag{3}$$

$$s^2_{\mathrm{pr2}} = \frac{s^2_y}{s^2_x - \beta_2(x)}[S^2_x - \beta_2(x)], \tag{4}$$

$$s^2_{\mathrm{pr3}} = \frac{s^2_y}{s^2_x\beta_2(x) - C_x}[S^2_x\beta_2(x) - C_x], \tag{5}$$

$$s^2_{\mathrm{pr4}} = \frac{s^2_y}{s^2_x C_x - \beta_2(x)}[S^2_x C_x - \beta_2(x)], \tag{6}$$

where $C_x = \frac{S_x}{\bar{X}}$ is the coefficient of variation for the population and $\beta_2(x)$ is the population kurtosis of the auxiliary variate.

The MSE of proposed estimators can be found using the first degree approximation in the Taylor series method defined by

$$\mathrm{MSE}(s^2_{\mathrm{pr}}) \cong \boldsymbol{d}\boldsymbol{\Sigma}\boldsymbol{d}', \tag{7}$$

where

$$\boldsymbol{d} = \left[\left.\frac{\partial h(a,b)}{\partial a}\right|_{S^2_y,S^2_x} \quad \left.\frac{\partial h(a,b)}{\partial b}\right|_{S^2_y,S^2_x} \right],$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} V(s^2_y) & \mathrm{cov}(s^2_y, s^2_x) \\ \mathrm{cov}(s^2_x, s^2_y) & V(s^2_x). \end{bmatrix} \quad [17].$$

Here $h(a,b) = h(s^2_y, s^2_x) = s^2_{\mathrm{pr1}}$. According to this definition, we obtain $\boldsymbol{d}$ for the first proposed estimator, s^2_{pr1}, as follows:

$$\boldsymbol{d} = \left[1 \quad -\tfrac{S^2_y}{S^2_x - C_x} \right].$$

We obtain the MSE of the first proposed estimator using (7) as

$$\mathrm{MSE}(s^2_{\mathrm{pr1}}) \cong V(s^2_y) - 2\frac{S^2_y}{S^2_x - C_x}\mathrm{cov}(s^2_y, s^2_x) + \left(\frac{S^2_y}{S^2_x - C_x}\right)^2 V(s^2_x), \tag{8}$$

where

$$\begin{aligned} V(s^2_y) &= \lambda S^4_y[\beta_2(y) - 1], \\ V(s^2_x) &= \lambda S^4_x[\beta_2(x) - 1], \\ \mathrm{cov}(s^2_y, s^2_x) &= \lambda S^2_y S^2_x(\theta - 1) \quad [11]. \end{aligned} \tag{9}$$

Using these definitions, we can write (8) as

$$\mathrm{MSE}(s^2_{\mathrm{pr}1}) \cong \lambda S_y^4\{\beta_2(y) - 1 - 2A_1(\theta - 1) + A_1^2[\beta_2(x) - 1]\}, \tag{10}$$

where $A_1 = \frac{S_x^2}{S_x^2 - C_x}$.

It is clear that A_1 is replaced with

$$A_2 = \frac{S_x^2}{S_x^2 - \beta_2(x)},$$
$$A_3 = \frac{S_x^2\beta_2(x)}{S_x^2\beta_2(x) - C_x},$$
$$A_4 = \frac{S_x^2 C_x}{S_x^2 C_x - \beta_2(x)},$$

for MSE equations of second, third and fourth proposed estimators, given in (4)–(6), respectively. In other words, the MSE equation of proposed estimators is

$$\mathrm{MSE}(s^2_{\mathrm{pr}i}) \cong \lambda S_y^4\{\beta_2(y) - 1 - 2A_i(\theta - 1) + A_i^2[\beta_2(x) - 1]\}, \quad i = 1, 2, 3, 4. \tag{11}$$

3. Efficiency comparisons

We compare the MSE of the proposed estimators, given in (11), with the MSE of the traditional ratio estimator, given in (2). We have the following condition:

$$\begin{aligned} &\mathrm{MSE}(s^2_{\mathrm{pr}i}) < \mathrm{MSE}(s^2_{\mathrm{ratio}}), \quad i = 1, 2, 3, 4, \\ &A_i^2\beta_2(x) - A_i^2 - 2A_i\theta + 2A_i - 1 - \beta_2(x) + 2\theta < 0. \end{aligned} \tag{12}$$

When the condition (12) is satisfied, the proposed estimators are more efficient than the traditional ratio estimator, given in (1).

When we analyze the condition (12), we see that this condition gets the value 0 if $A_i = 1$. This means that there is no difference between the proposed estimator and the traditional estimator for the MSE value. However, we know that $A_i \neq 1$ because $C_x \neq 0$ and $\beta_2\,(x) \neq 0$.

4. Variance estimator in stratified random sampling

In the stratified random sampling, the population variance, for the variate of interest, can be obtained as follows:

$$(N-1)S_{\mathrm{st},y}^2 = \sum_{h=1}^{\ell}\sum_{i=1}^{N_h}(y_{hi}-\overline{Y})^2 = \sum_{h=1}^{\ell}\sum_{i=1}^{N_h}[(y_{hi}-\overline{Y}_h)+(\overline{Y}_h-\overline{Y})]^2, \tag{13}$$

where ℓ is the number of stratum, N_h is the population size in stratum h and $\overline{Y}_h$ is the population mean of the variate of interest in stratum h, y_{hi}. Assuming that $N \cong N-1$ and $N_h \cong N_h - 1$, we can write (13) as

$$\begin{aligned} NS_{\mathrm{st},y}^2 &\cong \sum_{h=1}^{\ell} N_h S_{yh}^2 + \sum_{h=1}^{\ell} N_h(\overline{Y}_h-\overline{Y})^2, \\ S_{\mathrm{st},y}^2 &\cong \sum_{h=1}^{\ell}\omega_h S_{yh}^2 + \sum_{h=1}^{\ell}\omega_h(\overline{Y}_h-\overline{Y})^2, \end{aligned} \tag{14}$$

where $\omega_h = \frac{N_h}{N}$ is the stratum weight and S_{yh}^2 is the population variance of the variate of interest in stratum h. Assuming $n \cong n-1$ and $n_h \cong n_h - 1$, the estimator of the population variance, given in (14), is

$$s_{\mathrm{st},y}^2 = \sum_{h=1}^{\ell}\hat{\omega}_h s_{yh}^2 + \sum_{h=1}^{\ell}\hat{\omega}_h(\bar{y}_h-\bar{y}_{\mathrm{st}})^2, \tag{15}$$

where $\hat{\omega}_h = \frac{n_h}{n}$, s_{yh}^2 and $\bar{y}_h$ are the sample variance and the sample mean of the variate of interest in stratum h, respectively, and $\bar{y}_{\mathrm{st}} = \sum_{h=1}^{\ell}\omega_h\bar{y}_h$ is the estimator of the population mean for the variate of interest in the stratified random sampling [3, p. 103]. We know that $\frac{n_h}{n} = \frac{N_h}{N}$ in proportional allocation [4, p. 91]. Therefore we can take $\hat{\omega}_h = \omega_h$ for the rest of the paper. The MSE of the variance estimator, given in (15), is obtained as

$$\begin{aligned} \mathrm{MSE}(s_{\mathrm{st},y}^2) \cong{} & \sum_{h=1}^{\ell}\omega_h^2\lambda_h S_{yh}^4[\beta_2(y_h)-1] \\ & + 4\sum_{h=1}^{\ell}\omega_h^2(\overline{Y}_h-\overline{Y})\left(\lambda_h\mu_{30h} - \sum_{h=1}^{\ell}\omega_h\lambda_h\mu_{30h}\right) \\ & + 4\sum_{h=1}^{\ell}\omega_h^2(\overline{Y}_h-\overline{Y})^2\left(\lambda_h S_{yh}^2 - 2\sum_{h=1}^{\ell}\omega_h\lambda_h S_{yh}^2 + \sum_{h=1}^{\ell}\omega_h^2\lambda_h S_{yh}^2\right) \end{aligned} \tag{16}$$

(see Appendix A) where $\lambda_h = \frac{1}{n_h}$, $\mu_{rsh} = \frac{1}{N_h}\sum_{i=1}^{N_h}(y_{hi}-\overline{Y}_h)^r(x_{hi}-\overline{X}_h)^s$ and $\beta_2(y_h)$ is the population kurtosis of the variate of interest in stratum h. Here n_h is the sample size in stratum h and $\overline{X}_h$ is the population mean of the auxiliary variate in stratum h, x_{hi}.

5. Suggested estimators in stratified random sampling

Combined ratio estimator for the population mean is defined by

$$\bar{y}_{\mathrm{rc}} = \frac{\bar{y}_{\mathrm{st}}}{\bar{x}_{\mathrm{st}}}\overline{X} \quad [4],$$

where $\bar{x}_{\mathrm{st}} = \sum_{h=1}^{\ell}\omega_h\bar{x}_h$ is the estimator of the population mean of the auxiliary variate in the stratified random sampling. Here $\bar{x}_h$ is the sample mean of the auxiliary variate in stratum h. According to this definition, combined ratio estimator for the population variance can be expressed as

$$s_{\mathrm{rc}}^2 = \frac{s_{\mathrm{st},y}^2}{s_{\mathrm{st},x}^2}S_x^2, \tag{17}$$

where $s_{\mathrm{st},x}^2 = \sum_{h=1}^{\ell}\omega_h s_{xh}^2 + \sum_{h=1}^{\ell}\omega_h(\bar{x}_h - \bar{x}_{\mathrm{st}})^2$ is the estimator of the population variance of the auxiliary variate in the stratified random sampling. Here s_{xh}^2 is the sample variance of the auxiliary variate in stratum h. The MSE of the estimator, given in (17), is found as follows:

$$\begin{aligned}
\mathrm{MSE}(s_{\mathrm{rc}}^2) \cong \frac{S_x^4}{\delta^2}\Bigg\{ & \sum_{h=1}^{\ell}\omega_h^2\lambda_h S_{yh}^4[\beta_2(y_h)-1] + 4\sum_{h=1}^{\ell}\omega_h^2(\overline{Y}_h-\overline{Y})\Bigg(\lambda_h\mu_{30h} \\
& - \sum_{h=1}^{\ell}\omega_h\lambda_h\mu_{30h}\Bigg) + 4\sum_{h=1}^{\ell}\omega_h^2(\overline{Y}_h-\overline{Y})^2\Bigg(\lambda_h S_{yh}^2 - 2\sum_{h=1}^{\ell}\omega_h\lambda_h S_{yh}^2 \\
& + \sum_{h=1}^{\ell}\omega_h^2\lambda_h S_{yh}^2\Bigg) - 2\frac{v}{\delta}\sum_{h=1}^{\ell}\omega_h^2\lambda_h S_{yh}^2 S_{xh}^2(\theta_h-1) - 4\frac{v}{\delta} \\
& \times\sum_{h=1}^{\ell}\omega_h^2(\overline{X}_h-\overline{X})\Bigg(\lambda_h\mu_{21h} - \sum_{h=1}^{\ell}\omega_h\lambda_h\mu_{21h} - \frac{v}{\delta}\lambda_h\mu_{03h} \\
& + \frac{v}{\delta}\sum_{h=1}^{\ell}\omega_h\lambda_h\mu_{03h}\Bigg) - 4\frac{v}{\delta}\sum_{h=1}^{\ell}\omega_h^2(\overline{Y}_h-\overline{Y})\Bigg(\lambda_h\mu_{12h} \\
& - \sum_{h=1}^{\ell}\omega_h\lambda_h\mu_{12h}\Bigg) + \frac{v^2}{\delta^2}\sum_{h=1}^{\ell}\omega_h^2\lambda_h S_{xh}^4[\beta_2(x_h)-1] - 8\frac{v}{\delta}\sum_{h=1}^{\ell}\omega_h^2 \\
& \times(\overline{X}_h-\overline{X})(\overline{Y}_h-\overline{Y})\Bigg(\lambda_h S_{yxh} - 2\sum_{h=1}^{\ell}\omega_h\lambda_h S_{yxh} + \sum_{h=1}^{\ell}\omega_h^2\lambda_h S_{yxh}\Bigg) \\
& + 4\frac{v^2}{\delta^2}\sum_{h=1}^{\ell}\omega_h^2(\overline{X}_h-\overline{X})^2\Bigg(\lambda_h S_{xh}^2 - 2\sum_{h=1}^{\ell}\omega_h\lambda_h S_{xh}^2 + \sum_{h=1}^{\ell}\omega_h^2\lambda_h S_{xh}^2\Bigg)\Bigg\}
\end{aligned} \tag{18}$$

(see Appendix B) where $\theta_h = \frac{\mu_{22h}}{\mu_{20h}\mu_{02h}}$, $v = \sum_{h=1}^{\ell}\omega_h S_{yh}^2 + \sum_{h=1}^{\ell}\omega_h(\overline{Y}_h - \overline{Y})$ and $\delta = \sum_{h=1}^{\ell}\omega_h S_{xh}^2 + \sum_{h=1}^{\ell}\omega_h(\overline{X}_h - \overline{X})$. Here $\beta_2(x_h)$ is the population kurtosis of the auxiliary variate in stratum h, S_{xh}^2 is the population variance of the auxiliary variate in stratum h and S_{yxh} is the population covariance between the auxiliary variate and variate of interest in stratum h.

Motivated by Kadilar and Cingi [8], we propose following variance estimators in the stratified sampling:

$$s_{\mathrm{rcpr1}}^2 = \frac{s_{\mathrm{st},y}^2}{s_{\mathrm{st},x}^2 + C_x}(S_x^2 + C_x), \tag{19}$$

$$s_{\mathrm{rcpr2}}^2 = \frac{s_{\mathrm{st},y}^2}{s_{\mathrm{st},x}^2 + \beta_2(x)}[S_x^2 + \beta_2(x)], \tag{20}$$

$$s_{\mathrm{rcpr3}}^2 = \frac{s_{\mathrm{st},y}^2}{s_{\mathrm{st},x}^2\beta_2(x) + C_x}[S_x^2\beta_2(x) + C_x], \tag{21}$$

$$s_{\mathrm{rcpr4}}^2 = \frac{s_{\mathrm{st},y}^2}{s_{\mathrm{st},x}^2 C_x + \beta_2(x)}[S_x^2 C_x + \beta_2(x)] \tag{22}$$

whose MSE equations have the same form with (18). Naturally, δ and S_x^2 should be added by C_x for the MSE of the first proposed estimator, and $\beta_2(x)$ for the MSE of the second proposed estimator in (18). Similarly, δ and S_x^2 should be replaced by $\delta\beta_2(x) + C_x$ and $S_x^2\beta_2(x) + C_x$, respectively, and v should be multiplied by $\beta_2(x)$ for the MSE of the third proposed estimator in (18). It is obvious that $\beta_2(x)$ will be replaced by C_x, and vice versa, in the MSE equation of the third proposed estimator for the MSE of the fourth proposed estimator.

6. Numerical examples

We use data in Kadilar and Cingi [9,8] to compare efficiencies between the traditional and proposed estimators in the simple random sampling and in the stratified random sampling, respectively. These data concern the level of apple production (1 unit = 100 tonnes) as the variate of interest and number of apple trees (1 unit = 100 trees) as the auxiliary variate in 106 villages in the Marmarian Region and in 854 villages in 6 strata, respectively (as 1: Marmarian, 2: Agean, 3: Mediterranean, 4: Central Anatolia, 5: Black Sea, 6: East and Southeast Anatolia) in 1999 (Source: Institute of Statistics, Republic of Turkey).

Table 1
Data statistics of the population for the simple random sampling

$N = 106$
$n = 20$
$\rho = 0.82$
$C_y = 4.18$
$C_x = 2.02$
$\overline{Y} = 15.37$
$\overline{X} = 243.76$
$S_y = 64.25$
$S_x = 491.89$
$\beta_2(x) = 25.71$
$\beta_2(y) = 80.13$
$\lambda = 0.05$
$\theta = 33.30$

6.1. Numerical example for simple random sampling

In Table 1, we observe the statistics about the population. Using simple random sampling, we take the sample size as $n = 20$. We would like to remind that sample size has no effect on the efficiency comparisons of the estimators, as shown in Section 3. Note that the correlation (ρ) between the auxiliary variate and variate of interest is 0.82 for this data set.

We compute the MSE values of traditional, simple and proposed estimators using the Eqs. (2), (9) and (11), respectively, and these values are shown in Table 2. We observe that the traditional ratio estimator has the biggest MSE value, except the simple estimator, so we can say that all proposed estimators are more efficient than the traditional ratio estimator for this data set. However, this result is an expected result because the condition (12) is satisfied for all proposed estimators, having the value of ($-1.27\text{E}-04$) for the first proposed estimator, ($-1.61\text{E}-03$) for the second proposed estimator, ($-4.93\text{E}-06$) for the third proposed estimator and ($-7.99\text{E}-04$) for the fourth proposed

Table 2
MSE Values of variance estimators in the simple random sampling

Estimators	MSE
Simple (s_y^2)	67,423,045.30
Traditional (s_{ratio}^2)	33,431,595.28
Proposed 1 (s_{pr1}^2)	33,431,487.38
Proposed 2 (s_{pr2}^2)	33,430,220.74
Proposed 3 (s_{pr3}^2)	33,431,591.08
Proposed 4 (s_{pr4}^2)	33,430,914.10

Table 3
Data statistics of the population for the stratified random sampling

$N = 854$	$N_1 = 106$	$N_2 = 106$	$N_3 = 94$	$N_4 = 171$	$N_5 = 204$	$N_6 = 173$
$n = 140$	$n_1 = 9$	$n_2 = 17$	$n_3 = 38$	$n_4 = 67$	$n_5 = 7$	$n_6 = 2$
$\overline{X} = 376.00$	$\overline{X}_1 = 243.76$	$\overline{X}_2 = 274.22$	$\overline{X}_3 = 724.10$	$\overline{X}_4 = 743.65$	$\overline{X}_5 = 264.42$	$\overline{X}_6 = 98.44$
$\overline{Y} = 29.30$	$\overline{Y}_1 = 15.37$	$\overline{Y}_2 = 22.13$	$\overline{Y}_3 = 93.84$	$\overline{Y}_4 = 55.88$	$\overline{Y}_5 = 9.67$	$\overline{Y}_6 = 4.04$
$\beta_2(x) = 312.07$	$\beta_2(x_1) = 25.71$	$\beta_2(x_2) = 34.57$	$\beta_2(x_3) = 26.14$	$\beta_2(x_4) = 97.60$	$\beta_2(x_5) = 27.47$	$\beta_2(x_6) = 28.10$
$\beta_2(y) = 195.84$	$\beta_2(y_1) = 80.13$	$\beta_2(y_2) = 97.68$	$\beta_2(y_3) = 24.14$	$\beta_2(y_4) = 101.97$	$\beta_2(y_5) = 53.38$	$\beta_2(y_6) = 27.96$
$C_x = 3.85$	$C_{x1} = 2.02$	$C_{x2} = 2.10$	$C_{x3} = 2.22$	$C_{x4} = 3.84$	$C_{x5} = 1.72$	$C_{x6} = 1.91$
$C_y = 5.84$	$C_{y1} = 4.18$	$C_{y2} = 5.22$	$C_{y3} = 3.19$	$C_{y4} = 5.13$	$C_{y5} = 2.47$	$C_{y6} = 2.34$
$S_x = 1447.94$	$S_{x1} = 491.89$	$S_{x2} = 574.61$	$S_{x3} = 1607.57$	$S_{x4} = 2856.03$	$S_{x5} = 454.03$	$S_{x6} = 187.94$
$S_y = 171.06$	$S_{y1} = 64.25$	$S_{y2} = 115.52$	$S_{y3} = 299.07$	$S_{y4} = 286.43$	$S_{y5} = 23.90$	$S_{y6} = 9.46$
$\rho = 0.92$	$\rho_1 = 0.82$	$\rho_2 = 0.86$	$\rho_3 = 0.90$	$\rho_4 = 0.99$	$\rho_5 = 0.71$	$\rho_6 = 0.89$
$\delta = 2{,}107{,}597.26$	$\omega_1 = 0.12$	$\omega_2 = 0.12$	$\omega_3 = 0.11$	$\omega_4 = 0.20$	$\omega_5 = 0.24$	$\omega_6 = 0.20$
$v = 29{,}448.38$	$\lambda_1 = 0.11$	$\lambda_2 = 0.06$	$\lambda_3 = 0.03$	$\lambda_4 = 0.01$	$\lambda_5 = 0.14$	$\lambda_6 = 0.50$
	$\theta_1 = 33.30$	$\theta_2 = 57.40$	$\theta_3 = 20.80$	$\theta_4 = 99.53$	$\theta_5 = 21.09$	$\theta_6 = 23.08$

Table 4
MSE values of variance estimators in the stratified random sampling

Estimators	MSE
Simple (s_y^2)	2.40E+13
Stratified ($s_{\mathrm{st},y}^2$)	346,392,191.66
Combined ratio (s_{rc}^2)	104,956,935.47
Proposed 1 (s_{rcpr1}^2)	104,956,217.62
Proposed 2 (s_{rcpr2}^2)	104,898,788.62
Proposed 3 (s_{rcpr3}^2)	104,956,933.17
Proposed 4 (s_{rcpr4}^2)	104,941,830.88

estimator. It is worth to point out that we obtain $A_i > 1$; $i = 1,2,3,4$ for this data set.

6.2. Numerical example for stratified random sampling

In Table 3, we observe the statistics about the population. Using Neyman allocation in the stratified random sampling, we obtain the sample size for each stratum, n_h ($h = 1,2,\ldots,6$), as shown in Table 3 (see [8] for details).

We compute the MSE values of simple (s_y^2), traditional ($s_{\mathrm{st},y}^2$) and proposed ($s_{\mathrm{rc}}^2, s_{\mathrm{rcpr1}}^2, s_{\mathrm{rcpr2}}^2, s_{\mathrm{rcpr3}}^2, s_{\mathrm{rcpr4}}^2$) estimators using the Eqs. (9), (16) and (18), respectively, and these values are shown in Table 4. We observe that the simple estimator has an extremely big MSE value and all proposed estimators have a smaller MSE value than the traditional stratified ratio estimator for this data set.

7. Discussion and conclusion

From theoretical deductions in Section 3 and results of the numerical example, we inference that the proposed estimators are more efficient than the traditional ratio estimator when $A_i > 1$ in simple random sampling. In addition, we should denote that A_2 is never smaller than 1 because $\beta_2(x)$ is always positive. Therefore, the second proposed estimator is more efficient than the traditional estimator in all conditions. Besides, all proposed estimators in the stratified random sampling have also been found more efficient than the traditional estimator in an application where the stratified random sampling is used. In forthcoming studies, we hope to improve the proposed estimators, as in Kadilar and Cingi [10].

Appendix A

The MSE of the variance estimator in the stratified random sampling can be obtained using the first degree approximation in the Taylor series method defined by

$$\mathrm{MSE}(s^2_{\mathrm{st},y}) \cong \sum_{h=1}^{\ell} \boldsymbol{d}_h \boldsymbol{\Sigma}_h \boldsymbol{d}'_h, \tag{A.1}$$

where

$$\boldsymbol{d}_h = \left[\left.\frac{\partial h(a,b,c)}{\partial a}\right|_{S^2_{yh},\overline{Y}_h,\overline{Y}} \quad \left.\frac{\partial h(a,b,c)}{\partial b}\right|_{S^2_{yh},\overline{Y}_h,\overline{Y}} \quad \left.\frac{\partial h(a,b,c)}{\partial c}\right|_{S^2_{yh},\overline{Y}_h,\overline{Y}} \right],$$

$$\boldsymbol{\Sigma}_h = \begin{bmatrix} V(s^2_{yh}) & \mathrm{cov}(s^2_{yh},\bar{y}_h) & \mathrm{cov}(s^2_{yh},\bar{y}_{\mathrm{st}}) \\ \mathrm{cov}(\bar{y}_h,s^2_{yh}) & V(\bar{y}_h) & \mathrm{cov}(\bar{y}_h,\bar{y}_{\mathrm{st}}) \\ \mathrm{cov}(\bar{y}_{\mathrm{st}},s^2_{yh}) & \mathrm{cov}(\bar{y}_{\mathrm{st}},\bar{y}_h) & V(\bar{y}_{\mathrm{st}}) \end{bmatrix} \quad [17].$$

Here $h(a,b,c) = h(s^2_{yh},\bar{y}_h,\bar{y}_{\mathrm{st}}) = s^2_{\mathrm{st},y}$. Note that $\overline{Y}_{\mathrm{st}} = \sum_{h=1}^{\ell}\omega_h\overline{Y}_h = \overline{Y}$. According to this definition, we obtain d_h for the estimator, $s^2_{\mathrm{st},y}$ as follows:

$$\boldsymbol{d}_h = \left[\omega_h \quad 2\omega_h(\overline{Y}_h - \overline{Y}) \quad -2\omega_h(\overline{Y}_h - \overline{Y}) \right].$$

We obtain the MSE of this estimator using (A.1) as

$$\begin{aligned}\mathrm{MSE}(s^2_{\mathrm{st},y}) \cong{}& \sum_{h=1}^{\ell}\omega_h^2 V(s^2_{yh}) + 4\sum_{h=1}^{\ell}\omega_h^2(\overline{Y}_h - \overline{Y})[\mathrm{cov}(\bar{y}_h,s^2_{yh}) \\ & - \mathrm{cov}(\bar{y}_{\mathrm{st}},s^2_{yh})] + 4\sum_{h=1}^{\ell}\omega_h^2(\overline{Y}_h - \overline{Y})^2[V(\bar{y}_h) \\ & - 2\mathrm{cov}(\bar{y}_h,\bar{y}_{\mathrm{st}}) + V(\bar{y}_{\mathrm{st}})].\end{aligned} \tag{A.2}$$

Appendix B

The MSE of the combined ratio estimator for the population variance can be obtained using the first degree approximation in the Taylor series method as follows:

$$\boldsymbol{d}_h = \frac{S^2_x}{\delta}\begin{bmatrix} \omega_h & 2\omega_h(\overline{Y}_h - \overline{Y}) & -2\omega_h(\overline{Y}_h - \overline{Y}) & \frac{-\omega_h v}{\delta} \\ & \frac{-2\omega_h(\overline{X}_h - \overline{X})v}{\delta} & \frac{2\omega_h(\overline{X}_h - \overline{X})v}{\delta} & \\ & & & \end{bmatrix}$$

and

$$\boldsymbol{\Sigma}_h = \begin{bmatrix} V(s^2_{yh}) & \mathrm{cov}(s^2_{yh},\bar{y}_h) & \mathrm{cov}(s^2_{yh},\bar{y}_{\mathrm{st}}) & \mathrm{cov}(s^2_{yh},s^2_{xh}) & \mathrm{cov}(s^2_{yh},\bar{x}_h) & \mathrm{cov}(s^2_{yh},\bar{x}_{\mathrm{st}}) \\ \mathrm{cov}(\bar{y}_h,s^2_{yh}) & V(\bar{y}_h) & \mathrm{cov}(\bar{y}_h,\bar{y}_{\mathrm{st}}) & \mathrm{cov}(\bar{y}_h,s^2_{xh}) & \mathrm{cov}(\bar{y}_h,\bar{x}_h) & \mathrm{cov}(\bar{y}_h,\bar{x}_{\mathrm{st}}) \\ \mathrm{cov}(\bar{y}_{\mathrm{st}},s^2_{yh}) & \mathrm{cov}(\bar{y}_{\mathrm{st}},\bar{y}_h) & V(\bar{y}_{\mathrm{st}}) & \mathrm{cov}(\bar{y}_{\mathrm{st}},s^2_{xh}) & \mathrm{cov}(\bar{y}_{\mathrm{st}},\bar{x}_h) & \mathrm{cov}(\bar{y}_{\mathrm{st}},\bar{x}_{\mathrm{st}}) \\ \mathrm{cov}(s^2_{xh},s^2_{yh}) & \mathrm{cov}(s^2_{xh},\bar{y}_h) & \mathrm{cov}(s^2_{xh},\bar{y}_{\mathrm{st}}) & V(s^2_{xh}) & \mathrm{cov}(s^2_{xh},\bar{x}_h) & \mathrm{cov}(s^2_{xh},\bar{x}_{\mathrm{st}}) \\ \mathrm{cov}(\bar{x}_h,s^2_{yh}) & \mathrm{cov}(\bar{x}_h,\bar{y}_h) & \mathrm{cov}(\bar{x}_h,\bar{y}_{\mathrm{st}}) & \mathrm{cov}(\bar{x}_h,s^2_{xh}) & V(\bar{x}_h) & \mathrm{cov}(\bar{x}_h,\bar{x}_{\mathrm{st}}) \\ \mathrm{cov}(\bar{x}_{\mathrm{st}},s^2_{yh}) & \mathrm{cov}(\bar{x}_{\mathrm{st}},\bar{y}_h) & \mathrm{cov}(\bar{x}_{\mathrm{st}},\bar{y}_{\mathrm{st}}) & \mathrm{cov}(\bar{x}_{\mathrm{st}},s^2_{xh}) & \mathrm{cov}(\bar{x}_{\mathrm{st}},\bar{x}_h) & V(\bar{x}_{\mathrm{st}}) \end{bmatrix}.$$

Using (A.1),

$$\begin{aligned}
\mathrm{MSE}(s^2_{\mathrm{rc}}) \cong \frac{S_x^4}{\delta^2}\Bigg\{ & \sum_{h=1}^{\ell}\omega_h^2 V(s^2_{yh}) + 4\sum_{h=1}^{\ell}\omega_h^2(\overline{Y}_h-\overline{Y})\left[\mathrm{cov}(\bar{y}_h,s^2_{yh})-\mathrm{cov}(\bar{y}_{\mathrm{st}},s^2_{yh})\right] \\
& +4\sum_{h=1}^{\ell}\omega_h^2(\overline{Y}_h-\overline{Y})^2[V(\bar{y}_h)-2\mathrm{cov}(\bar{y}_{\mathrm{st}},\bar{y}_h)+V(\bar{y}_{\mathrm{st}})] \\
& -2\frac{v}{\delta}\sum_{h=1}^{\ell}\omega_h^2\mathrm{cov}(s^2_{yh},s^2_{xh}) - 4\frac{v}{\delta}\sum_{h=1}^{\ell}\omega_h^2(\overline{X}_h-\overline{X})\Big[\mathrm{cov}(\bar{x}_h,s^2_{yh}) \\
& -\mathrm{cov}(\bar{x}_{\mathrm{st}},s^2_{yh}) - \frac{v}{\delta}\mathrm{cov}\left(\bar{x}_h,s^2_{xh}\right)+\frac{v}{\delta}\mathrm{cov}\left(\bar{x}_{\mathrm{st}},s^2_{xh}\right)\Big] \\
& -4\frac{v}{\delta}\sum_{h=1}^{\ell}\omega_h^2(\overline{Y}_h-\overline{Y})\left[\mathrm{cov}\left(\bar{y}_h,s^2_{xh}\right)-\mathrm{cov}\left(\bar{y}_{\mathrm{st}},s^2_{xh}\right)\right] \\
& +\frac{v^2}{\delta^2}\sum_{h=1}^{\ell}\omega_h^2 V(s^2_{xh}) - 8\frac{v}{\delta}\sum_{h=1}^{\ell}\omega_h^2(\overline{X}_h-\overline{X})(\overline{Y}_h-\overline{Y})[\mathrm{cov}(\bar{y}_h,\bar{x}_h) \\
& -\mathrm{cov}(\bar{y}_{\mathrm{st}},\bar{x}_h)-\mathrm{cov}(\bar{y}_h,\bar{x}_{\mathrm{st}})+\mathrm{cov}(\bar{y}_{\mathrm{st}},\bar{x}_{\mathrm{st}})] \\
& +4\frac{v^2}{\delta^2}\sum_{h=1}^{\ell}\omega_h^2\left(\overline{X}_h-\overline{X}\right)^2[V(\bar{x}_h)-2\mathrm{cov}(\bar{x}_{\mathrm{st}},\bar{x}_h)+V(\bar{x}_{\mathrm{st}})]\Bigg\},
\end{aligned}$$

where

$$V\left(s^2_{yh}\right) = \lambda_h S^4_{yh}[\beta_2(y_h)-1], \quad \mathrm{cov}\left(\bar{y}_h,s^2_{yh}\right) = \lambda_h\mu_{30h},$$

$$\mathrm{cov}\left(\bar{y}_{\mathrm{st}},s^2_{yh}\right) = \sum_{h=1}^{\ell}\omega_h\lambda_h\mu_{30h}, \quad V(\bar{y}_h) = \lambda_h S^2_{yh},$$

$$\mathrm{cov}(\bar{y}_{\mathrm{st}},\bar{y}_h) = \sum_{h=1}^{\ell}\omega_h\lambda_h S^2_{yh}, \quad V(\bar{y}_{\mathrm{st}}) = \sum_{h=1}^{\ell}\omega_h^2\lambda_h S^2_{yh},$$

$$\mathrm{cov}\left(s^2_{yh},s^2_{xh}\right) = \lambda_h S^2_{yh}S^2_{xh}(\theta_h-1), \quad \mathrm{cov}\left(\bar{x}_h,s^2_{yh}\right) = \lambda_h\mu_{21h},$$

$$\mathrm{cov}\left(\bar{x}_{\mathrm{st}},s^2_{yh}\right) = \sum_{h=1}^{\ell}\omega_h\lambda_h\mu_{21h}, \quad \mathrm{cov}\left(\bar{x}_h,s^2_{xh}\right) = \lambda_h\mu_{03h},$$

$$\operatorname{cov}(\bar{x}_{\mathrm{st}}, s_{xh}^2) = \sum_{h=1}^{\ell} \omega_h \lambda_h \mu_{03h}, \quad \operatorname{cov}\left(\bar{y}_h, s_{xh}^2\right) = \lambda_h \mu_{12h},$$

$$\operatorname{cov}\left(\bar{y}_{\mathrm{st}}, s_{xh}^2\right) = \sum_{h=1}^{\ell} \omega_h \lambda_h \mu_{12h}, \quad V\left(s_{xh}^2\right) = \lambda_h S_{xh}^4 [\beta_2(x_h) - 1],$$

$$\operatorname{cov}(\bar{y}_h, \bar{x}_h) = \lambda_h S_{yxh}, \quad \operatorname{cov}(\bar{y}_{\mathrm{st}}, \bar{x}_h) = \operatorname{cov}(\bar{y}_h, \bar{x}_{\mathrm{st}}) = \sum_{h=1}^{\ell} \omega_h \lambda_h S_{yxh},$$

$$\operatorname{cov}(\bar{y}_{\mathrm{st}}, \bar{x}_{\mathrm{st}}) = \sum_{h=1}^{\ell} \omega_h^2 \lambda_h S_{yxh}, \quad V(\bar{x}_h) = \lambda_h S_{xh}^2,$$

$$\operatorname{cov}(\bar{x}_{\mathrm{st}}, \bar{x}_h) = \sum_{h=1}^{\ell} \omega_h \lambda_h S_{xh}^2, \quad V(\bar{x}_{\mathrm{st}}) = \sum_{h=1}^{\ell} \omega_h^2 \lambda_h S_{xh}^2.$$

References

[1] M.C. Agrawal, A.B. Sthapit, Unbiased ratio-type variance estimation, Statistics and Probability Letters 25 (1995) 361–364.

[2] A. Arcos, M. Rueda, M.D. Martinez, S. Gonzalez, Y. Roman, Incorporating the auxiliary information available in variance estimation, Applied Mathematics and Computation 160 (2005) 387–399.

[3] H. Cingi, Sampling Theory, Hacettepe University Press, 1994 (in Turkish).

[4] W.G. Cochran, Sampling Techniques, John Wiley and Sons, New York, 1977.

[5] A.K. Das, T.P. Tripathi, Use of auxiliary information in estimating the finite population variance, Sankhya 40 (1978) 139–148.

[6] M.R. Garcia, A.A. Cebrain, Variance estimation using auxiliary information: An almost unbiased multivariate ratio estimator, Metrika 45 (1997) 171–178.

[7] C.T. Isaki, Variance estimation using auxiliary information, Journal of the American Statistical Association 78 (1983) 117–123.

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

[9] C. Kadilar, H. Cingi, A study on the chain ratio-type estimator, Hacettepe Journal of Mathematics and Statistics 32 (2003) 105–108.

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

[11] M. Kendall, A. Stuart, The Advanced Theory of Statistics: Distribution Theory, vol. 1, Griffin, London, 1963.

[12] B. Prasad, H.P. Singh, Some improved ratio-type estimators of finite population variance in sample surveys, Communications in Statistics: Theory and Methods 19 (1990) 1127–1139.

[13] H.P. Singh, L.N. Upadhyaya, U.D. Namjoshi, Estimation of finite population variance, Current Science 57 (1988) 1331–1334.

[14] H.P. Singh, M.S. Kakran, A modified ratio estimator using known coefficient of kurtosis of an auxiliary character, 1993, unpublished..

[15] B.V.S. Sisodia, V.K. Dwivedi, A modified ratio estimator using coefficient of variation of auxiliary variable, Journal of Indian Society Agricultural Statistics 33 (1981) 13–18.

[16] L.N. Upadhyaya, H.P. Singh, Use of transformed auxiliary variable in estimating the finite population mean, Biometrical Journal 41 (1999) 627–636.

[17] K.M. Wolter, Introduction to Variance Estimation, Springer-Verlag, 1985.