Hacettepe Journal of Mathematics and Statistics
Volume 37 (1) (2008), 57 – 67

VARIANCE ESTIMATION IN SIMPLE RANDOM SAMPLING USING AUXILIARY INFORMATION

Sat Gupta* and Javid Shabbir†

Received 09 : 04 : 2008 : Accepted 04 : 06 : 2008

Abstract

In this paper we focus on investigating the precision of several variance estimators considered by different authors, such as those suggested by Isaki (*Variance estimation using auxiliary information*, J. Amer. Stat. Assoc. **78**, 117–123, 1983), and Kadilar and Cingi (*Ratio estimators for population variance in simple and stratified sampling*, Appl. Math. Comp. **173**, 1047–1058 and *Improvement in variance estimation using auxiliary information*, Hacet. J. Math. Stat. **35** (1), 111–115, 2006). We propose a new hybrid class of estimators and show that in some cases their efficiency is better than the aforementioned traditional ratio and regression estimators of Isaki, and Kadilar and Cingi. Four numerical examples are considered to further evaluate the performance of these estimators.

Keywords: Auxiliary variable, Bias, Mean square error, Efficiency.

2000 AMS Classification: 62 D 05.

1. Introduction

Consider a finite population $U = \{U_1, U_2, \ldots, U_i, \ldots, U_N\}$ consisting of N units. Let y and x be the study and the auxiliary variables with population means $\overline{Y}$ and $\overline{X}$ respectively. Let there be a sample of size n drawn from this population using simple random sampling without replacement (SRSWOR). Let $s_y^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2/(n-1)$ and $s_x^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2/(n-1)$ be the sample variances and $S_y^2 = \sum_{i=1}^{N}(y_i - \overline{Y})^2/(N-1)$ and $S_x^2 = \sum_{i=1}^{N}(x_i - \overline{X})^2/(N-1)$ the population variances of y and x respectively. Let $C_y = S_y/\overline{Y}$, and $C_x = S_x/\overline{X}$ be the coefficients of variation of y and x respectively, and

*Department of Mathematics and Statistics, University of North Carolina at Greensboro, Greensboro, NC 27412, USA. E-mail sngupta@uncg.edu

†Department of Statistics, Quaid-i-Azam University, Islamabad 45320, Pakistan.
E-mail jsqau@yahoo.com

ρ_{yx} the coefficient of correlation between y and x. We assume that all parameters of x are known.

Let $\delta_0 = (s_y^2 - S_y^2)/S_y^2$ and $\delta_1 = (s_x^2 - S_x^2)/S_x^2$ be such that $E(\delta_i) = 0$, $(i = 0, 1)$. Also, to the first order of approximation, we have $E(\delta_0^2) = \gamma(\beta_{2(y)} - 1)$, $\mathrm{E}(\delta_1^2) = \gamma(\beta_{2(x)} - 1)$ and $\mathrm{E}(\delta_0\delta_1) = \gamma(\lambda_{22} - 1)$, where $\gamma = \left(\frac{1}{n} - \frac{1}{N}\right)$, $\lambda_{rs} = \mu_{rs}/(\mu_{20}^{r/2}\mu_{02}^{s/2})$, $\mu_{rs} = \sum_{i=1}^{N}(y_i - \overline{Y})^r(x_i - \overline{X})^s/N$, and $\beta_{2(y)} = \mu_{40}/\mu_{20}^2$, $\beta_{2(x)} = \mu_{04}/\mu_{02}^2$ are the coefficients of kurtosis of y and x respectively.

Recall that the variance of the usual variance estimator $\widehat{S}_y^2 (= s_y^2)$ is given by

$$(1) \qquad \mathrm{Var}(\widehat{S}_y^2) = \gamma S_y^4(\beta_{2(y)} - 1).$$

Now we discuss some other variance estimators that are available in the literature.

(i) *The Isaki Estimators*

Isaki [4] suggested the following ratio and regression estimators. The ratio estimator is given by

$$(2) \qquad \widehat{S}_R^2 = s_y^2\left(\frac{S_x^2}{s_x^2}\right).$$

The bias and MSE of this estimator, to the first order of approximation, are given by

$$(3) \qquad \mathrm{Bias}(\widehat{S}_R^2) \cong \gamma S_y^2[(\beta_{2(x)} - 1) - (\lambda_{22} - 1)]$$

and

$$(4) \qquad \mathrm{MSE}(\widehat{S}_R^2) \cong \gamma S_y^4[(\beta_{2(y)} - 1) + (\beta_{2(x)} - 1) - 2(\lambda_{22} - 1)]$$

respectively.

Isaki [4] also discussed a regression estimator given by

$$(5) \qquad \widehat{S}_{Reg}^2 = s_y^2 + b^*(S_x^2 - s_x^2),$$

where b^* is the sample regression coefficient between s_y^2 and s_x^2. The corresponding population regression coefficient is given by $\beta = \frac{S_y^2(\lambda_{22}-1)}{S_x^2(\beta_{2(x)}-1)}$ (see Garcia and Cebrian [3]). To the first order of approximation, the estimator $\widehat{S}_{Reg}^2$ is unbiased and its variance is given by

$$(6) \qquad \mathrm{Var}(\widehat{S}_{Reg}^2) = \gamma S_y^4(\beta_{2(y)} - 1)(1 - \rho^2_{(s_y^2, s_x^2)}) = \mathrm{Var}(\widehat{S}_y^2)\left(1 - \rho^2_{(s_y^2, s_x^2)}\right),$$

where $\rho_{(s_y^2, s_x^2)} = (\lambda_{22} - 1)/[\sqrt{(\beta_{2(y)} - 1)}\sqrt{(\beta_{2(x)} - 1)}]$.

(ii) *The Kadilar and Cingi Estimators*

Kadilar and Cingi [5] suggested the following variance estimators by using various transformations.

(i) $\widehat{S}_{KC1}^2 = s_y^2\left(\frac{S_x^2 + C_x}{s_x^2 + C_x}\right)$,

(ii) $\widehat{S}_{KC2}^2 = s_y^2\left(\frac{S_x^2 + \beta_{2(x)}}{s_x^2 + \beta_{2(x)}}\right)$,

(iii) $\widehat{S}_{KC3}^2 = s_y^2\left(\frac{S_x^2\beta_{2(x)} + C_x}{s_x^2\beta_{2(x)} + C_x}\right)$,

(iv) $\widehat{S}_{KC4}^2 = s_y^2\left(\frac{S_x^2 C_x + \beta_{2(x)}}{s_x^2 C_x + \beta_{2(x)}}\right)$.

The bias and MSE of $\widehat{S}^2_{KCi}$ $(i = 1, 2, 3, 4)$, to the first order of approximation, are given by

(7) $$\text{Bias}(\widehat{S}^2_{KCi}) \cong \gamma S_y^2 \psi_i \left[\psi_i(\beta_{2(x)} - 1) - (\lambda_{22} - 1)\right]$$

and

(8) $$\text{MSE}(\widehat{S}_{KCi}) \cong \gamma S_y^4 \left[(\beta_{2(y)} - 1) + \psi_i^2(\beta_{2(x)} - 1) - 2\psi_i(\lambda_{22} - 1)\right],$$

where $\psi_1 = \frac{S_x^2}{S_x^2 + C_x}$, $\psi_2 = \frac{S_x^2}{S_x^2 + \beta_{2(x)}}$, $\psi_3 = \frac{S_x^2 \beta_{2(x)}}{S_x^2 \beta_{2(x)} + C_x}$, and $\psi_4 = \frac{S_x^2 C_x}{S_x^2 C_x + \beta_{2(x)}}$.

Kadilar and Cingi [6] also presented the following estimator

(9) $$\widehat{S}^2_{KC} = \alpha_1 s_y^2 + \alpha_2 \left(s_y^2 \frac{S_x^2}{s_x^2}\right) \tau,$$

where $\alpha_1 + \alpha_2 = 1$ and $\tau = \frac{1 + \gamma C_{yx}}{1 + \gamma C_x^2}$ is a constant (see Shabbir and Yaab [9]). Kadilar and Cingi report that the MSE of $\widehat{S}^2_{KC}$, to the first order of approximation, is given by

(10) $$\text{MSE}(\widehat{S}_{KC})_{opt} \cong \gamma S_y^4 \left[z^2(\beta_{2(y)} - 1) + \alpha_2^{*2}\tau^2(\beta_{2(x)} - 1) - 2\tau z \alpha_2^*(\lambda_{22} - 1)\right],$$

where $z = \alpha_1^* + \alpha_2^* \tau$,

$$\alpha_1^* = \frac{(\beta_{2(y)} - 1)(\tau - 1) + (\beta_{2(x)} - 1)\tau + (1 - 2\tau)(\lambda_{22} - 1)}{(\beta_{2(y)} - 1)\{(1 - \tau)^2/\tau\} + 2(\lambda_{22} - 1)(1 - \tau) + (\beta_{2(x)} - 1)\tau} \text{ and } \alpha_2^* = 1 - \alpha_1^*.$$

The corresponding bias of the estimator $\widehat{S}^2_{KC}$, to the first order of approximation, is given by

(11) $$\text{Bias}(\widehat{S}^2_{KC}) \cong (\alpha_1^* - 1)S_y^2 + \gamma S_y^2 \alpha_2^* \tau \left[(\beta_{2(x)} - 1) - (\lambda_{22} - 1)\right].$$

Kadilar and Cingi [6] have shown that the estimator $\widehat{S}^2_{KC}$ is more efficient than the ratio and regression estimators under the following conditions.

(i) $\text{MSE}(\widehat{S}^2_{KC})_{opt} < \text{MSE}(\widehat{S}^2_R)$ if

$$z^2(\beta_{2(y)} - 1) - \beta_{2(y)} - 2z\alpha_2^*\tau(\lambda_{22} - 1) + 2\lambda_{22} + \alpha_2^{*2}\tau^2(\beta_{2(x)} - 1) - \beta_{2(x)} < 0.$$

(ii) $\text{MSE}(\widehat{S}^2_{KC})_{opt} < \text{Var}(\widehat{S}^2_{Reg})$ if

$$(z^2 - 1)(\beta_{2(y)} - 1) - 2z\alpha_2^*\tau(\lambda_{22} - 1) + \alpha_2^{*2}\tau^2(\beta_{2(x)} - 1) + \frac{(\lambda_{22} - 1)^2}{(\beta_{2(x)} - 1)} < 0.$$

2. The proposed estimator

We now propose a hybrid estimator based on the following two mean estimators. The first one is the estimator considered by Sahahi and Ray [8] for estimating $\overline{Y}$. It is given by

(12) $$\widehat{\overline{Y}}_S = \bar{y}\left[2 - \left(\frac{\bar{x}}{\overline{X}}\right)^w\right],$$

where w is a constant.

The other estimator is proposed by Kaur [7], and is given by

(13) $$\widehat{\overline{Y}}_K = \left[k_1 \bar{y} + b_{yx}(\overline{X} - \bar{x})\right],$$

where k_1 is a constant and b_{yx} the sample regression coefficient between y and x.

A possible hybrid class of estimators of S_y^2, based on (12) and (13) is

(14) $$\widehat{S}^{2(\alpha)}_{\text{Pr}} = \left[d_1 s_y^2 + d_2\left(S_x^2 - s_x^2\right)\right]\left[2 - \left(\frac{s_x^2}{S_x^2}\right)^\alpha\right],$$

where d_1, d_2 and α are suitably chosen constants.

Writing (14) in terms of the δ_i 's $(i = 0, 1)$, we have

$$\widehat{S}^{2(\alpha)}_{\mathrm{Pr}} = \left[d_1 S_y^2(1+\delta_0) - d_2 S_x^2 \delta_1\right]\left[2 - (1+\delta_1)^{\alpha}\right]. \tag{15}$$

A first order approximation of (15) is given by

$$\widehat{S}^{2(\alpha)}_{\mathrm{Pr}} - S_y^2 \cong (d_1-1)S_y^2 + d_1 S_y^2\left\{\delta_0 - \alpha\delta_1 - \alpha\delta_0\delta_1 - \frac{\alpha(\alpha-1)}{2}\delta_1^2\right\} - d_2 S_x^2\left(\delta_1 - \alpha\delta_1^2\right). \tag{16}$$

From (16), the bias and MSE of $\widehat{S}^{2(\alpha)}_{\mathrm{Pr}}$ are given by

$$\begin{aligned}\mathrm{Bias}(\widehat{S}^{2(\alpha)}_{\mathrm{Pr}}) \cong (d_1-1)S_y^2 - d_1\alpha\gamma S_y^2\left\{(\lambda_{22}-1) + \frac{(\alpha-1)}{2}(\beta_{2(x)}-1)\right\} \\ + d_2\alpha\gamma S_x^2(\beta_{2(x)}-1)\end{aligned} \tag{17}$$

and

$$\begin{aligned}\mathrm{MSE}(\widehat{S}^{2(\alpha)}_{\mathrm{Pr}}) \cong \mathrm{E}\Big[(d_1-1)S_y^2 + d_1 S_y^2\left\{\delta_0 - \alpha\delta_1 - \alpha\delta_0\delta_1 - \frac{\alpha(\alpha-1)}{2}\delta_1^2\right\} \\ - d_2 S_x^2\left\{\delta_1 - \alpha\delta_1^2\right\}\Big]^2\end{aligned}$$

or

$$\begin{aligned}\mathrm{MSE}(\widehat{S}^{2(\alpha)}_{\mathrm{Pr}}) \cong{}& (d_1-1)^2 S_y^4 + d_1^2 S_y^4\gamma\left\{(\beta_{2(y)}-1) + \alpha^2(\beta_{2(x)}-1) - 2\alpha(\lambda_{22}-1)\right\} \\ &+ d_2^2 S_x^4\gamma(\beta_{2(x)}-1) - 2d_1(d_1-1) \\ &\times S_y^4\alpha\gamma\left\{(\lambda_{22}-1) + \frac{(\alpha-1)}{2}(\beta_{2(x)}-1)\right\} \\ &+ 2d_2(d_1-1)S_y^2 S_x^2\alpha\gamma(\beta_{2(x)}-1) \\ &- 2d_1 d_2 S_y^2 S_x^2\gamma\left\{(\lambda_{22}-1) - \alpha(\beta_{2(x)}-1)\right\}.\end{aligned} \tag{18}$$

From (18), we have

$$\begin{aligned}\mathrm{MSE}(\widehat{S}^{2(\alpha)}_{\mathrm{Pr}}) \cong{}& S_y^4 + d_1^2 S_y^4 A_1^{(\alpha)} + d_2^2 S_x^4 A_2 + 2d_1 d_2 S_y^2 S_x^2 A_3^{(\alpha)} \\ &- 2d_1 S_y^4 A_4^{(\alpha)} - 2d_2 S_y^2 S_x^2 A_5^{(\alpha)},\end{aligned} \tag{19}$$

where

$$\begin{aligned}A_1^{(\alpha)} &= 1 + \gamma\left\{(\beta_{2(y)}-1) + \alpha(\beta_{2(x)}-1) - 4\alpha(\lambda_{22}-1)\right\} \\ A_2 &= \gamma(\beta_{2(x)}-1), \\ A_3^{(\alpha)} &= \gamma\left\{2\alpha(\beta_{2(x)}-1) - (\lambda_{22}-1)\right\}, \\ A_4^{(\alpha)} &= 1 - \alpha\gamma\left\{(\lambda_{22}-1) + \frac{(\alpha-1)}{2}(\beta_{2(x)}-1)\right\}, \\ A_5^{(\alpha)} &= \alpha\gamma(\beta_{2(x)}-1).\end{aligned}$$

Setting $\frac{\partial MSE(\widehat{S}^{2(\alpha)}_{\mathrm{Pr}})}{\partial d_i} = 0$ $(i = 1, 2)$, we have $d_1 = \frac{A_2 A_4^{(\alpha)} - A_3^{(\alpha)} A_5^{(\alpha)}}{A_1^{(\alpha)} A_2 - A_3^{2(\alpha)}} = d_1^{(\alpha)}$ (say) and $d_2 = \frac{S_y^2}{S_x^2}\left\{\frac{A_1^{(\alpha)} A_5^{(\alpha)} - A_3^{(\alpha)} A_4^{(\alpha)}}{A_1^{(\alpha)} A_2 - A_3^{2(\alpha)}}\right\} = d_2^{(\alpha)}$ (say).

Substituting the optimum values of d_1 and d_2 in (19), we get the optimum MSE of $\widehat{S}^{2(\alpha)}_{\mathrm{Pr}}$ in the form

$$\mathrm{MSE}(\widehat{S}^{2(\alpha)}_{\mathrm{Pr}})_{opt} \cong S_y^4\left(1 - \frac{A_1^{(\alpha)} A_5^{2(\alpha)} + A_2 A_4^{2(\alpha)} - 2A_3^{(\alpha)} A_4^{(\alpha)} A_5^{(\alpha)}}{A_1^{(\alpha)} A_2 - A_3^{2(\alpha)}}\right). \tag{20}$$

The corresponding bias of $\widehat{S}^2_{\text{Pr}}$ is given by

$$\text{Bias}(\widehat{S}^{2(\alpha)}_{\text{Pr}}) \cong (d_1^{(\alpha)} - 1)S_y^2 - d_1^{(\alpha)}\alpha\gamma S_y^2\left\{(\lambda_{22} - 1) + \frac{(\alpha - 1)}{2}(\beta_{2(x)} - 1)\right\} + d_2^{(\alpha)}\alpha\gamma S_x^2(\beta_{2(x)} - 1). \tag{21}$$

Note that the expression in (20) is only an ideal optimal MSE value since $d_1^{(\alpha)}$ and $d_2^{(\alpha)}$ involve unknown parameters. However, we can use prior estimates of the population parameters λ_{22} and $\beta_{2(y)}$ in the corresponding variance estimator if good prior estimates are available. Upadhyaya and Singh [10] suggest that if such prior estimates are unknown, then these parameters can be estimated from the sample.

Now we consider three cases for different values of α that are of particular interest.

(i) For $\alpha = 0$, the proposed estimator becomes a generalized regression estimator. When we put $\alpha = 0$ in (14), (19), (20) and (21), we get

$$\widehat{S}^{2(0)}_{\text{Pr}} = \left[d_1 s_y^2 + d_2\left(S_x^2 - s_x^2\right)\right]. \tag{22}$$

The optimum MSE of $\widehat{S}^{2(0)}_{\text{Pr}}$ and the corresponding bias are given by

$$\text{Bias}(\widehat{S}^{2(0)}_{\text{Pr}}) \cong (d_1^{(0)} - 1)S_y^2 \tag{23}$$

and

$$\text{MSE}(\widehat{S}^{2(0)}_{\text{Pr}})_{opt} \cong S_y^4\left(1 - \frac{A_1^{(0)}A_5^{2(0)} + A_2A_4^{2(0)} - 2A_3^{(0)}A_4^{(0)}A_5^{(0)}}{A_1^{(0)}A_2 - A_3^{2(0)}}\right)$$

or

$$\text{MSE}(\widehat{S}^{2(0)}_{\text{Pr}})_{opt} \cong S_y^4\left(1 - \frac{1}{1 + \gamma(\beta_{2(y)} - 1)(1 - \rho^2_{(s_y^2, s_x^2)})}\right) = \frac{\text{Var}(\widehat{S}_y^2)(1 - \rho^2_{(s_y^2, s_x^2)})}{1 + \gamma(\beta_{2(y)} - 1)(1 - \rho^2_{(s_y^2, s_x^2)})}. \tag{24}$$

The optimum values of d_1 and d_2 are given by

$$d_1^{(0)} = \frac{A_2A_4^{(0)} - A_3^{(0)}A_5^{(0)}}{A_1^{(0)}A_2 - A_3^{2(0)}} \text{ and } d_2^{(0)} = \frac{S_y^2}{S_x^2}\left\{\frac{A_1^{(0)}A_5^{(0)} - A_3^{(0)}A_4^{(0)}}{A_1^{(0)}A_2 - A_3^{2(0)}}\right\},$$

where

$$A_1^{(0)} = 1 + \gamma(\beta_{2(y)} - 1),$$
$$A_2 = \gamma(\beta_{2(x)} - 1),$$
$$A_3^{(0)} = -\gamma(\lambda_{22} - 1),$$
$$A_4^{(0)} = 1,$$
$$A_5^{(0)} = 0.$$

(ii) For $\alpha = 1$, the second part of the proposed estimator becomes a transformed product estimator. When we put $\alpha = 1$ in (14), (19), (20) and (21), we get

$$\widehat{S}^{2(1)}_{\text{Pr}} = \left[d_1 s_y^2 + d_2\left(S_x^2 - s_x^2\right)\right]\left[2 - \left(\frac{s_x^2}{S_x^2}\right)\right], \tag{25}$$

The optimum MSE of $\widehat{S}^{2(1)}_{\text{Pr}}$ and the corresponding bias are given by

$$\text{Bias}(\widehat{S}^{2(1)}_{\text{Pr}}) \cong (d_1^{(1)} - 1)S_y^2 - d_1^{(1)}\gamma S_y^2(\lambda_{22} - 1) + d_2^{(1)}S_x^2\gamma(\beta_{2(x)} - 1) \tag{26}$$

and

$$\mathrm{MSE}(\widehat{S}^{2(1)}_{\mathrm{Pr}})_{opt} \cong S_y^4 \left(1 - \frac{A_1^{(1)}A_5^{2(1)} + A_2A_4^{2(1)} - 2A_3^{(1)}A_4^{(1)}A_5^{(1)}}{A_1^{(1)}A_2 - A_3^{2(1)}}\right),$$

or

$$\mathrm{MSE}(\widehat{S}^{2(1)}_{pr})_{opt} \cong S_y^4 \left(1 - \frac{(\beta_{2(x)}-1)U_1}{U_2}\right) = \frac{S_y^4\gamma U_1^*}{U_2}, \tag{27}$$

where

$$\begin{aligned}U_1 &= 1 - 3\gamma(\beta_{2(x)}-1) + \gamma^2(\beta_{2(y)}-1)(\beta_{2(x)}-1) + \gamma^2(\beta_{2(x)}-1)^2 \\ &\quad - \gamma^2(\lambda_{22}-1)^2,\\ U_2 &= (\beta_{2(x)}-1) + \gamma(\beta_{2(y)}-1)(\beta_{2(x)}-1) - 3\gamma(\beta_{2(x)}-1)^2 \\ &\quad - \gamma(\lambda_{22}-1)^2,\\ U_1^* &= (\beta_{2(y)}-1)(\beta_{2(x)}-1) - (\lambda_{22}-1)^2 - \gamma(\beta_{2(y)}-1)(\beta_{2(x)}-1)^2 \\ &\quad - \gamma(\beta_{2(x)}-1)^3 + \gamma(\beta_{2(x)}-1)(\lambda_{22}-1)^2.\end{aligned}$$

The optimum values of d_1 and d_2 are

$$d_1^{(1)} = \frac{A_2A_4^{(1)} - A_3^{(1)}A_5^{(1)}}{A_1^{(1)}A_2 - A_3^{2(1)}} \text{ and } d_2^{(1)} = \frac{S_y^2}{S_x^2}\left\{\frac{A_1^{(1)}A_5^{(1)} - A_3^{(1)}A_4^{(1)}}{A_1^{(1)}A_2 - A_3^{2(1)}}\right\},$$

where

$$\begin{aligned}A_1^{(1)} &= 1 + \gamma\left\{(\beta_{2y}-1) + (\beta_{2(x)}-1) - 4(\lambda_{22}-1)\right\}\\ A_2 &= \gamma(\beta_{2(x)}-1),\\ A_3^{(1)} &= \gamma\left\{2(\beta_{2(x)}-1) - (\lambda_{22}-1)\right\},\\ A_4^{(1)} &= 1 - \gamma(\lambda_{22}-1),\\ A_5^{(1)} &= \gamma(\beta_{2(x)}-1).\end{aligned}$$

(iii) For $\alpha = -1$, the second part of the proposed estimator becomes a transformed ratio estimator. When we put $\alpha = -1$ in (14), (19), (20) and (21), we get

$$\widehat{S}^{2(-1)}_{\mathrm{Pr}} = \left[d_1s_y^2 + d_2\left(S_x^2 - s_x^2\right)\right]\left[2 - \left(\frac{S_x^2}{s_x^2}\right)\right], \tag{28}$$

The optimum MSE of $\widehat{S}^{2(1)}_{\mathrm{Pr}}$ and the corresponding bias are given by

$$\begin{aligned}\mathrm{Bias}(\widehat{S}^{2(-1)}_{\mathrm{Pr}}) &\cong (d_1^{(1)}-1)S_y^2 + d_1^{(1)}\gamma S_y^2\left\{(\lambda_{22}-1) - (\beta_{2(x)}-1\right\}\\ &\quad - d_2^{(-1)}S_x^2\gamma(\beta_{2(x)}-1)\end{aligned} \tag{29}$$

and

$$\mathrm{MSE}(\widehat{S}^{2(-1)}_{\mathrm{Pr}})_{opt} \cong S_y^4\left(1 - \frac{A_1^{(-1)}A_5^{2(-1)} + A_2A_4^{2(-1)} - 2A_3^{(-1)}A_4^{(-1)}A_5^{(-1)}}{A_1^{(-1)}A_2 - A_3^{2(-1)}}\right),$$

or

$$\mathrm{MSE}(\widehat{S}^{2(-1)}_{pr})_{opt} \cong S_y^4\left(1 - \frac{(\beta_{2(x)}-1)U_3}{U_4}\right) = \frac{S_y^4\gamma U_3^*}{U_4}, \tag{30}$$

where

$$U_3 = 1 - 5\gamma(\beta_{2(x)} - 1) + \gamma^2(\beta_{2(y)} - 1)(\beta_{2(x)} - 1) + 4\gamma^2(\beta_{2(x)} - 1)^2 - \gamma^2(\lambda_{22} - 1)^2,$$

$$U_4 = (\beta_{2(x)} - 1) + \gamma(\beta_{2(y)} - 1)(\beta_{2(x)} - 1) - 5\gamma(\beta_{2(x)} - 1)^2 - \gamma(\lambda_{22} - 1)^2,$$

$$U_3^* = (\beta_{2(y)} - 1)(\beta_{2(x)} - 1) - (\lambda_{22} - 1)^2 - \gamma(\beta_{2(y)} - 1)(\beta_{2(x)} - 1)^2 - 4\gamma(\beta_{2(x)} - 1)^3 + \gamma(\beta_{2(x)} - 1)(\lambda_{22} - 1)^2.$$

The optimum values of d_1 and d_2 are

$$d_1^{(-1)} = \frac{A_2 A_4^{(-1)} - A_3^{(-1)} A_5^{(-1)}}{A_1^{(-1)} A_2 - A_3^{2(-1)}} \text{ and } d_2^{(-1)} = \frac{S_y^2}{S_x^2}\left\{\frac{A_1^{(-1)} A_5^{(-1)} - A_3^{(-1)} A_4^{(-1)}}{A_1^{(-1)} A_2 - A_3^{2(-1)}}\right\},$$

where

$$A_1^{(-1)} = 1 + \gamma\left\{(\beta_{2y} - 1) - (\beta_{2(x)} - 1) + 4(\lambda_{22} - 1)\right\}$$
$$A_2 = \gamma(\beta_{2(x)} - 1),$$
$$A_3^{(-1)} = -\gamma\left\{2(\beta_{2(x)} - 1) + (\lambda_{22} - 1)\right\},$$
$$A_4^{(-1)} = 1 + \gamma\left\{(\lambda_{22} - 1) - (\beta_{2(x)} - 1)\right\},$$
$$A_5^{(-1)} = -\gamma(\beta_{2(x)} - 1).$$

3. Comparison of the estimators

We now compare the proposed estimator with other estimators considered here. These comparisons lead to the following obvious conditions.

Case 1: $\alpha = 0$.

Condition (i):

$\mathrm{MSE}(\widehat{S}^{2(0)}_{\mathrm{Pr}})_{opt} < \mathrm{Var}(\widehat{S}^2_y)$ if

$$\mathrm{Var}(\widehat{S}^2_y) - \frac{\mathrm{Var}(\widehat{S}^2_y)(1 - \rho^2_{(s^2_y,\, s^2_x)})}{1 + A} = \frac{\mathrm{Var}(\widehat{S}^2_y)[A + \rho^2_{(s^2_y, S^2_x)}]}{(1 + A)} > 0,$$

where $A = \gamma(\beta_{2(y)} - 1)(1 - \rho^2_{(s^2_y,\, s^2_x)})$.

Condition (ii):

$\mathrm{MSE}(\widehat{S}^{2(0)}_{\mathrm{Pr}})_{opt} < \mathrm{MSE}(\widehat{S}^2_R)$ if

$$\mathrm{MSE}(\widehat{S}^2_R) - \frac{\mathrm{Var}(\widehat{S}^2_y)(1 - \rho^2_{(s^2_y,\, s^2_x)})}{1 + A} > 0, \text{ or}$$

$$\frac{\mathrm{MSE}(\widehat{S}^2_R)A + \mathrm{MSE}(\widehat{S}^2_R) - \mathrm{Var}(\widehat{S}^2_y)(1 - \rho^2_{(s^2_y,\, s^2_x)})}{(1 + A)} > 0, \text{ or}$$

$$\frac{\mathrm{MSE}(\widehat{S}^2_R)A + \gamma S^4_y\left\{(\beta_{2(x)} - 1) - 2(\lambda_{22} - 1) + \frac{(\lambda_{22} - 1)^2}{(\beta_{2(x)} - 1)}\right\}}{(1 + A)} > 0, \text{ or}$$

$$\frac{\mathrm{MSE}(\widehat{S}^2_R)A + \gamma S^4_y(\beta_{2(x)} - 1)\left(1 - \frac{(\lambda_{22} - 1)}{(\beta_{2(x)} - 1)}\right)^2}{(1 + A)} > 0.$$

Condition (iii):

$$\mathrm{MSE}(\widehat{S}^{2(0)}_{\mathrm{Pr}})_{opt} < \mathrm{Var}(\widehat{S}^2_{Reg}) \text{ if}$$

$$\mathrm{MSE}(\widehat{S}^2_{Reg}) - \frac{\mathrm{Var}(\widehat{S}^2_y)(1-\rho^2_{(s^2_y,s^2_x)})}{1+A} > 0, \text{ or}$$

$$\frac{Var(\widehat{S}^2_{Reg})A}{(1+A)} > 0,$$

using (6).

Conditions (iv-vii):

$$\mathrm{MSE}(\widehat{S}^{2(0)}_{\mathrm{Pr}})_{opt} < \mathrm{MSE}(\widehat{S}^2_{KCi}) \ (i=1,2,3,4) \text{ if}$$

$$\mathrm{MSE}(\widehat{S}^2_{KCi}) - \frac{\mathrm{Var}(\widehat{S}^2_y)(1-\rho^2_{(s^2_y,s^2_x)})}{1+A} > 0 \text{ or}$$

$$\frac{\mathrm{MSE}(\widehat{S}^2_{KCi})A + \mathrm{MSE}(\widehat{S}^2_{KCi}) - \mathrm{Var}(\widehat{S}^2_y)(1-\rho^2_{(s^2_y,s^2_x)})}{(1+A)} > 0, \text{ or}$$

$$\frac{\mathrm{MSE}(\widehat{S}^2_{KCi})A + \gamma S^4_y\left\{(\beta_{2(x)}-1)\psi^2_i - 2(\lambda_{22}-1)\psi_i + \frac{(\lambda_{22}-1)^2}{(\beta_{2(x)}-1)}\right\}}{(1+A)} > 0, \text{ or}$$

$$\frac{\mathrm{MSE}(\widehat{S}^2_{KCi})A + \gamma S^4_y(\beta_{2(x)}-1)\left(\psi_i - \frac{(\lambda_{22}-1)}{(\beta_{2(x)}-1)}\right)^2}{(1+A)} > 0.$$

The conditions (i)-(vii) above are always true.

Condition (viii):

$$\mathrm{MSE}(\widehat{S}^{2(0)}_{\mathrm{Pr}})_{opt} < \mathrm{MSE}(\widehat{S}^2_{KC})_{opt} \text{ if}$$

$$\mathrm{MSE}(\widehat{S}^2_{KC})_{opt} - \frac{\mathrm{Var}(\widehat{S}^2_y)(1-\rho^2_{(s^2_y,s^2_x)})}{1+A} > 0, \text{ or}$$

$$\frac{\mathrm{MSE}(\widehat{S}^2_{KC})_{opt}A + \mathrm{MSE}(\widehat{S}^2_{KC})_{opt} - \mathrm{Var}(\widehat{S}^2_y)(1-\rho^2_{(s^2_y,s^2_x)})}{(1+A)} > 0, \text{ or}$$

$$\frac{\mathrm{MSE}(\widehat{S}^2_{KC})_{opt}A + \gamma S^4_y B}{(1+A)} > 0,$$

where $B = (\beta_{2(y)}-1)(z^2+\rho^2_{(s^2_y,s^2_x)}-1) + \alpha^{*2}_2\tau^2(\beta_{2(x)}-1) - 2z\tau\alpha^*_2(\lambda_{22}-1)$.

Obviously the proposed estimator will always perform better than $\widehat{S}^2_y$, $\widehat{S}^2_R$, $\widehat{S}_{KCi}$ ($i = 1,2,3,4$) and $\widehat{S}^2_{Reg}$ since Conditions (i)-(vii) always hold. It will also perform better than $\widehat{S}_{KC}$ if Condition (viii) holds, which indeed may be the case in some situations as shown below for certain data sets that have been used in the literature.

Case 2: $\alpha = 1, -1$.

The proposed estimators $\widehat{S}^{2(1)}_{Pr}$ and $\widehat{S}^{2(-1)}_{Pr}$ are only conditionally better then the other estimators considered here. This will be so if the optimum MSEs of $\widehat{S}^{2(1)}_{Pr}$ and $\widehat{S}^{2(-1)}_{Pr}$ are smaller than the MSEs of the estimators $\widehat{S}^2_i$ $i = y, R, KCi$ $(i = 1,2,3,4), KC$ and Reg.

4. Description of data and results

We use the following data to numerically compare the different estimators.

Data 1: [Source: Kadilar and Cingi [6]]

The data consist of 104 villages in the East Anatolia Region of Turkey in 1999. The variates are defined as:

y is the level of apple production (in 100 tones)

and

x is the number of apple trees.

For this data, we have

$$N = 104,\ n = 20,\ \gamma = 0.04038,\ \overline{Y} = 6.254,\ \overline{X} = 13931.683,$$
$$S_y = 11.67,\ S_x = 23029.072,\ \rho_{yx} = 0.865,\ \rho_{(s_y^2,\ s_x^2)} = 0.836757,$$
$$\beta_{2(y)} = 16.523,\ \beta_{2(x)} = 17.516,\ \lambda_{22} = 14.398,\ \tau = 0.99766,$$
$$\alpha_1^* = 0.18777,\ d_1^{(0)} = 0.84179,\ d_2^{(0)} = 0,\ d_1^{(1)} = 0.41072,$$
$$d_2^{(1)} = 0,\ d_1^{(-1)} = 0.46619,\ d_2^{(-1)} = 0.$$

Data 2: [Source: Das [2]]

The data consist of 278 villages/towns/wards under Gajole Police Station of Malda district of West Bengal, India. The variates are defined as:

y is the number of agricultural laborers in 1971

and

x is the number of agricultural laborers in 1961.

For this data, we have:

$$N = 278,\ n = 30,\ \gamma = 0.02974,\ \overline{Y} = 39.068,\ \overline{X} = 25.111,$$
$$S_y = 56.457167,\ S_x = 40.674797,\ \rho_{yx} = 0.7213,\ \rho_{(s_y^2,\ s_x^2)} = 0.840475,$$
$$\beta_{2(y)} = 25.8969,\ \beta_{2(x)} = 38.8898,\ \lambda_{22} = 26.8142,\ \tau = 0.974196,$$
$$\alpha_1^* = 0.30808,\ d_1^{(0)} = 0.82143,\ d_2^{(0)} = 1.07818,\ d_1^{(1)} = 0.57957,$$
$$d_2^{(1)} = 0.45414,\ d_1^{(-1)} = 0.53897,\ d_2^{(-1)} = 0.85759.$$

Data 3: [Source: Cochran [1, p. 325]]

The data consist of 100 blocks in a large city of which 10 were chosen. The variates are defined as:

y is the number of persons per block

and

x is the number of rooms per block.

For this data, we have:

$$N = 100,\ n = 10,\ \gamma = 0.09,\ \overline{Y} = 101.1,\ \overline{X} = 58.8,$$
$$S_y = 14.6595,\ S_x = 7.53228,\ \rho_{yx} = 0.65,\ \rho_{(s_y^2,\ s_x^2)} = 0.419701,$$
$$\beta_{2(y)} = 2.3523,\ \beta_{2(x)} = 2.2387,\ \lambda_{22} = 1.5432,\ \tau = 0.99961,$$
$$\alpha_1^* = 0.56103,\ d_1^{(0)} = 0.90887,\ d_2^{(0)} = 1.50966,\ d_1^{(1)} = 1.01464,$$
$$d_2^{(1)} = -2.21336,\ d_1^{(-1)} = 1.22602,\ d_2^{(-1)} = 7.53651.$$

Data 4: [Source: Kadilar and Cingi [5]]

The data consist of 106 villages in the Marmarian Region of Turkey in 1999. The variates are defined as:

y is the level of apple production (1 unit = 100 tonnes) and

x is the number of apple trees (1 unit = 100 trees).

For this data, we have:

$$N = 106, n = 20,\ \gamma = 0.04056, \overline{Y} = 15.37,\ \overline{X} = 243.76,$$
$$S_y = 64.25,\ S_x = 491.89,\ \rho_{yx} = 0.82,\ \rho_{(s_y^2,\ s_x^2)} = 0.730458,$$
$$\beta_{2(y)} = 80.13,\ \beta_{2(x)} = 25.71,\ \lambda_{22} = 33.30,\ \tau = 0.974196,$$
$$\alpha_1^* = -0.17269,\ d_1^{(0)} = 0.40048,\ d_2^{(0)} = 0.00893,\ d_1^{(1)} = 1.97072,$$
$$d_2^{(1)} = -0.00623,\ d_1^{(-1)} = 0.79816,\ d_2^{(-1)} = 0.02797.$$

We obtain the Percent Relative Efficiency (PRE) of the various estimators as compared to $\widehat{S}_y^2$. These are given in Table 1.

Table 1. PRE of the various estimators with respect to $\widehat{S}_y^2$

Estimator	Data 1	Data 2	Data 3	Data 4
$\widehat{S}_y^2$	100.000	100.000	100.000	100.000
$\widehat{S}_R^2$	296.071	223.124	89.878	201.656
$\widehat{S}_{KC1}^2$	296.071	223.597	90.065	201.656
$\widehat{S}_{KC2}^2$	296.071	234.354	93.032	201.648
$\widehat{S}_{KC3}^2$	296.071	223.137	89.961	201.656
$\widehat{S}_{KC4}^2$	296.071	230.088	108.555	201.652
$\widehat{S}_{KC}^2$	334.786	353.047	121.423	169.493
$\widehat{S}_{Reg}^2$	333.514	340.597	121.381	214.394
$\widehat{S}_{\mathrm{Pr}}^{2(0)}$	396.196	414.641	133.552	535.346
$\widehat{S}_{\mathrm{Pr}}^{2(1)}$	133.311	123.456	121.580	162.318
$\widehat{S}_{\mathrm{Pr}}^{2(-1)}$	78.318	64.041	167.789	200.665

In Table 1, the efficiencies of the estimators $\widehat{S}_R^2$ and $\widehat{S}_{KCi}^2$ $(i = 1, 2, 3, 4)$ are the same in Data 1 because the values of ψ_i $(i = 1, 2, 3, 4)$ are all equal to one. Also, one can observe that the estimators $\widehat{S}_{KCi}^2$ $(i = 1, 2, 3, 4)$ and $\widehat{S}_{KC}^2$ may not always be more efficient than the regression estimator, but that one of the proposed estimators $(\widehat{S}_{\mathrm{Pr}}^{2(0)})$ perform better than the regression estimator. This is to be expected based on the fact that condition (iii) for $\alpha = 0$ always holds. The estimator $(\widehat{S}_{\mathrm{Pr}}^{2(0)})$ is more efficient than $\widehat{S}_{KCi}^2$ $(i = 1, 2, 3, 4)$ and $\widehat{S}_{KC}^2$ also. However, the estimators $\widehat{S}_{\mathrm{Pr}}^{2(1)}$ and $\widehat{S}_{\mathrm{Pr}}^{2(-1)}$ are not always superior to the other estimators.

Acknowledgement

The authors appreciate the feedback provided by the referees that helped improve the presentation of the paper. The second author also acknowledges with thanks the support provided by the University of North Carolina at Greensboro.

References

[1] Cochran, W. G. *Sampling Techniques*, 3rd edn. (Wiley & Sons, 1977).

[2] Das, A. K. *Contribution to the theory of sampling strategies based on auxiliarly information* (Ph.D. thesis submitted to Bidhan Chandra Krishi Vishwavidyalaya, Mohanpur, Nadia, west Bengal, India, 1988).

[3] Garcia, M. R. and Cebrian, A. A. *Repeated substitution method: the ratio estimator for the population variance*, Metrika. **43**, 101–105, 1996.

[4] Isaki, C. T. *Variance estimation using auxiliary information*, Journal of American Statistical Association **78**, 117–123, 1983.

[5] Kadilar, C. and Cingi, H. *Ratio estimators for population variance in simple and stratified sampling*, Applied Mathematics and Computation **173**, 1047–1058, 2006.

[6] Kadilar, C. and Cingi, H. *Improvement in variance estimation using auxiliary information*, Hacettepe Journal of Mathematics and Statistics **35** (1), 111–115, 2006.

[7] Kaur, P. *An efficient regression type estimator in survey sampling*, Biom. Journal **27** (1), 107–110, 1985.

[8] Sahahi, A. and Ray, S. K. *An efficient estimator using auxiliary information*, Metrika **27**, 271–275, 1980.

[9] Shabbir, J. and Yaab, Z. *Improvement over transformed auxiliary variables in estimating the finite population mean*, Biometrical Journal **45** (6), 723–729, 2003.

[10] Upadhyaya, L. N. and Singh, H. P. *Almost unbiased ratio and product-type estimators of finite population variance in sample surveys*, Statistics in Transition **7** (5), 1087–1096, 2006.