



The Ratio-type Estimators of Variance with Minimum Average Square Error

Mohammadreza Faghihi

Shahid Beheshti University

Abstract. The ratio-type estimators have been introduced for estimating the mean and total population, but in recent years based on the ratio methods several estimators for population variance have been proposed. In this paper two families of estimators have been suggested and their approximation mean square error (MSE) have been developed. In addition, the efficiency of these variance estimators are compared. We also show that their performance is better than traditional ratio estimator. Finally, a computational approach to find the optimum values of parameters is introduced and by using rice fields data of Amol county of Iran, an empirical study is carried out to show that this approach is the best. In this manner, since the values of estimators MSE are approximated, to obtain the approximately unbiased estimator with minimum variance, average square error (ASE) is used.

Keywords. Ratio estimator; ratio-type estimator; auxiliary information; variance estimation; mean square error; average square error.

MSC 2010: 62D05, 62G05, 65C60.

1 Introduction

The auxiliary information in sample surveys helps to maximize the precision of estimators. Mostly, the researchers have introduced estimators for population mean by using the auxiliary information, with ratio methods. Diana (1993) suggested a family of estimators for the population mean in

* Corresponding author

the stratified random sampling and following Diana (1993), Koyuncu and Kadilar (2010) proposed the general family of estimators, Khoshnevisan et al. (2007) also suggested a family of estimators for population mean. Following Khoshnevisan et al. (2007) and Koyuncu and Kadilar (2010), we propose families of estimators for population variance. By using a computational approach, we compare a computational estimator with these estimators and estimators that Isaki (1983) and Kadilar and Cingi (2006) proposed for population variance.

Consider a finite population $U = \{u_1, u_2, \dots, u_N\}$ of size N and let y_j and x_j , respectively, be the study and auxiliary variables associated with each unit u_j ($j = 1, 2, \dots, N$) of the population. In order to estimate the population variance of the study variable y , assuming the knowledge of population variance S_X^2 of the auxiliary variable x , Isaki (1983) proposed the ratio estimator as

$$s_{ratio}^2 = \frac{s_y^2}{s_x^2} S_X^2 \quad (1)$$

where s_y^2 and s_x^2 are the unbiased estimators of S_Y^2 and S_X^2 .

Theorem 1. *Approximate mean square error (MSE) of estimator (1) can be written as*

$$MSE(s_{ratio}^2) \cong \lambda S_Y^4 \{\beta_2(y) + \beta_2(x) - 2\theta\} \quad (2)$$

where

$$\beta_2(y) = \frac{\mu_{40}}{\mu_{20}^2}, \quad \beta_2(x) = \frac{\mu_{04}}{\mu_{02}^2}, \quad \lambda = \frac{1}{n},$$

$$\mu_{rs} = \frac{1}{N} \sum_{j=1}^N (Y_j - \bar{Y})^r (X_j - \bar{X})^s, \quad \theta = \frac{\mu_{22}}{\mu_{20}\mu_{02}}.$$

Proof. If e_0 and e_1 defined as

$$e_0 = \frac{s_y^2 - S_Y^2}{S_Y^2} \quad (3)$$

and

$$e_1 = \frac{s_x^2 - S_X^2}{S_X^2}, \quad (4)$$

then

$$E(e_0) = 0, \quad E(e_1) = 0,$$

$$E(e_0^2) = \frac{\text{var}(s_y^2)}{S_Y^4},$$

$$E(e_1^2) = \frac{\text{var}(s_x^2)}{S_X^4},$$

and

$$E(e_1 e_2) = \frac{\text{cov}(s_x^2, s_y^2)}{S_Y^2 S_X^2}.$$

With assumptions $|e_0| < 1$, $|e_1| < 1$ we get

$$s_{ratio}^2 = S_Y^2(1 + e_0)(1 + e_1)^{-1}. \quad (5)$$

Expanding the right hand side of (5) and retaining terms up to the second powers of e 's, we have

$$s_{ratio}^2 \cong S_Y^2(1 + e_0 - e_1 + e_1^2 - e_0 e_1). \quad (6)$$

Subtracting S_Y^2 from both sides of (6), squaring both sides and taking expectations, and eliminating the terms with higher than the second degree we get the approximate *MSE* as

$$\begin{aligned} E(s_{ratio}^2 - S_Y^2)^2 &\cong S_Y^4 E(e_0^2 + e_1^2 - 2e_0 e_1) \\ &= S_Y^4 \left\{ \frac{\text{var}(s_y^2)}{S_Y^4} + \frac{\text{var}(s_x^2)}{S_X^4} - 2 \frac{\text{cov}(s_x^2, s_y^2)}{S_Y^2 S_X^2} \right\}. \end{aligned}$$

To obtain the variance and covariance of typical variances, we use the following moments

$$m_r = \frac{1}{n} \sum_{i=1}^n (x_i - m'_1)^r,$$

and

$$m'_r = \frac{1}{n} \sum_{i=1}^n x_i^r.$$

We have

$$\begin{aligned} (x_i - m'_1)^r &= \sum_{j=0}^r \binom{r}{j} x_i^j (-m'_1)^{r-j} \\ &= x_i^r - r m'_1 x_i^{r-1} + \frac{r(r-2)}{2} x_i^{r-2} (m'_1)^2 - \dots \\ &= x_i^r - r m'_1 x_i^{r-1} + O\left(\frac{1}{n^2}\right) \end{aligned}$$

and eliminating the terms with coefficient less than $\frac{1}{n}$, the moment variance m_r , can be approximated by

$$\begin{aligned} \text{var}(m_r) &= E(m_r^2) - \{E(m_r)\}^2 \cong E \left(\frac{\sum_i x_i^r - \frac{r \sum_{i \neq j} x_i x_j^{r-1}}{n}}{n} \right)^2 - \{E(m_r)\}^2 \\ &= \frac{1}{n^2} E \left(\sum_i x_i^{2r} + \sum_{i \neq j} x_i^r x_j^r + \frac{r^2}{n^2} \sum_{i \neq j} x_i^2 x_j^{2r-2} \right. \\ &\quad \left. + \frac{r^2}{n^2} \sum_{i \neq j \neq k} x_i^2 x_j^{r-1} x_k^{r-1} - \frac{2r}{n} \sum_{i \neq j} x_i^{r+1} x_j^{r-1} \right) - \{E(m_r)\}^2. \end{aligned}$$

With assumption $E(m_r) = \mu_r$, we can get

$$\text{var}(m_r) \cong \frac{1}{n} (\mu_{2r} - \mu_r^2 + r^2 \mu_2 \mu_{r-1}^2 - 2r \mu_{r-1} \mu_{r+1}), \quad (7)$$

and similarly, we obtain

$$\text{cov}(m_r, m_s) \cong \frac{1}{n} (\mu_{r+s} - \mu_r \mu_s + r s \mu_{r-1} \mu_{s-1} + r \mu_{r-1} \mu_{s+1} - s \mu_{r+1} \mu_{s-1}), \quad (8)$$

(Kendall and Stuart, 1969).

Substituting $m_2 = s_y^2$ in (7), we can get

$$\begin{aligned} \text{var}(s_y^2) &\cong \frac{1}{n} (\mu_4 - \mu_2^2 + 4\mu_2 \mu_1^2 - 4\mu_1 \mu_3) \\ &= \frac{1}{n} (\mu_{40} - S_Y^4 + 4\mu_{20} \mu_{10}^2 - 4\mu_{10} \mu_{30}) \\ &= \frac{1}{n} (\mu_{40} - S_Y^4) \\ &= \frac{1}{n} S_Y^4 \left(\frac{\mu_{40}}{S_Y^4} - 1 \right) \\ &= \frac{1}{n} S_Y^4 (\beta_2(y) - 1). \end{aligned} \quad (9)$$

Similarly, the variance of s_x^2 can be written as

$$\text{var}(s_x^2) \cong \lambda S_X^4 \{\beta_2(x) - 1\}, \quad \lambda = \frac{1}{n}. \quad (10)$$

Substituting $m_{02} = s_x^2$ and $m_{20} = s_y^2$ in (8), the covariance of sample variances would be

$$\begin{aligned} \text{cov}(s_x^2, s_y^2) &\cong \frac{1}{n}(\mu_{22} - \mu_{20}\mu_{02} + 4\mu_{1,1}\mu_{10}\mu_{01} - 2\mu_{21}\mu_{01} - 2\mu_{10}\mu_{12}) \\ &= \lambda S_Y^2 S_X^2 (\theta - 1). \end{aligned} \quad (11)$$

Hence, the *MSE* of estimator will be as (2). \square

Kadilar and Cingi (2006) proposed four estimators for population variance as

$$\begin{aligned} s_{p1}^2 &= \frac{s_y^2}{s_x^2 - C_x} (S_X^2 - C_x), \\ s_{p2}^2 &= \frac{s_y^2}{s_x^2 - \beta_2(x)} \{S_X^2 - \beta_2(x)\}, \\ s_{p3}^2 &= \frac{s_y^2}{s_x^2 \beta_2(x) - C_x} \{S_X^2 \beta_2(x) - C_x\}, \\ s_{p4}^2 &= \frac{s_y^2}{s_x^2 C_x - \beta_2(x)} \{S_X^2 C_x - \beta_2(x)\}, \end{aligned}$$

where C_x is coefficient of variation of axillary variable. The *MSE* of quadruplet estimators has been represented as

$$MSE(s_{pi}^2) \cong \lambda S_Y^4 [\beta_2(y) - 1 - 2A_i(\theta - 1) + A_i^2 \{\beta_2(x) - 1\}], \quad i = 1, 2, 3, 4 \quad (12)$$

where

$$\begin{aligned} A_1 &= \frac{S_X^2}{S_X^2 - C_x}, & A_2 &= \frac{S_X^2}{S_X^2 - \beta_2(x)}, \\ A_3 &= \frac{S_X^2 \beta_2(x)}{S_X^2 \beta_2(x) - C_x}, & A_4 &= \frac{S_X^2}{S_X^2 C_x - \beta_2(x)}. \end{aligned}$$

It has been shown that s_{p2}^2 is the best estimator among quadruplet estimators. To compare the efficiency of the proposed estimators s_{pi}^2 with the ratio estimator, using (2) and (12), we will have

$$MSE(s_{ratio}^2) - MSE(s_{pi}^2) > 0, \quad i = 1, 2, 3, 4,$$

if

$$A_i^2 \beta_2(x) - A_i^2 - 2A_i \theta + 2A_i - 1 - \beta_2(x) + 2\theta < 0.$$

In such a case, the quadruplet estimators are more efficient than ratio estimator. Note that quadruplet estimators are dependent to the scale of auxiliary variable.

2 The First Suggested Family of Estimators

Khoshnevisan et al. (2007) proposed a general family of estimators for population mean as

$$t = \bar{y} \left\{ \frac{a\bar{X} + b}{\alpha(a\bar{x} + b) + (1 - \alpha)(a\bar{X} + b)} \right\}^g$$

where $a (\neq 0)$ and b are either real numbers or functions of the known parameters of the auxiliary variable x .

Following them, we define a family of estimators for population variance as

$$\hat{S}_{F1}^2 = s_y^2 \left\{ \frac{aS_X^2 + b}{\alpha(as_x^2 + b) + (1 - \alpha)(aS_X^2 + b)} \right\}^g.$$

Theorem 2. *The approximate bias and MSE of estimator \hat{S}_{F1}^2 are*

$$B(\hat{S}_{F1}^2) \cong \left\{ \frac{g(g+1)}{2} (\lambda')^2 \alpha^2 \frac{\text{var}(s_x^2)}{S_X^4} - (\lambda') \alpha g \frac{\text{cov}(s_y^2, s_x^2)}{S_Y^2 S_X^2} \right\}, \quad (13)$$

and

$$MSE(\hat{S}_{F1}^2) \cong S_Y^4 \left\{ \frac{\text{var}(s_y^2)}{S_Y^4} + (\lambda')^2 \alpha^2 g^2 \frac{\text{var}(s_x^2)}{S_X^4} - 2(\lambda') \alpha g \frac{\text{cov}(s_x^2, s_y^2)}{S_Y^2 S_X^2} \right\}, \quad (14)$$

where

$$\lambda' = \frac{aS_X^2}{aS_X^2 + b}.$$

In addition, the minimum value of approximate MSE is

$$MSE^*(\hat{S}_{F1}^2) \cong \text{var}(s_y^2) \{1 - \rho(s_x^2, s_y^2)^2\}, \quad (15)$$

in which $\rho(s_x^2, s_y^2)$ is the correlation coefficient of s_x^2 and s_y^2 .

Proof. To obtain the bias and MSE of \hat{S}_{F1}^2 , by using e_0 and e_1 , defined in (3) and (4), we have

$$s_y^2 = S_Y^2(1 + e_0), \quad s_x^2 = S_X^2(1 + e_1).$$

Expressing \hat{S}_{F1}^2 in terms of e 's, we have

$$\hat{S}_{F1}^2 = S_Y^2(1 + e_0)(1 + \lambda' \alpha e_1)^{-g} \quad (16)$$

where

$$\lambda' = \frac{aS_X^2}{aS_X^2 + b}.$$

We assume that $|\lambda' \alpha e_1| < 1$ so $(1 + \lambda' \alpha e_1)^{-g}$ is expandable. Expanding the right hand side of (16) and retaining terms up to the second powers of e 's, we have

$$\hat{S}_{F1}^2 \cong S_Y^2 \left\{ 1 + e_0 - (\lambda') \alpha g e_1 + \frac{g(g+1)}{2} (\lambda')^2 \alpha^2 e_1^2 - \lambda' \alpha g e_0 e_1 \right\}.$$

Subtracting S_Y^2 from both sides and then taking expectation of both sides, we get the bias of estimator \hat{S}_{F1}^2 as

$$B(\hat{S}_{F1}^2) \cong \left\{ \frac{g(g+1)}{2} (\lambda')^2 \alpha^2 \frac{\text{var}(s_x^2)}{S_X^4} - (\lambda') \alpha g \frac{\text{cov}(s_y^2, s_x^2)}{S_Y^2 S_X^2} \right\}.$$

Squaring difference of \hat{S}_{F1}^2 and S_Y^2 and then taking expectations, we get approximate MSE of the estimator \hat{S}_{F1}^2 as

$$\begin{aligned} MSE(\hat{S}_{F1}^2) &\cong S_Y^4 E\{e_0^2 + (\lambda')^2 \alpha^2 g^2 e_1^2 - 2(\lambda') \alpha g e_0 e_1\} \\ &= S_Y^4 \left\{ \frac{\text{var}(s_y^2)}{S_Y^4} + (\lambda')^2 \alpha^2 g^2 \frac{\text{var}(s_x^2)}{S_X^4} - 2(\lambda') \alpha g \frac{\text{cov}(s_x^2, s_y^2)}{S_Y^2 S_X^2} \right\}. \end{aligned}$$

In order to obtain the minimum value of (14) with respect to α , by taking

$$\frac{\partial MSE}{\partial \alpha} = 0,$$

we have

$$\alpha_{opt} = \frac{\text{cov}(s_y^2, s_x^2) S_X^2}{\text{var}(s_x^2) \lambda' g S_Y^2}. \quad (17)$$

Substitution of (17) in (14) yields the minimum value of $MSE(\hat{S}_{F1}^2)$ as

$$\begin{aligned}
 MSE^*(\hat{S}_{F1}^2) &\cong S_Y^4 \left\{ \frac{\text{var}(s_y^2)}{S_Y^4} - \frac{\text{cov}(s_x^2, s_y^2)^2}{S_Y^4 \text{var}(s_x^2)} \right\} \\
 &= \text{var}(s_y^2) \{1 - \rho(s_x^2, s_y^2)^2\}.
 \end{aligned}$$

□

To show some family members of estimator \hat{S}_{F1}^2 , Table 1 presents some of the important known estimators of the population variance which can be obtained by suitable choice of constants α , a , b and g in \hat{S}_{F1}^2 .

Table 1. Some family members of estimators \hat{S}_{F1}^2 .

Estimators	α	a	b	g
s_y^2	0	a	b	g
s_{ratio}^2	1	1	0	1
s_{p2}^2	1	1	$-\beta_2(x)$	1

In addition, a large number of estimators can be generated from the proposed family of estimators \hat{S}_{F1}^2 just by putting other values of α , g , a , and b . Obviously, the approximate bias and MSE of a given member of the family can be obtained by just substituting the values of α , g , a , and b in (13) and (14), respectively.

3 The Second Suggested Family of Estimators

Koyuncu and Kadilar (2010) introduced a family of estimators to estimate the population mean in stratified sampling as

$$\bar{y}_N = \bar{y}_{st} \left(\frac{\bar{x}_{st}}{\bar{X}} \right)^\delta \left\{ w + (1 - w) \left(\frac{\bar{x}_{st} A_{st} + B_{st}}{\bar{X} A_{st} + B_{st}} \right)^\varepsilon \right\}^\eta,$$

where $\bar{y}_{st} = \sum_{h=1}^L W_h \bar{y}_h$, $A_{st} = \sum_{h=1}^L W_h A_h$ and $B_{st} = \sum_{h=1}^L W_h B_h$. Here A_h and B_h could be the population information of the auxiliary variable for the h th stratum such as S_{xh} , coefficient of variation C_{xh} , skewness $\beta_{1h}(x)$, kurtosis $\beta_{2h}(x)$, and correlation coefficient $\rho_h(xy)$.

Following Diana (1993) and Koyuncu and Kadilar (2010) with using some known population parameters, we define a general family of estimators for

the population variance as (with new assumption $\eta = 1$ for avoiding too many parameters)

$$\hat{S}_{F2}^2 = s_y^2 \left(\frac{s_x^2}{S_X^2} \right)^\delta \left\{ w + (1-w) \left(\frac{s_x^2 A + B}{S_X^2 A + B} \right)^\varepsilon \right\},$$

where A and B could be the population information of the auxiliary variable such as C_X and $\beta_1(x)$ and the parameters δ , ε and w can take finite values. When these parameters are arbitrarily chosen, many estimators will be obtained. Also, when one parameter is considered as “free parameter”, it is possible to obtain some subclasses of estimators. In other words, we can minimize the MSE equation according to the free parameter.

Theorem 3. *The approximate bias and MSE of estimator \hat{S}_{F2}^2 are*

$$B(\hat{S}_{F2}^2) \cong S_Y^2 \left\{ c_1 \frac{\text{cov}(s_x^2, s_y^2)}{S_X^2 S_Y^2} + c_2 \frac{\text{var}(s_x^2)}{S_X^4} \right\},$$

and

$$MSE(\hat{S}_{F2}^2) \cong S_Y^4 \left\{ \frac{\text{var}(s_y^2)}{S_Y^4} + \frac{c_1^2 \text{var}(s_x^2)}{S_X^4} + 2c_1 \frac{\text{cov}(s_x^2, s_y^2)}{S_X^2 S_Y^2} \right\},$$

where

$$c_i = \sum_{j=0}^i a_j b_{i-j} (d)^{i-j},$$

in which

$$a_i = \frac{\delta(\delta-1)\dots(\delta-i+1)}{i!} \quad \text{for } i = 1, 2, \dots \quad \text{and } a_0 = 1,$$

$$b_i = \frac{(1-w)\varepsilon(\varepsilon-1)\dots(\varepsilon-i+1)}{i!} \quad \text{for } i = 1, 2, \dots \quad \text{and } b_0 = 1,$$

and

$$d = \frac{S_X^2 A}{S_X^2 A + B}.$$

In addition, the minimum value of approximate MSE is

$$MSE^*(\hat{S}_{F2}^2) \cong \text{var}(s_y^2) \{1 - \rho(s_x^2, s_y^2)^2\} \quad (18)$$

Proof. Expressing the estimator $\hat{S}_{F_2}^2$, in terms of e_i ($i = 0, 1$), we obtain

$$\hat{S}_{F_2}^2 = S_Y^2(1 + e_0)(1 + e_1)^\delta \{w + (1 - w)(1 + e_1d)^\varepsilon\}.$$

Since $\hat{S}_{F_2}^2$ is given by terms $(1 + e_1)^\delta$ and $\{w + (1 - w)(1 + e_1d)^\varepsilon\}$, so we can get

$$\hat{S}_{F_2}^2 = S_Y^2(1 + e_0)(a_0 + a_1e_1 + a_2e_1^2 + \dots)(b_0 + b_1de_1 + b_2d^2e_1^2 + \dots).$$

Retaining terms up to the second degree of e_i s, we can write

$$B(\hat{S}_{F_2}^2) \cong S_Y^2 \left\{ c_1 \frac{\text{cov}(s_x^2, s_y^2)}{S_X^2 S_Y^2} + c_2 \frac{\text{var}(s_x^2)}{S_X^4} \right\},$$

and

$$MSE(\hat{S}_{F_2}^2) \cong S_Y^4 \left\{ \frac{\text{var}(s_y^2)}{S_Y^4} + \frac{c_1^2 \text{var}(s_x^2)}{S_X^4} + 2c_1 \frac{\text{cov}(s_x^2, s_y^2)}{S_X^2 S_Y^2} \right\}.$$

Note that the optimum value of c_1 is obtained as $c_{1(opt)} = \frac{-\text{cov}(s_x^2, s_y^2)S_X^2}{\text{var}(s_x^2)S_Y^2}$. Using this optimal value, the minimum MSE of the estimator is

$$\begin{aligned} MSE^*(\hat{S}_{F_2}^2) &\cong S_Y^4 \left\{ \frac{\text{var}(s_y^2)}{S_Y^4} - \frac{\text{cov}(s_x^2, s_y^2)^2}{S_Y^4 \text{var}(s_x^2)} \right\} \\ &= \text{var}(s_y^2) \{1 - \rho(s_x^2, s_y^2)^2\}. \end{aligned}$$

□

Some known family members of the estimators $\hat{S}_{F_2}^2$ are shown in Table 2. Therefore estimators $\hat{S}_{F_1}^2$ and $\hat{S}_{F_2}^2$ can be called generalized estimators, because known estimators are special cases of them.

Table 2. Some family members of estimator $\hat{S}_{F_2}^2$.

Estimators	δ	w	A	B	ε
s_y^2	0	1	A	B	ε
s_{ratio}^2	-1	0	A	B	ε
s_{p2}^2	0	0	1	$-\beta_2(x)$	-1
$\hat{S}_{F_1}^2 _{g=-1}$	0	$1 - \alpha$	a	b	1

Remind that A and B can be any population information of the auxiliary variable such as C_X and $\beta_1(x)$.

4 Computational Approach

In analytical approach, approximating the mean square error of estimators which are obtained by retaining terms up to the second degree, decreases the accuracy of MSE . As a reminder, for families \hat{S}_{F1}^2 and \hat{S}_{F2}^2 the same minimum value for MSE is obtained. On the other hand, increasing the number of estimator parameters makes obtaining the minimum value of MSE more difficult. By using analytical approach, we have to compare the performance of estimators only through their approximate MSE 's which have not enough accuracy. In order to solve these problems and to obtain the best estimator with known parameters, we propose a computational approach.

In this approach, we assume that the values of population variables X and Y are available e.g. from last census data, and the aim is determination of the optimum value of estimator parameters such as \hat{S}_{F2}^2 parameters, δ , w , ε . Note that, in approximate MSE also some population parameters, like S_Y^2 , are needed which we assume they are available from previous studies or last census data. So, suppose that k simple random sample of size n have been taken from population and k values of variance estimation have been calculated by applying given values of estimator parameters (such as δ , w , ε). If we show these obtained estimations with \hat{S}_i^2 , since the population variance is known, say S^2 , we can use estimated average square error

$$EASE(\hat{S}^2) = \frac{1}{k} \sum_{i=1}^k (\hat{S}_i^2 - S^2)^2 \quad (19)$$

as a criterion for estimator performance evaluation. It is expected that the average squares of differences between estimated values and the real variance of population or average square error ($EASE$) tends to real MSE by increasing k . But the difference between $EASE$ and approximate MSE could be high. Now by using an optimization method, values of parameters of estimator which minimize the $EASE$ can be obtained.

To illustrate the algorithm of this computational approach, suppose that we have the census data for both variables X and Y and estimation of \hat{S}_{F2}^2 is of interest, then the algorithm has the following steps

1. Take k simple random samples of size n from the population, such that we get an array, say D , with k matrices of $n \times 2$,
2. give some initial values to the parameters of \hat{S}_{F2}^2 , i.e. δ, w, ε . With these values, k estimated variances for samples in D can be obtained and hence the $EASE$ can be calculated,
3. use an optimization algorithm e.g. `nlminb` function in R, with $EASE$ as objective function, initial values of step 2 and array D as data, to get the optimum values of parameters δ, w, ε .

Therefore by using this algorithm, the optimum values of parameters δ, w , and ε and minimized $EASE$ will be obtained.

To apply the computational approach, we use the Amol county data from agricultural census of Iran in 2003. The used data contain the rice production (in tones) as study variable (Y) and the area of paddy fields (in hectares) as auxiliary variable (X) for 229 paddies. The summary statistics of these population data are given in Table 3.

Table 3. The summary statistics of Amol county.

Statistic	Value
N	229
S_Y^2	120.9
S_X^2	40984.67
$\beta_2(y)$	70.48
$\beta_2(x)$	27.76
C_X	1.164393
ρ	0.9103379

Based on simple random sampling, we select a sample of size $n = 20$ out of the 229 rice paddies, 50,000 times. The results of computational approach are shown in Table ???. Note that, only the optimum values of free parameters used for \hat{S}_{F1}^2 and \hat{S}_{F2}^2 estimators are obtained by minimizing the $EASE$. To get these optimum values, the `nlminb` function in R is used. The initial values and boundaries of parameters for the families of estimators \hat{S}_{F1}^2 and \hat{S}_{F2}^2 are respectively ($\alpha = 0.9, g = 2$) and ($\delta = -1, w = 0.8, \varepsilon = 2$). Remind that the values of a, b, A , and B for \hat{S}_{F1}^2 and \hat{S}_{F2}^2 could be any population information which are selected by researcher, so in this study they were chosen as $a = A = C_X = 1.164393$ and $b = B = \rho_{X,Y} = 0.9103379$.

Table 4. The *EASE* values of compared estimators.

Estimator	<i>EASE</i>
s_{ratio}^2	2473.504
s_{p2}^2	2471.16
\hat{S}_{F1}^2	2112.893
\hat{S}_{F2}^2	2046.642

The results show that the family \hat{S}_{F2}^2 has the minimum *EASE* value and so it is recommended as the best estimator of variance between the candidate estimators. The optimum values of parameters are shown in Tables 5 and 6.

Table 5. The optimum values of parameters of \hat{S}_{F1}^2 estimator.

Parameter	α	g
Optimum Value	0.8939228	1.3303677

Table 6. The optimum values of parameters of \hat{S}_{F2}^2 estimator.

Parameter	δ	w	ε
Optimum Value	-1.0075	1.0352	1.6711

Using the optimum values of parameters, \hat{S}_{F2}^2 estimator can be used in sample surveys implementing after census to estimate variance of Y .

Note that, ratio-type estimators are biased, so if we want to have an unbiased estimator we can subtract the value of biasness from the estimator. The bias of estimators can be estimated with respect to the notations of (19) by

$$\hat{B}(\hat{S}^2) = \frac{1}{k} \sum_{i=1}^k (\hat{S}^2_i - S^2).$$

Remind that for calculating the bias values of proposed estimators, initially the optimal values of parameters should be replaced in estimators. The estimated bias values are shown in Table 7.

Table 7. The estimated bias values of estimators.

Estimator	Bias
s_{ratio}^2	-20.11995
s_{p2}^2	-20.02178
\hat{S}_{F1}^2	-21.42566
\hat{S}_{F2}^2	-24.17231

Therefore the final approximately unbiased ratio-type estimator with minimum *EASE* would be

$$\hat{S}_{F2}^2 = s_y^2 \left(\frac{s_x^2}{S_X^2} \right)^{-1.0075} \left\{ 1.0352 - 0.0352 \left(\frac{s_x^2 A + B}{S_X^2 A + B} \right)^{1.6711} \right\} + 24.17231.$$

5 Conclusion

In this section we illustrate the algorithm of computational approach and we will discuss about the results. As previously mentioned, the optimum values of \hat{S}_{F2}^2 estimator parameters can be obtained by using last census data in this algorithm, and then we can use this estimator for the sampling in the following years. The steps of algorithm are as follows:

1. Take k simple random samples of size n from the population, such that we have an array, say D , with k matrices of $n \times 2$,
2. give some initial values to the parameters of \hat{S}_{F2}^2 , i.e. δ, w, ε . Note that by these values, k estimated variances for samples in D can be obtained and hence the *EASE* can be calculated,
3. use an optimization algorithm e.g. `nlm` function in R, with *EASE* as objective function, initial values of step 2 and array D as data, to get the optimum values of parameters δ, w, ε .
4. Using the optimum values of parameters in step 3, the estimator \hat{S}_{F2}^2 can be calculated for k samples in D , by which the estimated bias, $\hat{B}(\hat{S}_{F2}^2)$, can be obtained.

Therefor if another census data is used, the optimal values obtained by the above steps applying the census data can be used to get the approximated unbiased estimator of variance.

Further work could be considered as

1. The optimum values of parameters depend on the number of simulation k , so k is needed to be large enough.
2. In this study we consider simple random sampling design, but in later researches we can use the other sampling designs.
3. The performance of `nlinb` function depends on the initial values of parameters, so it should be carefully chosen.
4. On the other hand, the optimum values of parameters also depend on the value of n and the optimum value of $EASE$ will certainly be reduced by increasing n , but the value of n should be equal to the used sample size in the sampling design.

References

- Diana, G. (1993). A Class of Estimators of the Population Mean in Stratified Random Sampling. *Statistica*, **1**, 59-66.
- Isaki, C. (1983). Variance Estimation Using Auxiliary Information. *Journal of the American Statistical Association*, **78**, 117-123.
- Kadilar, C. and Cingi, H. (2006). Ratio Estimator for the Population Variance in Simple and Stratified Random Sampling. *Applied Mathematics and Computation*, **173**, 1047-1059.
- Kendall, M. and Stuart, A. (1969). *The Advanced Theory of Statistics*, Volume 1. Hafner, New York.
- Khoshnevisan, M., Singh, R. and Smarandache, F. (2007). A General Family of Estimators for Estimating Population Mean Using Known Value of Some Population Parameters. *Technical Report*, <http://arxiv.org/pdf/math/>.
- Koyuncu, N. and Kadilar, C. (2010). On the Family of Estimators of Population Mean in Stratified Random Sampling. *Pakistan Journal of Statistics*, **26**, 427-443.

Mohammadreza Faghihi

Department of Statistics,
Shahid Beheshti University,
Tehran, Iran.

email: m.faghihi@sbu.ac.ir

