Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v11n1p196

Unbiased estimator of finite population variance based on a ratio type estimator

By Al-Jararha, Aljadeed

Published: 26 April 2018

Electronic Journal of Applied Statistical Analysis
Vol. 11, Issue 01, April 2018, 196-207
DOI: 10.1285/i20705948v11n1p196

Unbiased estimator of finite population variance based on a ratio type estimator

Jehad Al-Jirarha*and A. Aljadeed

Yarmouk University, Department of Statistics
Irbid, Jordan

Published: 26 April 2018

In this paper, an estimator of the finite population variance S_y^2 is proposed. The proposed estimator is exactly unbiased estimator for S_y^2; further, the mean squared error (MSE) of the proposed estimator is derived. Empirical studies from real data sets are used to compare the proposed estimator and other estimators proposed in the literature. The proposed estimator is stable among other families and practically has minimum MSE among other estimators.

keywords: Hartely and Ross Estimator, Ratio Type Estimators, Finite Population Variance, Auxiliary Variable, Mean Squared Error, Simple Random Sample.

1 Introduction

Consider the finite population U of N units indexed by the set $\{1, 2, \ldots, N\}$. For the ith unit, let y_i be the value of the interest variable Y, and x_i be the value of the auxiliary variable X. For $i = 1, \ldots, N$, define $r_i = y_i/x_i$ be the values of the variable R. Under simple random sample without replacement (SRSWOR) design, draw a random sample s of size n from U. The i^{th} unit $(y_i, x_i, r_i) \in s$ is assumed to be known.

The finite population ratio θ_{yx} is defined by

$$\begin{aligned} \theta_{yx} &= \frac{\bar{y}_u}{\bar{x}_u} \\ &= \frac{t_y}{t_x} \end{aligned} \quad (1)$$

where $t_y = \sum_{i=1}^{N} y_i$ and $\bar{y}_u = t_y/N$ are the finite population total and mean for the variable of interest Y, respectively. Furthermore, $t_x = \sum_{i=1}^{N} x_i$ and $\bar{x}_u = t_x/N$ are the finite population total and mean for the auxiliary variables X, respectively.

*Corresponding author: jjararha@gmail.com

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

A well known exactly unbiased estimator for θ_{yx} is the HR estimator proposed by Hartley and Ross (1954) and is given by

$$\hat{\theta}_{yx} = \bar{r}_s + \frac{n\,(N-1)}{N\,(n-1)\,\bar{x}_u}\left(\bar{y}_s - \bar{r}_s\bar{x}_s\right), \tag{2}$$

where $\bar{x}_u$ is assumed to be known. $\bar{y}_s$, $\bar{x}_s$, and $\bar{r}_s$ are the sample means of Y, X, and R, respectively.

The HR estimator has been used and generalized in different directions. Recently, many authors are interested in the HR estimator. Under general sampling design Al-Jararha (2008) has generalized the HR estimator. When the information of the auxiliary variable are available, Rao and Swain (2014) has constructed an alternative HR unbiased estimator for $\bar{y}_u$. Other authors have been considering this estimator in different directions.

To our knowledge, the HR is used to estimate the finite population ratio and can be modified to estimate parameters can be written in term of the population ratio. Our aim is to adopt the HR estimator for estimating S_y^2.

In this paper, the main notations are given in Section 2. Section 3 is devoted to the literature review. The proposed estimator and the main results are given in Section 4. Section 5 and Section 6 are devoted to the empirical studies and Concluding Remarks, respectively.

2 Notations

Let $S_y^2 = \sum_{i=1}^{N}\left(y_i - \bar{y}_u\right)^2/\left(N-1\right)$, $S_x^2 = \sum_{i=1}^{N}\left(x_i - \bar{x}_u\right)^2/\left(N-1\right)$ and $S_r^2 = \sum_{i=1}^{N}\left(r_i - \bar{r}_u\right)^2/\left(N-1\right)$ be the finite population variances for Y, X, and R, respectively. The covariance between X and Y is defined by $S_{yx} = \sum_{i=1}^{N}\left(y_i - \bar{y}_u\right)\left(x_i - \bar{x}_u\right)/\left(N-1\right)$. Further, define $C_y = S_y/\bar{y}_u$, $C_x = S_x/\bar{x}_u$, and $C_{yx} = S_{yx}/\bar{y}_u\bar{x}_u$. On the sample level, let $s_y^2 = \sum_{i=1}^{n}\left(y_i - \bar{y}_s\right)^2/\left(n-1\right)$, $s_x^2 = \sum_{i=1}^{n}\left(x_i - \bar{x}_s\right)^2/\left(n-1\right)$, and $s_r^2 = \sum_{i=1}^{n}\left(r_i - \bar{r}_s\right)^2/\left(n-1\right)$ be the sample variances of the variables Y, X, and R, respectively. The sample covariance between X and Y is $s_{yx} = \sum_{i=1}^{n}\left(y_i - \bar{y}_s\right)\left(x_i - \bar{x}_s\right)/\left(n-1\right)$. Similar formulas can be defined for other variables.

Let

$$\xi_x = \tfrac{\bar{x}_s - \bar{x}_u}{\bar{x}_u}, \quad \xi_y = \tfrac{\bar{y}_s - \bar{y}_u}{\bar{y}_u} \qquad \text{and} \qquad \xi_r = \tfrac{\bar{r}_s - \bar{r}_u}{\bar{r}_u}$$

Therefore,

$$E\left(\xi_x\right) = E\left(\xi_y\right) = E\left(\xi_r\right) = 0$$

and

$$E\left(\xi_x^2\right) = \tfrac{1-f}{n}C_{xx}, \qquad E\left(\xi_y^2\right) = \tfrac{1-f}{n}C_{yy}, \quad E\left(\xi_r^2\right) = \tfrac{1-f}{n}C_{rr}, \quad E\left(\xi_y\xi_x\right) = \tfrac{1-f}{n}C_{yx},$$
$$\text{and } E\left(\xi_y\xi_r\right) = \tfrac{1-f}{n}C_{yr},$$

where $f = n/N$. Furthermore, define the following terms:

$$\beta^*_{2(x)} = \beta_{2(x)} - 1, \qquad \beta^*_{2(y)} = \beta_{2(y)} - 1, \qquad \beta_{2(x)} = \tfrac{\mu_{x.4}}{\mu^2_{x.2}},$$
$$\beta_{2(y)} = \tfrac{\mu_{y.4}}{\mu^2_{y.2}}, \qquad \lambda^*_{wv.st} = \lambda_{wv.st} - 1, \qquad \lambda_{wv.st} = \tfrac{\mu_{wv.st}}{\mu^{s/2}_{w.2}\mu^{t/2}_{v.2}},$$
$$\mu_{w.s} = \tfrac{1}{N}\sum_{i=1}^{N}\left(w_i - \bar{w}_u\right)^s, \quad \text{and} \quad \mu_{wv.st} = \tfrac{1}{N}\sum_{i=1}^{N}\left(w_i - \bar{w}_u\right)^s\left(v_i - \bar{v}_u\right)^t,$$

for s and t being non-negative integers.

3 Literature Review

The traditional sample variance s_y^2 is an unbiased estimator for the population variance S_y^2. However, this estimator dose not use the availability of the auxiliary information. The MSE of s_y^2 is

$$\begin{aligned} MSE\left(s_y^2\right) &= var\left(s_y^2\right) \\ &= \frac{1}{n}S_y^4\beta^*_{2(y)}. \end{aligned} \tag{3}$$

Gupta (1983) considered the two sampling designs, SRSWOR and simple random sample with replacement (SRSWR) design, and proposed unbiased estimators for estimating S_y^2, the estimators are using the availability of the auxiliary information. Further, Gupta (1983) considered two cases: when $\bar{x}_u$ is known and $\bar{x}_u$ and S_x^2 are known. When the design is SRSWOR and the case $\bar{x}_u$ and S_x^2 are known, the estimator is defined by

$$s^2_{gpta} = \frac{n(N-1)}{N(n-1)}\left[\frac{N-1}{N}s_y^2 + \bar{v}S_x^2 + \frac{(N-n)}{N(n-1)}\bar{y}_s^2 - \frac{n(N-1)}{N}s_x^2 - \frac{n^2(N-1)}{N(n-1)}\left(\bar{x}_s - \bar{x}_u\right)^2\right], \tag{4}$$

where $v_i = y_i^2/\left(x_i - \bar{x}_u\right)^2$, and $\bar{v} = \sum_{i=1}^n v_i/n$. However, the MSE was not reported for this estimator.

Prasad and Singh (1992) proposed a family of estimators for estimating S_y^2. The proposed family is defined by

$$s^2_{pras} = s_y^2 - a\frac{s_x^2}{S_x^2} + a. \tag{5}$$

Where a, is a constant or function of known parameters of the auxiliary variable, the family s^2_{pras} is unbiased for S_y^2 with

$$\begin{aligned} var\left(s^2_{pras}\right) &= var\left(s_y^2\right) + a^2\frac{var\left(s_x^2\right)}{S_x^4} - 2a\frac{cov\left(s_y^2,\, s_x^2\right)}{S_x^2} \\ &= \frac{1}{n}S_y^4\beta^*_{2(y)} + \frac{a^2}{n}\beta^*_{2(x)} - 2\frac{a}{n}S_y^2\lambda^*_{yx.22}. \end{aligned} \tag{6}$$

However, Prasad and Singh (1992) considered $a = 1$.

Singh and Solanki (2013) summarized and proposed classes of estimators for estimating S_y^2. Different estimators and classes of estimators are proposed for estimating S_y^2. The class of estimators

$$t = s_y^2\left(\frac{aS_x^2 + b}{as_x^2 + b}\right) \tag{7}$$

is proposed to estimate S_y^2, where a and b are either constants or functions of known parameters of the auxiliary variable X. The minimum MSE of the class of estimators t is given by

$$\begin{aligned} MSE_{min}(t) &= \frac{1}{n}S_y^4\beta^*_{2(y)}\left(1 - \rho^{*2}_{S_y^2S_x^2}\right) \\ &= var\left(s_y^2\right)\left(1 - \rho^{*2}_{S_y^2S_x^2}\right), \end{aligned} \tag{8}$$

where

$$\rho^*_{S_y^2S_x^2} = \frac{\lambda^*_{yx.22}}{\sqrt{\beta^*_{2(x)}\beta^*_{2(y)}}} \tag{9}$$

is the correlation coefficient between s_x^2 and s_y^2.

Estimators t_i for $i = 0, 1, \ldots, 9$ discussed by Singh and Solanki (2013) are members of the t class. Such estimators are given in Table (1).

Table 1: Different estimators for estimating S_y^2.

Estimator	Remark
$t_0 = s_y^2$	The sample variance estimator.
$t_1 = s_y^2 \left(\frac{S_x^2}{s_x^2}\right)$	Isaki (1983).
$t_2 = s_y^2 \left(\frac{S_x^2+\beta_{2(x)}}{s_x^2+\beta_{2(x)}}\right)$	Upadhyaya and Singh (1996)
$t_3 = s_y^2 \left(\frac{S_x^2-C_x}{s_x^2-C_x}\right)$	Kadilar and Cingi (2006)
$t_4 = s_y^2 \left(\frac{S_x^2-\beta_{2(x)}}{s_x^2-\beta_{2(x)}}\right)$	Kadilar and Cingi (2006)
$t_5 = s_y^2 \left(\frac{S_x^2\beta_{2(x)}-C_x}{s_x^2\beta_{2(x)}-C_x}\right)$	Kadilar and Cingi (2006)
$t_6 = s_y^2 \left(\frac{S_x^2 C_x-\beta_{2(x)}}{s_x^2 C_x-\beta_{2(x)}}\right)$	Kadilar and Cingi (2006)
$t_7 = s_y^2 \left(\frac{S_x^2+C_x}{s_x^2+C_x}\right)$	Singh and Solanki (2013)
$t_8 = s_y^2 \left(\frac{S_x^2\beta_{2(x)}+C_x}{s_x^2\beta_{2(x)}+C_x}\right)$	Singh and Solanki (2013)
$t_9 = s_y^2 \left(\frac{S_x^2 C_x+\beta_{2(x)}}{s_x^2 C_x+\beta_{2(x)}}\right)$	Singh and Solanki (2013)

The difference type estimator

$$t_d = s_y^2 + \omega_2 \left(S_x^2 - s_x^2\right), \tag{10}$$

where ω_2 is suitably chosen constant, has the same minimum MSE as the t class, i.e. the

$$MSE_{min}\left(t_d\right) = MSE_{min}\left(t\right),$$

and $MSE_{min}\left(t\right)$ is defined by Eq. (8).

The class of estimators for estimating S_y^2 proposed by Singh et al. (1988) defined by

$$t_s = \omega_1 s_y^2 + \omega_2 \left(S_x^2 - s_x^2\right), \tag{11}$$

where ω_1 and ω_2 are suitably chosen constants. The minimum MSE of t_s is given by

$$MSE_{min}\left(t_s\right) = \frac{MSE_{min}\left(t_d\right)}{1 + \left(MSE_{min}\left(t_d\right)/S_y^4\right)} \tag{12}$$

where $MSE_{min}\left(t_d\right)$ is defined by Eq.(8).

Singh and Solanki (2013) proposed the following classes of estimators

$$t_{sg} = \left[\omega_1 s_y^2 - \omega_2 \left(\frac{s_x^2 - S_x^2}{S_x^2}\right)\right] \left(\frac{aS_x^2 + b}{as_x^2 + b}\right), \tag{13}$$

and

$$T=\omega_1 s_y^2\left[\frac{cS_x^2-ds_x^2}{(c-d)\,S_x^2}\right]^{\gamma}+\omega_2 s_y^2\left[\frac{(a+b)\,S_x^2}{aS_x^2+bs_x^2}\right]^{\delta}, \tag{14}$$

for estimating S_y^2, where ω_1 and ω_2 are suitably chosen constants and a, b, c, and d are either constants or functions of known parameters of the auxiliary variable X. The minimum MSE of t_{sg} and T are given by

$$MSE_{min}\left(t_{sg}\right)=\frac{n^{-1}S_y^4\beta^*_{2(y)}\left(1-\rho^{*2}_{S_y^2S_x^2}\right)\left(1-n^{-1}\eta^{*2}\beta^*_{2(x)}\right)}{1+n^{-1}\left[\beta^*_{2(y)}\left(1-\rho^{*2}_{S_y^2S_x^2}\right)-\eta^{*2}\beta^*_{2(x)}\right]} \tag{15}$$

and

$$MSE_{min}\left(T\right)=S_y^4\left[1-\frac{BD^2-2CDE+AE^2}{AB-C^2}\right], \tag{16}$$

respectively, where

$$A = 1+\frac{1}{n}\left[\beta^*_{2(y)}+\gamma\eta_2\beta^*_{2(x)}\left(\eta_2\left(2\gamma-1\right)-4\lambda\right)\right], \tag{17}$$

$$B = 1+\frac{1}{n}\left[\beta^*_{2(y)}+\delta\eta_1\beta^*_{2(x)}\left(\eta_1\left(2\delta+1\right)-4\lambda\right)\right], \tag{18}$$

$$C = 1+\frac{1}{n}\left[\beta^*_{2(y)}+\beta^*_{2(x)}\left(\frac{\eta}{2}-2\left(\delta\eta_1+\gamma\eta_2\right)\lambda\right)\right], \tag{19}$$

$$D = 1+\frac{1}{n}\gamma\eta_2\beta^*_{2(x)}\left(\frac{\eta_2\left(\gamma-1\right)}{2}-\lambda\right), \tag{20}$$

$$E = 1+\frac{1}{n}\delta\eta_1\beta^*_{2(x)}\left(\frac{\left(\delta+1\right)\eta_1}{2}-\lambda\right), \tag{21}$$

$$\eta_1 = \frac{b}{a+b}, \tag{22}$$

$$\eta_2 = \frac{d}{c-d}, \tag{23}$$

$$\eta^* = \frac{aS_x^2}{aS_x^2+b}, \tag{24}$$

$$\lambda = \frac{\lambda^*_{xy.22}}{\beta^*_{2(x)}}. \tag{25}$$

The estimators $t_{sg1},\ldots,t_{sg10}$ are members from the t_{sg} class and defined in Table(1), Singh and Solanki (2013) and $T_1,\ldots,T_{10}$ are members from the T class and defined in Table(2) from the same reference.

Under the constrain $\omega_1+\omega_2=1$, the class of estimators T reduces to the new class of estimators

$$T^*=\omega_1 s_y^2\left[\frac{cS_x^2-ds_x^2}{(c-d)\,S_x^2}\right]^{\gamma}+\left(1-\omega_1\right)s_y^2\left[\frac{(a+b)\,S_x^2}{aS_x^2+bs_x^2}\right]^{\delta}, \tag{26}$$

with minimum MSE given by Eq.(8).

Under two-phase sampling and based on the auxiliary variable in the presence of non-response, Bhushan (2012) proposed a class of modified exponential-ratio type estimators for estimating the population mean. Shabbir and Khan (2013) the prior information about the auxiliary variable are used to introduce ideas for estimating the mean of the variable Y, unbiasedly.

Most of the classes of estimators mentioned in this summary, are originally proposed to estimate the finite population mean, total, and ratio modified to estimate S_y^2. To our knowledge, none of them adopted the approach of well known behaves estimator proposed by Hartley and Ross (1954) for estimating the population ratio θ_{yx}. The proposed estimator will modify the HR estimator to estimate S_y^2.

4 Proposed Estimator

From Eq. (2), rewrite the HR estimator in terms ξ_x, ξ_y, and ξ_r, we have

$$\hat{\theta}_{yx} = \bar{r}_u + \frac{n(N-1)}{N(n-1)\bar{x}_u}(\bar{y}_u - \bar{r}_u\bar{x}_u) + a_y\xi_y + a_r\xi_r + a_x\xi_x(1+\xi_r), \tag{27}$$

where $a_y = \frac{n(N-1)}{N(n-1)}\frac{\bar{y}_u}{\bar{x}_u}$, $a_r = -\frac{N-n}{N(n-1)}\bar{r}_u$, and $a_x = -\frac{n(N-1)}{N(n-1)}\bar{r}_u$. It is known that $\hat{\theta}_{yx}$ is exactly unbiased estimator for θ_{yx}. This is also clear from Eq. (27), since

$$\begin{aligned} E(\xi_x\xi_r) &= \frac{1-f}{n}C_{xr} && (28)\\ &= \frac{1-f}{n}\frac{N}{N-1}\frac{\bar{y}_u - \bar{r}_u\bar{x}_u}{\bar{r}_u\bar{x}_u} && (29)\end{aligned}$$

Therefore,

$$E\left(\hat{\theta}_{yx}\right) = \bar{r}_u + \frac{n(N-1)}{N(n-1)\bar{x}_u}(\bar{y}_u - \bar{r}_u\bar{x}_u) + a_x\frac{1-f}{n}\frac{N}{N-1}\frac{\bar{y}_u - \bar{r}_u\bar{x}_u}{\bar{r}_u\bar{x}_u} = \bar{y}_u/\bar{x}_u = t_y/t_x.$$

In this case, the MSE of $\hat{\theta}_{yx}$ is the same as the variance of $\hat{\theta}_{yx}$. Therefore, taking the variance of both sides of Eq. (27), we have

$$\begin{aligned} MSE\left(\hat{\theta}_{yx}\right) &= Var\left(\hat{\theta}_{yx}\right)\\ &= \frac{1-f}{n}\left\{a_y^2C_{yy} + a_r^2C_{rr} + a_x^2C_{xx} + 2a_xa_yC_{xy} + 2a_xa_rC_{xr} + 2a_ya_rC_{yr}\right\}\\ &= \frac{1-f}{n}\left\{a_y^2C_{yy} + a_r^2C_{rr} + a_x^2C_{xx}\right.\\ &+ \left. 2a_xa_yC_{xy} + 2a_ya_rC_{yr} + 2\frac{N}{N-1}\frac{\bar{y}_u - \bar{r}_u\bar{x}_u}{\bar{r}_u\bar{x}_u}a_xa_r\right\}. && (30)\end{aligned}$$

Remark 4.1. *The ratio estimator for estimating $\bar{y}_u$ based on the HR estimator is*

$$\bar{y}_{hr} = \bar{x}_u\bar{r}_s + \frac{n(N-1)}{N(n-1)}(\bar{y}_s - \bar{r}_s\bar{x}_s). \tag{31}$$

This estimator is exactly unbiased. The MSE of $\bar{y}_{hr}$ is given by

$$MSE(\bar{y}_{hr}) = \bar{x}_u^2 MSE\left(\hat{\theta}_{yx}\right), \tag{32}$$

where $MSE\left(\hat{\theta}_{yx}\right)$ is given by Eq. (30).

The ratio parameter of our interest is

$$\begin{aligned}\theta_{S_y^2 S_x^2} &= \frac{S_y^2}{S_x^2} \\ &= \frac{\sum_{i=1}^{N}\left(y_i-\bar{y}_u\right)^2/\left(N-1\right)}{\sum_{i=1}^{N}\left(x_i-\bar{x}_u\right)^2/\left(N-1\right)} \\ &= \frac{t_{\breve{y}}}{t_{\breve{x}}} \qquad (33)\end{aligned}$$

which is a function of the population totals $t_{\breve{y}}$ and $t_{\breve{x}}$, where $t_{\breve{y}} = \sum_{i=1}^{N}\breve{y}_i = \sum_{i=1}^{N}\left(y_i-\bar{y}_u\right)^2$ and $t_{\breve{x}} = \sum_{i=1}^{N}\breve{x}_i = \sum_{i=1}^{N}\left(x_i-\bar{x}_u\right)^2$.

Therefore, to estimate $\theta_{S_y^2 S_x^2}$ based on the HR approach, substitute $\left(y_i-\bar{y}_s\right)^2$, $\left(x_i-\bar{x}_s\right)^2$, and $\left(y_i-\bar{y}_s\right)^2/\left(x_i-\bar{x}_s\right)^2$ in Eq. (2) instead of y_i, x_i, and r_i, respectively. Hence

$$\widehat{\theta}_{S_y^2 S_x^2} = \bar{\breve{r}}_s + \frac{1}{S_x^2}\left(s_y^2 - \bar{\breve{r}}_s s_x^2\right) \qquad (34)$$

is the HR approach for estimating the finite population ratio $\theta_{S_y^2 S_x^2}$, where $\bar{\breve{r}}_s = \sum_{i=1}^{n}\left[\left(y_i-\bar{y}_s\right)^2/\left(x_i-\bar{x}_s\right)^2\right]/n$ for all $x_i \in s$ and $x_i \neq \bar{x}_s$. The estimator $\widehat{\theta}_{S_y^2 S_x^2}$ is exactly unbiased estimator for $\theta_{S_y^2 S_x^2}$.

The finite population variance S_y^2 can be estimated by the proposed estimator

$$s_{yhr}^2 = \frac{N-1}{N} S_x^2 \widehat{\theta}_{S_y^2 S_x^2},$$

hence

$$\begin{aligned} s_{yhr}^2 &= \frac{N-1}{N} S_x^2 \widehat{\theta}_{S_y^2 S_x^2} \\ &= \frac{N-1}{N}\left[s_y^2 + \bar{\breve{r}}_s\left(S_x^2 - s_x^2\right)\right] \qquad (35)\end{aligned}$$

be the proposed estimator based on HR approach for estimating S_y^2, where S_x^2 is assumed to be known.

Eq. (32) can be used to obtain the MSE of s_{yhr}^2 by using $\breve{x}$, $\breve{y}$, and $\breve{r}$ instead of x, y, and r, respectively. Therefore,

$$\begin{aligned} MSE\left(s_{yhr}^2\right) &= Var\left(s_{yhr}^2\right) \\ &= \frac{1-f}{n}\left(\frac{N-1}{N}\right)^2 S_x^4\left\{a_{\breve{y}}^2 C_{\breve{y}\breve{y}} + a_{\breve{r}}^2 C_{\breve{r}\breve{r}} + a_{\breve{x}}^2 C_{\breve{x}\breve{x}}\right. \\ &+ \left. 2a_{\breve{x}}a_{\breve{y}}C_{\breve{x}\breve{y}} + 2a_{\breve{x}}a_{\breve{r}}C_{\breve{x}\breve{r}} + 2a_{\breve{y}}a_{\breve{r}}C_{\breve{y}\breve{r}}\right\}. \qquad (36)\end{aligned}$$

Where

$a_{\breve{y}} = \frac{n(N-1)}{N(n-1)}\frac{S_y^2}{S_x^2}$ $\quad a_{\breve{r}} = -\frac{N-n}{N(n-1)}\bar{\breve{r}}_u$ $\quad a_{\breve{x}} = -\frac{n(N-1)}{N(n-1)}\bar{\breve{r}}_u$

$\bar{\breve{y}}_u = \mu_{y.2}$ $\quad \bar{\breve{x}}_u = \mu_{x.2}$ $\quad \bar{\breve{r}}_u = \frac{1}{N}\sum_{i=1}^{N}\frac{(y_i-\bar{y}u)^2}{(x_i-\bar{x}u)^2}$

$$C_{\breve{x}\breve{x}} = \frac{N}{N-1}\beta^*_{2(x)},\quad C_{\breve{y}\breve{y}} = \frac{N}{N-1}\beta^*_{2(y)},\quad C_{\breve{r}\breve{r}} = \frac{N}{N-1}\left[\frac{\frac{1}{N}\sum_{i=1}^{N}\frac{(y_i-\bar{y}u)^4}{(x_i-\bar{x}u)^4}}{\bar{\breve{r}}_u^2} - 1\right],$$

$$C_{\breve{y}\breve{x}} = \frac{N}{N-1}\lambda^*_{yx.22},\quad C_{\breve{y}\breve{r}} = \frac{N}{N-1}\left[\frac{\frac{1}{N}\sum_{i=1}^{N}\frac{(y_i-\bar{y}u)^4}{(x_i-\bar{x}u)^2}}{\bar{\breve{r}}_u\mu_{y.2}} - 1\right],\quad \text{and}\quad C_{\breve{x}\breve{r}} = \frac{N}{N-1}\left[\frac{\mu_{y.2}}{\bar{\breve{r}}_u\mu_{x.2}} - 1\right].$$

Analytically, it is not an easy task to compare the MSE's of the other estimators mentioned in this article and the proposed approach. Therefore, the empirical MSE will be computed from real data sets for the proposed estimator and other estimators mentioned in Section (3).

5 Empirical Studies

The comparisons between the families of estimators discussed by Singh and Solanki (2013), Prasad and Singh (1992) and the proposed approach will be made based on the empirical relative MSEs studies from real data sets.

Remark 5.1. *The empirical relative efficiency for each family is the ratio of the maximum MSE to the minimum MSE of the estimators in that family. Further, the empirical relative efficiency of the proposed estimator is the ratio of* $MSE\left(s^2_{yhr}\right)$ *to the minimum MSE of each family. The empirical relative efficiencies are given in the boxes at the end of Tables (2), (3) and (4).*

Consider the Apple data set considered by Kadilar and Cingi (2007). The summary of the data set as reported by Singh and Solanki (2013) is given in the following table.

$N = 104$	$n = 20$	$f = 0.192$	$\beta_{2(y)} = 16.532$	$\lambda_{yx.22} = 14.398$
$\bar{Y} = 6.254$	$S_y = 11.670$	$C_y = 1.866$	$\beta_{2(x)} = 17.516$	$\rho^*_{s_y^2 s_x^2} = 0.837$
$\bar{X} = 13931.683$	$S_x = 23026.133$	$C_x = 1.653$	$\rho_{yx} = 0.865$	$\lambda = 0.811$

To compute the MSE of the proposed approach, the complete data set is needed to define the R variable. However, the results are summarized in Table (2) and will be used in the comparisons.

Consider the Loblolly data set included in the R library data sets package. The variable of interest Y is the tree heights (ft) and the auxiliary variable X is the tree ages (year). The Loblolly data set consists from $N = 84$ observations. The sample size ($n = 16$) gives approximately the same $f = 0.192$ for the Apple data set. The Loblolly data set is summarized by the following table.

$N = 84$	$n = 16$	$f = 0.1905$	$\bar{Y} = 32.3644$	$\bar{X} = 13$
$S_y^2 = 427.3979$	$S_x^2 = 62.4096$	$C_y = 0.6388$	$C_x = 0.6077$	$\rho_{yx} = 0.9899$
$\rho^*_{s_y^2 s_x^2} = 0.9366$	$\beta^*_{2(y)} = 0.5422$	$\beta^*_{2(x)} = 0.6361$	$\lambda^*_{yx.22} = 0.5500$	$\bar{\breve{r}}_u = 8.1881$
$C_{\breve{r}\breve{r}} = 0.4907$	$C_{\breve{y}\breve{r}} = -0.0871$	$C_{\breve{x}\breve{r}} = -0.1656$		

The computations from Loblolly data set are summarized in Table (3).

Table 2: Computations from Apple data sets when ($n = 20$).

t family Defined by Eq(7)	t_{sg} family Defined by Eq(13)	T family Defined by Eq(14)	estimator
$MSE\left(t_0\right) = \underbrace{14395.577}_{max}$	$MSE\left(t_{sg1}\right) = \underbrace{1847.70}_{min}$	$MSE\left(T_1\right) = 672.903$	$MSE_{min}\left(t\right) = 4316.321$
$MSE\left(t_1\right) = 4862.205$	$MSE\left(t_{sg2}\right) = 1847.799$	$MSE\left(T_2\right) = \underbrace{77.054}_{min}$	$MSE\left(t_d\right) = 4310.482$
$MSE\left(t_2\right) = 4862.205$	$MSE\left(t_{sg3}\right) = 1847.799$	$MSE\left(T_3\right) = 79.434$	$MSE\left(t_s\right) = 3497.623$
$MSE\left(t_3\right) = 4862.205$	$MSE\left(t_{sg4}\right) = 1847.799$	$MSE\left(T_4\right) = 751.039$	$MSE\left(t_g\right) = 2615.177$
$MSE\left(t_4\right) = 4862.205$	$MSE\left(t_{sg5}\right) = 1847.799$	$MSE\left(T_5\right) = 1350.204$	$MSE\left(T^*\right) = 4316.321$
$MSE\left(t_5\right) = 4862.205$	$MSE\left(t_{sg6}\right) = \underbrace{1848.469}_{max}$	$MSE\left(T_6\right) = \underbrace{1534.483}_{max}$	$MSE\left(s^2_{pras}\right) = 14393.404$
$MSE\left(t_6\right) = 4862.205$	$MSE\left(t_{sg7}\right) = 1847.799$	$MSE\left(T_7\right) = 1438.673$	
$MSE\left(t_7\right) = 4862.205$	$MSE\left(t_{sg8}\right) = 1847.799$	$MSE\left(T_8\right) = 966.107$	
$MSE\left(t_8\right) = 4862.205$	$MSE\left(t_{sg9}\right) = 1847.813$	$MSE\left(T_9\right) = 717.981$	
$MSE\left(t_9\right) = 4862.205$	$MSE\left(t_{sg10}\right) = 1847.799$	$MSE\left(T_{10}\right) = 911.452$	
$\frac{MSE(t_0)}{MSE(t_1)} = 2.9607$	$\frac{MSE(t_{sg6})}{MSE(t_{sg1})} = 1.0004$	$\frac{MSE(T_6)}{MSE(T_2)} = 19.9144$	

Table 3: Computations from Loblolly data sets when ($n = 16$).

t family Defined by Eq(7)	t_{sg} family Defined by Eq(13)	T family Defined by Eq(14)	estimator
$MSE\left(t_0\right) = \underbrace{6190.2}_{max}$	$MSE\left(t_{sg1}\right) = 757.26$	$MSE\left(T_1\right) = \underbrace{707.17}_{min}$	$MSE_{min}\left(t\right) = 760.54$
$MSE\left(t_1\right) = 893.51$	$MSE\left(t_{sg2}\right) = 813.07$	$MSE\left(T_2\right) = 748.47$	$MSE\left(t_d\right) = 760.54$
$MSE\left(t_2\right) = 848.05$	$MSE\left(t_{sg3}\right) = 690.9$	$MSE\left(T_3\right) = 758.1$	$MSE\left(t_s\right) = 757.38$
$MSE\left(t_3\right) = 913.54$	$MSE\left(t_{sg4}\right) = 824.57$	$MSE\left(T_4\right) = 1194.58$	$MSE\left(t_g\right) = 745.36$
$MSE\left(t_4\right) = 951.69$	$MSE\left(t_{sg5}\right) = \underbrace{838.2}_{max}$	$MSE\left(T_5\right) = 760.35$	$MSE\left(T^*\right) = 760.54$
$MSE\left(t_5\right) = 864.32$	$MSE\left(t_{sg6}\right) = 534.66$	$MSE\left(T_6\right) = 7985.42$	$MSE\left(s^2_{pras}\right) = 6160.86$
$MSE\left(t_6\right) = 996.88$	$MSE\left(t_{sg7}\right) = 691.69$	$MSE\left(T_7\right) = 747.79$	$MSE\left(s^2_{yhr}\right) = 240.92$
$MSE\left(t_7\right) = 875.24$	$MSE\left(t_{sg8}\right) = 835.96$	$MSE\left(T_8\right) = 746.50$	
$MSE\left(t_8\right) = 882.14$	$MSE\left(t_{sg9}\right) = \underbrace{254.01}_{min}$	$MSE\left(T_9\right) = 758.53$	
$MSE\left(t_9\right) = \underbrace{824.65}_{min}$	$MSE\left(t_{sg10}\right) = 823.74$	$MSE\left(T_{10}\right) = \underbrace{12088.74}_{max}$	
$\frac{MSE(t_0)}{MSE(t_9)} = 7.5065$ $\frac{MSE\left(s^2_{yhr}\right)}{MSE(t_9)} = 0.2922$	$\frac{MSE(t_{sg5})}{MSE(t_{sg9})} = 3.2999$ $\frac{MSE\left(s^2_{yhr}\right)}{MSE(t_{sg9})} = 0.9485$	$\frac{MSE(T_{10})}{MSE(T_1)} = 17.0945$ $\frac{MSE\left(s^2_{yhr}\right)}{MSE(T_1)} = 0.3407$	$\frac{MSE\left(s^2_{yhr}\right)}{MSE(t_g)} = 0.3232$

Consider the data set given by Cochran (1977), Table 6.9, Page 182. Let Y be numbers of not inoculated children and X be numbers of placebo. The data is summarized by the following table.

$N = 34$	$n = 10$	$f = 0.2941$	$\bar{Y} = 8.3706$	$\bar{X} = 4.9235$
$S_y^2 = 74.0428$	$S_x^2 = 25.3855$	$C_y = 1.028$	$C_x = 1.0233$	$\rho_{yx} = 0.9014$
$\rho^*_{s_y^2 s_x^2} = 0.58$	$\beta^*_{2(y)} = 7.9325$	$\beta^*_{2(x)} = 5.3911$	$\lambda^*_{yx.22} = 4.9802$	$\bar{\breve{r}}_u = 4.6354$
$C_{\breve{r}\breve{r}} = 3.6483$	$C_{\breve{y}\breve{r}} = 8.1729$	$C_{\breve{x}\breve{r}} = -0.382$		

The computations are summarized in Table (4).

Table 4: Computations for Table 6.9, Cochran (1977), when ($n = 10$).

t family Defined by Eq(7)	t_{sg} family Defined by Eq(13)	T family Defined by Eq(14)	estimator
$MSE(t_0) = \underbrace{4348.85}_{max}$	$MSE(t_{sg1}) = 1211.36$	$MSE(T_1) = 451.83$	$MSE_{min}(t) = 1826.69$
$MSE(t_1) = 1843.86$	$MSE(t_{sg2}) = 1227.44$	$MSE(T_2) = 838.4$	$MSE(t_d) = 1826.69$
$MSE(t_2) = 1872.79$	$MSE(t_{sg3}) = 1215.5$	$MSE(T_3) = 1243.05$	$MSE(t_s) = 1370.16$
$MSE(t_3) = 1868.00$	$MSE(t_{sg4}) = \underbrace{1194.85}_{min}$	$MSE(T_4) = 1079.92$	$MSE(t_g) = 1166.82$
$MSE(t_4) = 2330.09$	$MSE(t_{sg5}) = 1195.89$	$MSE(T_5) = \underbrace{1292.14}_{max}$	$MSE(T^*) = 1826.69$
$MSE(t_5) = 1831.03$	$MSE(t_{sg6}) = \underbrace{1241.08}_{max}$	$MSE(T_6) = \underbrace{196.63}_{min}$	$MSE\left(s^2_{pras}\right) = 4275.64$
$MSE(t_6) = 2305.57$	$MSE(t_{sg7}) = 1222.27$	$MSE(T_7) = 1272.25$	$MSE\left(s^2_{yhr}\right) = 719.32$
$MSE(t_7) = \underbrace{1830.84}_{min}$	$MSE(t_{sg8}) = 1235.21$	$MSE(T_8) = 1383.2$	
$MSE(t_8) = 1841.15$	$MSE(t_{sg9}) = 1219.21$	$MSE(T_9) = 1090.25$	
$MSE(t_9) = 1870.12$	$MSE(t_{sg10}) = 1221.7$	$MSE(T_{10}) = 383.07$	
$\frac{MSE(t_0)}{MSE(t_7)} = 2.38$ $\frac{MSE\left(s^2_{yhr}\right)}{MSE(t_7)} = 0.59$	$\frac{MSE(t_{sg6})}{MSE(t_{sg4})} = 1.04$ $\frac{MSE\left(s^2_{yhr}\right)}{MSE(t_{sg4})} = 0.60$	$\frac{MSE(T_5)}{MSE(T_6)} = 6.57$ $\frac{MSE\left(s^2_{yhr}\right)}{MSE(T_6)} = 3.66$	$\frac{MSE\left(s^2_{yhr}\right)}{MSE(t_g)} = 0.62$

6 Concluding Remarks

In this section, the main results for the proposed approach and other approaches are summarized. Based on Tables (2), (3) and (4), we have the following concluding remarks.

1. For the t family, the MSE of the estimators do not attain $MSE_{min}(t)$ in all examples. Indeed, the estimators in this family are depending on the choice of the parameters of the auxiliary variable as can be seen from Tables (3) and (4).

2. The families are not stable between themselves. Since there are significant difference between the minimum MSE and the maximum MSE of the estimators for each family. This is also clear from the relative efficiency for each family; for example, the relative efficiency for the T family are 19.9144, 17.0945, and 6.57 from Tables (2), (3) and (4), respectively.

3. There is no clear way to depend on a specified family and on a specified estimator from that family. Among the families (t, t_{sg}, and T): In Table (2), the minimum MSE among all other estimators is for the estimator T_2 from the T family. From Table (3), this minimum is for t_{sg9} from the t_{sg} family. Further, T_6 has the minimum MSE as can be seen from Table (4). At the same time, T_6 has the maximum MSE from Table (2).

4. From Table (3), The proposed estimator s_{yhr}^2 has the minimum MSE among all estimators and families and the 3^{rd} best on Table (4). At the same time, the fact that T_6 and T_{10} are not stable estimators can not be ignored since T_6 has the maximum MSE in Table (2) and T_{10} has the maximum MSE in Table (3). Based on this, the emphasis on using s_{yhr}^2 as an estimator for S_y^2 should be raised.

5. The relative efficiency of the s_{yhr}^2 is more stable than the relative efficiency between the families themselves.

6. The estimator s_{yhr}^2 is simple to apply in practical applications and needs only one assumption S_x^2 to be known. However, other estimators and families need more assumptions.

As final conclusions, the proposed estimator s_{yhr}^2 is derived based on the Hartley and Ross (1954) approach. s_{yhr}^2 is exactly unbiased, stable for estimating the finite population variance S_y^2, has minimum MSE among the discussed estimators and families, and behaves as the original estimator proposed by Hartley and Ross (1954).

Acknowledgement

The valuable comments and suggestions from the referees are appreciated and improved this paper.

References

Al-Jararha, J. (2008). *Unbiased Ratio Estimation For Finite Populations.* PhD thesis, Colorado State University, Fort Collins, Co.

Bhushan, S. (2012). Some efficient sampling straregies based on ratio type estimator. *Electronic Journal of Applied Statistical Analysis*, 5(1):74–88.

Cochran, W. G. (1977). *Sampling Techniques.* John Wiley & Sons, New York, 3rd edition.

Gupta, R. (1983). Unbiased estimation of finite population variance using auxiliary information. *Statistics and Probability Letters*, 1(3):121–124.

Hartley, H. O. and Ross, A. (1954). Unbiased ratio estimates. *Nature*, 174:270–271.

Isaki, C. (1983). Variance estimation using auxiliary information. *J Am Stat Assoc*, 78:117–123.

Kadilar, C. and Cingi, H. (2006). Ratio estimators for the population variance insimple and stratified random sampling. *Appl Math Comput*, 35(1):111–115.

Kadilar, C. and Cingi, H. (2007). Improvement in variance estimation in simple random sampling. *Communication of Statistics Theory and Methods*, 36(11):2075–2081.

Prasad, B. and Singh, H. P. (1992). Unbiased estimators of finite population variance using auxiliary information in sample surveys. *Communications in Statistics-Theory and Methods*, 21(5):1367–1376.

Rao, T. J. and Swain, A. K. P. C. (2014). A note on the hartley-ross unbiased ratio estimator. *Communications in Statistics-Theory and Methods*, 43:3162–3169.

Shabbir, J. and Khan, N. (2013). Some modified exponential-ratio type estimators in the presence of non-response under two-phase sampling scheme. *Electronic Journal of Applied Statistical Analysis*, 6(1):1–17.

Singh, H., Upadhyaya, L., and Namjoshi, U. (1988). Estimation of finite population variance. *Current Science*, 57:1331–1334.

Singh, H. P. and Solanki, R. S. (2013). A new procedure for variance estimation in simple random sampling using auxiliary information. *Stat Papers*, 54:479–497.

Upadhyaya, L. and Singh, H. (1996). An estimator of population variance that utilizes the kurtosis of an auxiliary variable in sample surveys. *Vikram Math J*, 19:14–17.