Communications of the Korean Statistical Society
2010, Vol. 17, No. 2, 165–181

Efficient Use of Auxiliary Variables in Estimating Finite Population Variance in Two-Phase Sampling

Housila P. Singh[a], Sarjinder Singh[b], Jong-Min Kim[1,c]

[a]School of Studies in Statistics, Vikram University
[b]Department of Mathematics, Texas A&M University-Kingsville
[c]Statistics Discipline Division of Science and Mathematics, University of Minnesota - Morris

Abstract

This paper presents some chain ratio-type estimators for estimating finite population variance using two auxiliary variables in two phase sampling set up. The expressions for biases and mean squared errors of the suggested classes of estimators are given. Asymptotic optimum estimators(AOE's) in each class are identified with their approximate mean squared error formulae. The theoretical and empirical properties of the suggested classes of estimators are investigated. In the simulation study, we took a real dataset related to pulmonary disease available on the CD with the book by Rosner, (2005).

Keywords: Finite population variance, auxiliary variables, two phase sampling, bias and mean squared error.

1. Introduction

Consider a finite population $U = (U_1, U_2, \dots, U_N)$. Let y and x_1 denote the study variable and auxiliary variable, taking values y_i and x_{1i} respectively on U_i, $i = 1, 2, \dots, N$. Let $(\bar{Y}, \sigma_0^2, C_0)$ and $(\bar{X}, \sigma_1^2, C_1)$ denote the population mean, variance and coefficient of variation of y and x_1 respectively. Das and Tripathi (1978) have considered the problem of estimating the population variance of y using information on auxiliary variable x_1 and suggested six estimators in three different situations, where $\bar{X}_1$ or σ_1^2 or C_1^2 is known and studied their properties. The studies relating to estimation of finite population variance are also made by, among others, Srivastava and Jhajj (1980), Isaki (1983) and Searls and Intarapanich (1990).

Sometimes even if population mean $\bar{X}_1$ of x_1 is not known, the population mean $\bar{X}_2$ of another auxiliary variable x_2 closely related to x_1 but compared to x_1 remotely related to y (*i.e.* $\rho_{01} > \rho_{02}$) is available. Employing two phase sampling procedure and motivated by Chand (1975), several authors including Kiregyera (1980), Mukerjee *et al.* (1987), Srivastava *et al.* (1989), Upadhyaya *et al.* (1990) and Singh *et al.* (1994) have suggested some chain ratio-type estimators for estimating population mean $\bar{Y}$ of y. Suppose a preliminary large sample of size n units is drawn by simple random sampling without replacement(SRSWOR) and the auxiliary variables x_1 and x_2 are measured. In the second phase, a sub sample of size $m(< n)$ units is drawn using SRSWOR and the two variables y and x_1 are observed. Adopting the same procedure as adopted by Chand (1975) and Srivastava (1967), Gupta, Singh, and Mangat (Gupta *et al.*, 1992–1993) suggested the following classes estimators in two different situations and studied their properties up to the first order of approximation. The forms of the estimators proposed by them are given below:

[1] Corresponding author: Associate Professor, Statistics Discipline Division of Science and Mathematics, University of Minnesota - Morris Morris, MN 56267, USA. E-mail: jongmink@morris.umn.edu

(i) When population mean $\bar{X}_2$ of the variable x_2 is known,

$$d_1 = s_0^2\left(\frac{\bar{x}_1}{\bar{x}_1^*}\right)^{g_1}\left(\frac{\bar{x}_2^*}{\bar{X}_2}\right)^{g_2}. \tag{1.1}$$

(ii) When population mean squared $S_2^2 = (N-1)^{-1}\sum_{j=1}^{N}(x_{2j}-\bar{X}_2)^2$ is known,

$$d_2 = s_0^2\left(\frac{s_1^2}{s_1^{*2}}\right)^{I_1}\left(\frac{s_2^{*2}}{S_2^2}\right)^{I_2}, \tag{1.2}$$

where (g_1, g_2) and (I_1, I_2) are constants and

$$\bar{y} = m^{-1}\sum_{j=1}^{m} y_j,\ \ \bar{x}_1 = m^{-1}\sum_{j=1}^{m} x_{1j},\ \ \bar{x}_1^* = n^{-1}\sum_{j=1}^{n} x_{1j},\ \ \bar{x}_2^* = n^{-1}\sum_{j=1}^{n} x_{2j},\ \ s_0^2 = (m-1)^{-1}\sum_{j=1}^{m}(y_j-\bar{y})^2,$$

$$s_1^2 = (m-1)^{-1}\sum_{j=1}^{m}(x_{1j}-\bar{x}_1)^2,\ \ s_1^{*2} = (n-1)^{-1}\sum_{j=1}^{n}(x_{1j}-\bar{x}_1^*)^2,\ \ s_2^{*2} = (n-1)^{-1}\sum_{j=1}^{n}(x_{2j}-\bar{x}_2^*)^2.$$

Assuming that population size N is large enough as compared to sample sizes m and n so that the finite population correction terms can be ignored, to the first degree of approximation, the minimum mean squared errors(MSEs) of d_1 and d_2 are respectively given by

$$\min\ \mathrm{MSE}(d_1) = \sigma_0^4\left[m^{-1}\delta_{400}^* - \lambda\delta_{210}^2 - n^{-1}\delta_{201}^2\right] \tag{1.3}$$

and

$$\min\ \mathrm{MSE}(d_2) = \sigma_0^4\left[m^{-1}\delta_{400}^* - \lambda\left(\frac{\delta_{220}^{*2}}{\delta_{040}^*}\right) - n^{-1}\left(\frac{\delta_{202}^{*2}}{\delta_{004}^*}\right)\right], \tag{1.4}$$

where

$$\delta_{400}^* = (\delta_{400}-1),\quad \delta_{220}^* = (\delta_{220}-1),\quad \delta_{040}^* = (\delta_{040}-1),\quad \delta_{004}^* = (\delta_{004}-1),\quad \delta_{202}^* = (\delta_{202}-1),$$

$$\delta_{pqr} = \frac{\mu_{pqr}}{\mu_{200}^{p/2}\mu_{020}^{q/2}\mu_{002}^{r/2}},\quad \mu_{pqr} = \frac{1}{N}\sum_{j=1}^{N}\left(y_j-\bar{Y}\right)^p\left(x_{1j}-\bar{X}_1\right)^q\left(x_{2j}-\bar{X}_2\right)^r,$$

(p, q, r) being non-negative integers and $\lambda = 1/m - 1/n$.

Recently several researchers have paid attention to two-phase sampling, and a few of them as listed as: Farrell and Singh (2010), Rueda *et al.* (2007) and Singh *et al.* (2006). In particular, the objective of this paper is to propose some improved chain ratio-type estimators for finite population variance σ_0^2 of y supposing that information on another auxiliary variable x_2 is available on all units of the population, for instance, see Mukerjee *et al.* (1987). In Section 2, several special cases of the class of estimators defined in Section 3 are investigated theoretically. The simulation study results are given in Section 4.

2. The Suggested Class of Estimators

In Section 2.1, we consider a situation when the population mean $\bar{X}_2$ of the second auxiliary variable is known where as the population mean $\bar{X}_1$ of the first auxiliary variable remains unknown. In Section 2.2, we consider a situation when the population variance S_2^2 of the second auxiliary variable is known, but the population variance S_1^2 of the first auxiliary variable remains unknown. In Section 2.3, we consider a situation when the coefficient of variation C_2 of x_2 may be known.

2.1. When population mean $\bar{X}_2$ is known

Utilising prior knowledge on population mean $\bar{X}_2$ of $\bar{x}_2$, we define a chain ratio-type estimator for σ_0^2 as

$$\Sigma_1 = s_0^2\left(\frac{\bar{x}_1}{\bar{x}_1^*}\right)^{p_1}\left(\frac{\bar{x}_2^*}{\bar{X}_2}\right)^{p_2}\left(\frac{\bar{x}_2}{\bar{X}_2}\right)^{p_3}, \tag{2.1}$$

where $\bar{x}_2 = m^{-1}\sum_{j=1}^{m} x_{2j}$ and $p_i\,(i = 1,2,3)$ are suitably chosen constants. For a choice of constants as: $(p_1 = g_1, p_2 = g_2, p_3 = 0)$, Σ_1 reduces to the estimator d_1.

To the first degree of approximation, the bias and MSE of Σ_1 are, respectively, given by

$$B(\Sigma_1) = \sigma_0^2\left[\lambda p_1 C_1\left(\frac{p_1-1}{2}C_1 + \rho_{12}C_2p_3 + \delta_{210}\right) + \left(\frac{p_2C_2}{n}\right)\left(\frac{p_2-1}{n}C_1 + p_3C_2 + \delta_{201}\right) + \left(\frac{p_3C_2}{m}\right)\left(\frac{p_3-1}{2}C_2 + \delta_{201}\right)\right] \tag{2.2}$$

and

$$\mathrm{MSE}(\Sigma_1) = \sigma_0^4\left[\lambda p_1 C_1\left(p_1C_1 + 2p_3\rho_{12}C_2 + 2\delta_{210}\right) + \frac{1}{m}\left\{\delta_{400}^* + p_3C_2\left(p_3C_2 + 2\delta_{201}\right)\right\} \frac{1}{n}p_2C_2\left(p_2C_2 + 2p_3C_2 + 2\delta_{201}\right)\right], \tag{2.3}$$

where ρ_{12} is the correlation coefficient between variables x_1 and x_2 and $C_i\,(i = 0,1,2)$ is the coefficient of variation of y, x_1 and x_2.

The $\mathrm{MSE}(\Sigma_1)$ at (2.3) is minimized for:

$$p_1 = \frac{\rho_{12}\delta_{201} - \delta_{210}}{C_1\left(1-\rho_{12}^2\right)} \tag{2.4}$$

$$p_2 = \frac{\rho_{12}(\rho_{12}\delta_{201} - \delta_{210})}{C_2\left(1-\rho_{12}^2\right)} \tag{2.5}$$

and

$$p_3 = \frac{(\rho_{12}\delta_{210} - \delta_{201})}{C_2\left(1-\rho_{12}^2\right)}. \tag{2.6}$$

On substituting (2.4)–(2.6) in (2.3, the minimum MSE of Σ_1 is, therefore, given by

$$\min \mathrm{MSE}(\Sigma_1) = \frac{\sigma_0^4}{m}\delta_{400}^*\left[1 - \frac{n-m}{n}\gamma_{0.12}^2 - \frac{m}{n}\rho^2\right], \tag{2.7}$$

where

$$\gamma_{0.12}^2 = \frac{\delta_{201}^2 + \delta_{210}^2 - 2\rho_{12}\delta_{210}\delta_{201}}{\delta_{400}^*\left(1-\rho_{12}^2\right)}$$

is the squared of the coefficient of multiple correlation of $(y - Y)^2$ with $(x_1 - \bar{X}_1)$ and $(x_2 - \bar{X}_2)$, and $\rho^2 = \delta_{201}^2/\delta_{400}^*$ is the squared of the correlation coefficient between $(y - Y)^2$ and $(x_2 - \bar{X}_2)$. It can be further easily be seen that the minimum MSE of the difference type estimator:

$$\Sigma_{d1} = s_y^2 + p_1(\bar{x}_1^* - \bar{x}_1) + p_2\left(\bar{X}_2 - \bar{x}_2^*\right) + p_3\left(\bar{X}_2 - \bar{x}_2\right) \tag{2.8}$$

is the same as that of Σ_1 given in (2.7).

From (1.3) and (2.7), we have

$$\min \text{MSE}(d_1) - \min \text{MSE}(\Sigma_1 \text{ or } \Sigma_{d1}) = \lambda\sigma_0^4 \frac{(\delta_{201} - \rho_{12}\delta_{210})^2}{1 - \rho_{12}^2} > 0 \tag{2.9}$$

unless $\delta_{201} = \rho_{12}\delta_{210}$.

It follows that the proposed estimator Σ_1 (or Σ_{d1}) is more efficient than d_1.

2.2. When population mean squared S_2^2 is known

We consider the estimator for σ_y^2 as

$$\sum{}_2 = s_0^2\left(\frac{s_1^2}{s_1^{*2}}\right)^{q_1}\left(\frac{s_2^{*2}}{S_2^2}\right)^{q_2}\left(\frac{s_2^2}{S_2^2}\right)^{q_3}, \tag{2.10}$$

where $q_i\,(i = 1, 2, 3)$ are constants and $s_2^2 = (m-1)^{-1}\sum_{j=1}^{m}(x_{2j} - \bar{x}_2)^2$. For $q_1 = h_1$, $q_2 = h_2$ and $q_3 = 0$, Σ_2 reduces to d_2.

To the first degree of approximation, the bias and MSE of Σ_2 are respectively given by

$$B(\Sigma_2) = \sigma_0^2\left[\lambda q_1\left\{\left(\frac{q_1 - 1}{2}\right)\delta_{040}^* + q_3\delta_{022}^* + \delta_{220}^*\right\} + \frac{q_2}{m}\left\{\left(\frac{q_2 - 1}{2}\right)\delta_{004}^* + \delta_{202}^*\right\}\right.$$
$$\left. + \frac{q_2}{n}\left\{\left(\frac{q_2 - 1}{2}\right)\delta_{004}^* + q_3\delta_{004}^* + \delta_{202}^*\right\}\right] \tag{2.11}$$

and

$$\text{MSE}(\Sigma_2) = \sigma_0^4\left[\lambda q_1\left(q_1\delta_{040}^* + 2q_3\delta_{022}^* + 2\delta_{220}^*\right) + \frac{1}{m}\left\{\delta_{400}^* + q_3\left(q_3\delta_{004}^* + 2\delta_{202}^*\right)\right\}\right.$$
$$\left. + \frac{1}{n}q_2\left(q_2\delta_{004}^* + 2q_3\delta_{004}^* + 2\delta_{202}^*\right)\right]. \tag{2.12}$$

The MSE(Σ_2) is minimized for

$$q_1 = \frac{\delta_{022}^*\delta_{202}^* - \delta_{004}^*\delta_{220}^*}{\delta_{040}^*\delta_{004}^* - \delta_{022}^{*2}} \tag{2.13}$$

$$q_2 = \left(\frac{\delta_{022}^*}{\delta_{004}^*}\right)\frac{\delta_{202}^*\delta_{022}^* - \delta_{220}^*\delta_{004}^*}{\delta_{040}^*\delta_{004}^* - \delta_{022}^{*2}} \tag{2.14}$$

and

$$q_3 = \frac{\delta_{022}^*\delta_{220}^* - \delta_{202}^*\delta_{040}^*}{\delta_{040}^*\delta_{004}^* - \delta_{022}^{*2}}. \tag{2.15}$$

Thus the minimum MSE of Σ_2 is given by

$$\min \text{MSE}(\Sigma_2) = \frac{\sigma_0^4}{m}\delta^*_{400}\left[1 - \left(\frac{n-m}{n}\right)\gamma^{*2}_{0.12} - \frac{m}{n}\rho^{*2}\right], \tag{2.16}$$

where

$$\gamma^{*2}_{0.12} = \frac{\delta^*_{004}\delta^{*2}_{220} - 2\delta^*_{220}\delta^*_{022}\delta^*_{202} + \delta^*_{040}\delta^{*2}_{202}}{\delta^*_{400}\left(\delta^*_{040}\delta^*_{004} - \delta^{*2}_{022}\right)}$$

is the squared of the coefficient of multiple correlation of $(y - \bar{Y})^2$ with $(x_1 - \bar{X}_1)^2$ and $(x_2 - \bar{X}_2)^2$ and $\rho^{*2} = \delta^{*2}_{202}/(\delta^*_{400}\delta^*_{004})$ is the squared of the correlation coefficient between $(y - \bar{Y})^2$ and $(x_2 - \bar{X}_2)^2$.

It is to be noted that the difference-type estimator:

$$\Sigma_{d2} = s_0^2 + q_1\left(s_1^{*2} - s_1^2\right) + q_2\left(S_2^2 - s_2^{*2}\right) + q_3\left(S_2^2 - s_2^2\right) \tag{2.17}$$

attains the same minimum MSE as that of Σ_2 given in (2.16).

From (1.4) and(2.16), we have

$$\min \text{MSE}(d_2) - \min \text{MSE}(\Sigma_2 \text{ or } \Sigma_{d2}) = \lambda\sigma_0^4\frac{\left(\delta^*_{004}\delta^*_{220} - \delta^*_{202}\delta^*_{022}\right)}{\delta^*_{004}\left(\delta^*_{040}\delta^*_{004} - \delta^{*2}_{022}\right)} > 0 \tag{2.18}$$

provided $\delta^*_{004}\delta^*_{220} \neq \delta^*_{202}\delta^*_{022}$, whence it follows that the proposed estimator Σ_2 (or Σ_{d2}) is better than the estimator d_2.

2.3. When population coefficient of variation C_2 is known

In many situations of practical importance information regarding population mean $\bar{X}_2$ or mean squared S_2^2 of the auxiliary variable x_2 may not be available, but the coefficient of variation C_2 of x_2 may be available as it is a very stable quantity, for instance, see Searls (1964), Murthy (1967, pp. 96–99), Gleser and Healy (1976) and Lee (1981). The survey statistician may utilize this information in obtaining estimators for σ_0^2, better than the usual sample mean squared s_0^2. Thus using the knowledge of C_2, we define the following class of estimators for σ_0^2 as

$$\Sigma_3 = s_0^2\left(\frac{\hat{C}_1^2}{\hat{C}_1^{*2}}\right)^{k_1}\left(\frac{\hat{C}_2^{*2}}{C_2^2}\right)^{k_2}\left(\frac{\hat{C}_2^2}{C_2^2}\right)^{k_3}, \tag{2.19}$$

where k_i $(i = 1, 2, 3)$, are constants, $\hat{C}_1^2 = s_1^2/\bar{x}_1^2$, $\hat{C}_2^2 = s_2^2/\bar{x}_2^2$, $\hat{C}_1^{*2} = s_1^{*2}/\bar{x}_1^{*2}$ and $\hat{C}_2^{*2} = s_2^{*2}/\bar{x}_2^{*2}$.

To the first degree of approximation, it can easily be seen that

$$\begin{aligned} B(\Sigma_3) = \sigma_0^2\Bigg[&\lambda\left\{h(y, x_1) + \left(\frac{k_1 - 1}{2}\right)h(x_1) + k_3 h(x_1, x_2)\right\}k_1 \\ &+ \frac{k_2}{n}\left\{h(y, x_2) + k_3 h(x_2) + \left(\frac{k_2 - 1}{2}\right)h(x_2)\right\} + \frac{k_3}{m}\left\{\left(\frac{k_3 - 1}{2}\right)h(x_2) + h(y, x_2)\right\}\Bigg] \end{aligned} \tag{2.20}$$

and

$$\begin{aligned} \text{MSE}(\Sigma_3) = \sigma_0^4\Bigg[&\lambda k_1\left\{k_1 h(x_1) + 2k_3 h(x_1, x_2) + 2g(y, x_1)\right\} + \frac{1}{m}\left\{\delta^*_{400} + k_3\left(k_3 h(x_2) + 2g(y, x_2)\right)\right\} \\ &+ \frac{k_2}{n}\left\{(k_2 + 2k_3)h(x_2) + 2g(y, x_2)\right\}\Bigg], \end{aligned} \tag{2.21}$$

where

$$h(x_1) = \left(4C_1^2 - 4\delta_{030}C_1 + \delta_{040}^*\right), \qquad h(x_2) = \left(4C_2^2 - 4\delta_{003}C_2 + \delta_{004}^*\right),$$
$$h(y,x_1) = \left[3C_1^2 - 2C_1(\delta_{030} + \delta_{210}) + \delta_{220}^*\right], \qquad h(y,x_2) = \left[3C_2^2 - 3C_2(\delta_{003} + \delta_{201}) + \delta_{202}^*\right],$$
$$h(x_1,x_2) = \left[4\rho_{12}C_1C_2 - 2\delta_{012}C_1 - 2\delta_{021}C - 2 + \delta_{220}^*\right],$$
$$g(y,x_1) = \left(\delta_{220}^* - 2\delta_{210}C_1\right), \qquad g(y,x_2) = \left(\delta_{202}^* - 2\delta_{201}C_2\right).$$

The MSE(Σ_3) is minimized for

$$k_1 = \frac{h(x_1,x_2)g(y,x_2) - h(x_2)g(y,x_1)}{h(x_1)h(x_2) - \{h(x_1,x_2)\}^2}, \tag{2.22}$$

$$k_2 = \frac{h(x_1,x_2)}{h(x_2)}\left[\frac{h(x_1,x_2)g(y,x_2) - h(x_2)g(y,x_1)}{h(x_1)h(x_2) - \{h(x_1,x_2)\}^2}\right], \tag{2.23}$$

$$k_3 = \frac{h(x_1,x_2)g(y,x_1) - h(x_1)g(y,x_2)}{h(x_1)h(x_2) - \{h(x_1,x_2)\}^2}. \tag{2.24}$$

On substituting (2.4)–(2.6) in (2.3, the minimum MSE of Σ_3 is, therefore, given by

$$\min \text{MSE}(\Sigma_3) = \frac{\sigma_0^4}{m}\delta_{400}^*\left[1 - \left(\frac{n-m}{n}\right)\gamma_{0.12}^{**2} - \frac{m}{n}\rho^{**2}\right], \tag{2.25}$$

where

$$\gamma_{0.12}^{**2} = \frac{h(x_1)g^2(y,x_2) - 2h(x_1,x_2)g(y,x_1)g(y,x_2) + h(x_2)g^2(y,x_1)}{\delta_{400}^*\{h(x_1)h(x_2) - h^2(x_1,x_2)\}} \tag{2.26}$$

and

$$\rho^{**2} = \frac{g^2(y,x_2)}{h(x_2)}\delta_{400}^*.$$

2.4. Results in trivariate normal population

Let (y, x_1, x_2) have a trivariate normal distribution with mean $(\bar{Y}, \bar{X}_1, \bar{X}_2)$ and covariance matrix Σ in which the variances are denoted by σ_0^2, σ_1^2 and σ_2^2 and correlation coefficients by ρ_{01}, ρ_{02} and ρ_{12}. Also we have $\delta_{004}^* = 2$, $\delta_{220}^* = 2\rho_{01}^2$, $\delta_{202}^* = 2\rho_{02}^2$, $\delta_{022}^* = 2\rho_{12}^2$ and $\delta_{012} = \delta_{201} = \delta_{210} = 0$ Thus the expressions in (1.3), (1.4), (2.3), (2.12), (2.16) and (2.25), respectively, reduces to

$$\min \text{MSE}(d_1) = \frac{2\sigma_0^4}{m} = V\left(s_0^2\right) \tag{2.27}$$

$$\min \text{MSE}(d_2) = \frac{2\sigma_0^4}{m}\left[1 - \frac{n-m}{n}\rho_{01}^4 - \frac{m}{n}\rho_{02}^4\right] \tag{2.28}$$

$$\min \text{MSE}(\Sigma_1) = V\left(s_0^2\right)\frac{2\sigma_0^4}{m} \tag{2.29}$$

$$\min \text{MSE}(\Sigma_2) = \frac{2\sigma_0^4}{m}\left[1 - \left(\frac{n-m}{n}\right)\left(\frac{\rho_{01}^4 - 2\rho_{01}^2\rho_{02}^2\rho_{12}^2 + \rho_{02}^4}{1-\rho_{12}^4}\right) - \frac{m}{n}\rho_{02}^4\right] \tag{2.30}$$

$$\min \text{MSE}(\Sigma_3) = \left[1 - \frac{m}{n}\frac{\rho_{02}^4}{1+2C_2^2} - \frac{\left(1+2C_1^2\right)\rho_{02}^4 - 2\rho_{12}(\rho_{12} + 2C_1C_2)\rho_{01}^2\rho_{02}^2 + \left(1+2C_2^2\right)\rho_{01}^4}{\left(1+2C_1^2\right)\left(1+2C_2^2\right) - \rho_{12}^2(\rho_{12} + 2C_1C_2)^2}\right]. \tag{2.31}$$

From the above expressions, we make the following observations:

I. It is clear from (2.27) and (2.29) that in trivariate normal populations, the use of auxiliary information in the form of sample means $(\bar{x}_1, \bar{x}_1^*, \bar{x}_2, \bar{x}_2^*)$ and known population mean does not contribute towards the reduction in variance of usual estimator s_0^2. In practice, therefore, one should not pick up the estimators from the classes d_1 and Σ_1. The estimators s_0^2, d_1 and Σ_1 are equally efficient. This result is same as reported by Singh *et al.* (1999) while constructing calibrated estimators of variance in survey sampling.

II. Expression (2.30) clearly demonstrates that there is considerable reduction in the variance of the usual estimator s_0^2 by using Σ_2. Thus merely the additional knowledge of the coefficients of correlation ρ_{ij} $(i \neq j = 0, 1, 2)$ would enable us to use the suggested class of estimators Σ_2. The proposed estimator Σ_2 is more efficient than s_0^2 and d_2 in case of trivariate normal population too.

III. The reduction in variance of s_0^2 is seen by using the estimator Σ_3 which requires the additional knowledge on well known parameters C_1, C_2 and ρ_{ij} $(i \neq j = 0, 1, 2)$ of the auxiliary characters x_1 and x_2.

Thus it is worth noting that one should pick up estimators Σ_2 and Σ_3 in case of trivariate normal population.

3. A General Class of Estimators

When both population mean $\bar{X}_2$ and population mean squared S_2^2 of auxiliary variate x_2 are known, we first suggest a class of estimators for estimating the population variance σ_0^2 as:

$$d_3 = s_0^2\left(\frac{\bar{x}_1}{\bar{x}_1^*}\right)^{\alpha}\left(\frac{s_1^2}{s_1^{*2}}\right)^{\beta}\left(\frac{\bar{x}_2^*}{\bar{X}_2}\right)^{\tau}\left(\frac{s_2^{*2}}{S_2^2}\right)^{\phi}, \tag{3.1}$$

which may be further generalized as:

$$\Sigma_4 = s_0^2\left(\frac{\bar{x}_1}{\bar{x}_1^*}\right)^{\alpha}\left(\frac{s_1^2}{s_1^{*2}}\right)^{\beta}\left(\frac{\bar{x}_2^*}{\bar{X}_2}\right)^{\tau}\left(\frac{s_2^{*2}}{S_2^2}\right)^{\phi}\left(\frac{\bar{x}_2}{\bar{X}_2}\right)^{\gamma}\left(\frac{s_2^2}{S_2^2}\right)^{\psi} \tag{3.2}$$

$(\alpha, \beta, \tau, \phi, \gamma, \psi)$ are suitably chosen constants. For $\gamma = \psi = 0$ in (3.2), Σ_4 reduces to d_3.

Keeping the form of d_3 one can define a class of estimators for σ_0^2 a

$$\Sigma_h = h\left(s_0^2, \bar{x}_1, \bar{x}_1^*, s_1^2, s_1^{*2}, \bar{x}_2^*, s_2^{*2}\right), \tag{3.3}$$

where $h(\bullet)$ is a function of $(s_0^2, \underset{\sim}{u})$ with $\underset{\sim}{u} = (\bar{x}_1, \bar{x}_1^*, s_1^2, s_1^{*2}, \bar{x}_2^*, s_2^{*2})$ such that

(i) $(s_0^2, \underset{\sim}{u})$ assume the value in a closed convex subspace, Q, of the six dimensional space containing the point

$$\left(\sigma_0^2, \bar{X}_1, \bar{X}_1, S_1^2, S_1^2, \bar{X}_2, S_2^2\right) = \left(\sigma_0^2, \bar{X}_1, S_1^2, \bar{X}_2, S_2^2\right) = \left(\sigma_0^2, \underset{\sim}{P}\right).$$

(ii) The function $h(s_0^2, \underset{\sim}{u})$ is continuous and bounded in Q.

(iii) $h(\sigma_0^2, \underset{\sim}{P}) = 1$, where $h(\sigma_0^2, \underset{\sim}{P})$ denotes the first order partial derivative with respect to

$$s_0^2 \text{ at } \left(s_0^2, \underset{\sim}{u}\right) = \left(\sigma_0^2, \underset{\sim}{P}\right).$$

(iv) The first and second order partial derivatives of $h(s_0^2, \underset{\sim}{u})$ exist and are continuous and bounded in Q.

It is to noted that the estimator Σ_h includes the estimators d_1 and d_2, but it fails to include the estimators Σ_i, $i = 1, 2, 3, 4$. We, therefore, define a more general class of estimators of σ_0^2 as

$$\Sigma_g = g\left(s_0^2, \underset{\sim}{v}\right), \tag{3.4}$$

where $\underset{\sim}{v} = (\bar{x}_1, \bar{x}_1^*, s_1^2, s_1^{*2}, \bar{x}_2^*, s_2^{*2}, \bar{x}_2, s_2^2)$ and $g(s_0^2, \underset{\sim}{v})$ is function of $(s_0^2, \underset{\sim}{v})$ such that $g(\sigma_0^2, \underset{\sim}{P}) = \sigma_0^2$ satisfying the following conditions:

(i) Whatever be the sample chosen, let $(s_0^2, \underset{\sim}{v})$ assume values in a closed convex sub-space, S, of the eight dimensional real space containing the point $(\sigma_0^2, \underset{\sim}{P})$.

(ii) The function $g(s_0^2, \underset{\sim}{v})$ is continuous and bounded in S.

(iii) The first and second order partial derivatives of $g(s_0^2, \underset{\sim}{v})$ exists and are also continuous in S.

(iv) $g(\sigma_0^2, \underset{\sim}{P}) = 1$, where $(\sigma_0^2, \underset{\sim}{P})$ denote, the first order partial derivative with respect to s_0^2 at $(s_0^2, \underset{\sim}{v}) = (\sigma_0^2, \underset{\sim}{P})$. It may be noted here that for constructing Σ_h in (4.3) and Σ_g in (4.4) we adopt the approach by Srivastava (1980).

First, we shall derive the expressions for bias and MSE of Σ_h and then of Σ_g. Expanding $h(s_0^2, \underset{\sim}{u})$ about the point $(\sigma_0^2, \underset{\sim}{P})$ in a second order Taylor's series and taking expectations it is found that the bias of Σ_h is of order n^{-1}. Denoting the first order partial derivatives of $h(s_0^2, \underset{\sim}{u})$ with respect to $(s_0^2, \underset{\sim}{u})$ at the point $(\sigma_0^2, \underset{\sim}{P})$ by $(h_0, h_1, h_2, h_3, h_4, h_5, h_6)$ respectively and noting that $h_1 = -h_2$, $h_3 = -h_4$, we see that the MSE of Σ_h to the first degree of approximation is given by

$$\text{MSE}(\Sigma_h) = \left[\frac{1}{m}\mu_{400}^* + (h^*)^T \underset{\sim}{\Sigma} \left(\underset{\sim}{h}^*\right) + 2\, \underset{\sim}{b}^T \underset{\sim}{h}^*\right] \tag{3.5}$$

which is minimized for

$$\underset{\sim}{h}^* = -\underset{\sim}{\Sigma}^{-1} \underset{\sim}{b}, \tag{3.6}$$

where

$$\underset{\sim}{\Sigma}^{-1} = \begin{bmatrix} \lambda\mu_{020}, & \lambda\mu_{030}, & 0, & 0 \\ \lambda\mu_{030}, & \lambda\mu_{040}^*, & 0, & 0 \\ 0, & 0, & 1/n\mu_{002}, & 1/n\mu_{003} \\ 0, & 0, & 1/n\mu_{003}, & 1/n\mu_{004}^* \end{bmatrix}, \quad \left(\underset{\sim}{h}^*\right)^T = (h_1, h_3, h_5, h_6),$$

$$\underset{\sim}{b}^T = \left(\lambda\mu_{210}, \lambda\mu_{220}^*, \frac{1}{n}\mu_{201}, \frac{1}{n}\mu_{202}^*\right), \quad \mu_{220}^* = \mu_{220} - \mu_{200}\mu_{020}, \quad \mu_{202}^* = \mu_{202} - \mu_{200}\mu_{002},$$

$$\mu_{040}^* = \mu_{040} - \mu_{020}^2, \quad \mu_{004}^* = \mu_{004} - \mu_{002}^2, \quad \mu_{400}^* = \mu_{400} - \mu_{200}^2$$

and $\underset{\sim}{\Sigma}$ is assumed to be positive definite.

Substituting for h^* from (3.6) in (3.5), the minimum value of MSE of Σ_h is given by

$$\min \text{MSE}(\Sigma_h) = \frac{1}{m}\mu^*_{400} - \underset{\sim}{b}^T \underset{\sim}{\Sigma}^{-1} \underset{\sim}{b}. \tag{3.7}$$

Since $\underset{\sim}{\Sigma}$ is positive definite, the term $\underset{\sim}{b}^T \underset{\sim}{\Sigma}^{-1} \underset{\sim}{b}$ is non-negative. To obtain the bias and MSE of the general class Σ_g in (3.4), we denote the first order partial derivatives of $g(s_0^2, \underset{\sim}{v})$ with respect to $(s_0^2, \underset{\sim}{v})$ at the point $(\sigma_0^2, \underset{\sim}{P})$ by $(g_0, g_1, g_2, g_3, g_4, g_5, g_6, g_7, g_8)$, $(\underset{\sim}{g}^*)^T = (g_1, g_3, g_5, g_6, g_7, g_8)$, $\underset{\sim}{b}^{*T} = (\lambda\mu_{210}, \lambda\mu^*_{220}, 1/n\mu_{201}, 1/n\mu^*_{202}, 1/m\mu_{201}, 1/m\mu^*_{202})$

and $\underset{\sim}{\Sigma}^* = \begin{bmatrix} \lambda\mu_{020}, & \lambda\mu_{030}, & 0, & 0, & \lambda\mu_{011}, & \lambda\mu_{012} \\ \lambda\mu_{030}, & \lambda\mu^*_{040}, & 0, & 0, & \lambda\mu_{021}, & \lambda\mu^*_{022} \\ 0, & 0, & \mu_{002}/n, & \mu_{003}/n, & \mu_{002}/n, & \mu^*_{004}/n \\ 0, & 0, & \mu_{003}/n, & \mu^*_{004}/n, & \mu_{003}/n, & \mu^*_{004}/n \\ \lambda\mu_{011}, & \lambda\mu_{021}, & \mu_{002}/n, & \mu_{003}/n, & \mu_{002}/m, & \mu_{003}/m \\ \lambda\mu_{012}, & \lambda\mu^*_{022}, & \mu_{003}/n, & \mu^*_{004}/n, & \mu_{003}/m, & \mu^*_{004}/m \end{bmatrix}$ assumed to be positive definite.

Expanding $g(s_0^2, \underset{\sim}{v})$ about the point $(\sigma_0^2, \underset{\sim}{P})$ in a second order Taylor's series and taking expectations , it is found that the bias of Σ_g is of order n^{-1}.

Noting that $g_1 = -g_2$, $g_3 = -g_4$, $g_0(\sigma_0^2, \underset{\sim}{P}) = 1$, we find the MSE of Σ_g to the first degree of approximation, as

$$\text{MSE}(\Sigma_g) = \frac{\mu^*_{400}}{m} + \left(\underset{\sim}{g}^*\right)^T \underset{\sim}{\Sigma}^* \left(\underset{\sim}{g}^*\right) + 2\,\underset{\sim}{b}^{*T} \underset{\sim}{g}^* \tag{3.8}$$

which is minimized for

$$\underset{\sim}{g}^* = -\underset{\sim}{\Sigma}^{*-1} \underset{\sim}{b}^*. \tag{3.9}$$

Thus the minimum MSE of Σ_g is given by

$$\min \text{MSE}(\Sigma_g) = \frac{\mu^*_{400}}{m} - \underset{\sim}{b}^{*T} \underset{\sim}{\Sigma}^{*-1} \underset{\sim}{b}^*. \tag{3.10}$$

As Σ^* is positive definite, the term $\underset{\sim}{b}^{*T} \underset{\sim}{\Sigma}^{*-1} \underset{\sim}{b}^*$ is non-negative.

We have from (3.7) and (3.10) that

$$\min \text{MSE}(\Sigma_h) - \min \text{MSE}(\Sigma_g) = \left(\underset{\sim}{b}^{*T} \underset{\sim}{\Sigma}^{*-1} \underset{\sim}{b}^* - \underset{\sim}{b}^T \underset{\sim}{\Sigma}^{-1} \underset{\sim}{b}\right) > 0. \tag{3.11}$$

It follows that Σ_g is more efficient than Σ_h. Thus the estimators Σ_g and Σ_4 are better than Σ_3.

In case of trivariate normal populations the minimu MSE's of Σ_h and Σ_g are same as that of d_2 and Σ_2 respectively given in (2.22) and (2.24).

Remark 1. If $n = N$ and x_2 is considered to be a non zero constant, the estimator Σ_h and Σ_g reduce to the class of estimators:

$$\Sigma_s = t\left(s_0^2, \bar{x}_1, \bar{X}_1, s_1^2, S_1^2\right)$$

reported by Srivastava and Jhajj (1980).

Table 1: Format for FEV.DAT

Column	Variable	Format or code
1–5	ID Number	
6–8	Age (years)	
10–15	FEV (liters)	X.XXX
17–20	Height (inches)	XX.X
22	Gender	0 = female/ 1 = male
24	Smoking Status	0 = noncurrent smoker/1 = current smoker

Table 2: Descriptive parameters of the three variables

Variable	N	Mean	SE Mean	Min	Q1	Q2	Q3	Max
Age	654	9.931	0.116	3.000	8.000	10.000	12.000	19.000
FEV	654	2.637	0.034	0.791	1.977	2.547	3.121	5.793
Height	654	61.144	0.223	46.000	57.000	61.000	65.000	74.000

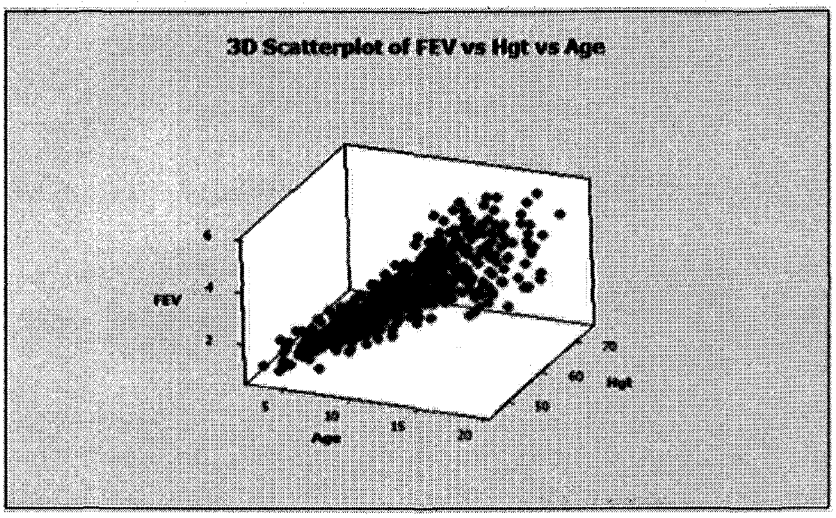

Figure 1: *3D scatter plot of the real population used in the study*

Remark 2. If x_2 is to be considered to be non-zero constant, the estimator Σ_h and Σ_g reduce to the estimator

$$\Sigma_s^* = t^*\left(s_0^2, \bar{x}_1, \bar{x}_1^*, s_1^2, s_1^{*2}\right)$$

of which the estimator $\Sigma_s^{**} = s_0^2 h^*(\bar{x}_1/\bar{x}_1^*,\ s_1^2/s_1^{*2})$ suggested by Singh (1990) is a particular case.

Remark 3. As it is assumed that x_2-values are known for all the units in the population, the estimators $d_1, d_2, \Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4$ and the classes of estimators Σ_h and Σ_g require the same data, namely, the y-values over a sample, say s_m of size m and the x_1-values over a sample, say s_n of size n, therefore all these estimators are equally costly. Hence variance/MSE will be the only criterion for a meaningful comparison so that Σ_g will be best of these estimators.

4. Empirical Example

In the simulation study, we took a real dataset related to pulmonary disease available on the CD with the book by Rosner (2005). In this dataset, FEV(forced expiratory volume) is an index of pul-

Table 3: RE(%) comparison of various estimators for different values on n and m

Ratio $(m/n) \approx$		30%	40%	50%	60%	70%	80%	90%
n	m	3	5	6	7	9	10	11
13	s_0^2	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	d_1	103.52	103.54	103.55	103.56	103.58	103.59	103.60
	d_2	127.26	132.44	135.19	138.06	144.18	147.45	150.86
	Σ_1	103.90	103.84	103.82	103.79	103.73	103.71	103.68
	Σ_2	152.33	153.46	154.04	154.63	155.79	156.39	156.99
	Σ_3	136.33	135.82	135.56	135.30	134.79	134.54	134.28
n	m	7	10	13	15	18	20	23
26	s_0^2	100.00	100.00	100.00	100.00	100.00	100.00	100.0
	d_1	103.52	103.54	103.55	103.57	103.58	103.59	103.61
	d_2	128.52	132.44	136.61	139.54	144.18	147.45	152.63
	Σ_1	103.89	103.85	103.80	103.78	103.74	103.71	103.67
	Σ_2	152.62	153.46	154.34	154.92	155.79	156.39	157.29
	Σ_3	136.20	135.82	135.43	135.17	134.79	134.54	134.16
n	m	11	15	19	23	27	31	35
39	s_0^2	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	d_1	103.53	103.54	103.55	103.57	103.58	103.59	103.61
	d_2	128.94	132.44	136.14	140.04	144.18	148.57	153.23
	Σ_1	103.88	103.85	103.81	103.77	103.73	103.69	103.66
	Σ_2	152.71	153.47	154.24	155.02	155.79	156.59	157.39
	Σ_3	136.16	135.82	135.47	135.13	134.79	134.45	134.12
n	m	15	20	26	31	36	41	46
52	s_0^2	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	d_1	103.53	103.54	103.56	103.57	103.58	103.59	103.61
	d_2	129.16	132.44	136.61	140.29	144.18	148.28	152.63
	Σ_1	103.88	103.85	103.81	103.77	103.74	103.70	103.67
	Σ_2	152.75	153.47	154.33	155.06	155.79	156.54	157.29
	Σ_3	136.14	135.82	135.43	135.11	134.79	134.47	134.16
n	m	19	26	32	39	45	52	58
65	s_0^2	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	d_1	103.53	103.54	103.56	103.57	103.58	103.59	103.61
	d_2	129.28	132.98	136.33	140.44	144.18	148.79	152.99
	Σ_1	103.88	103.84	103.81	103.77	103.74	103.69	103.66
	Σ_2	152.78	153.58	154.28	155.09	155.79	156.63	157.35
	Σ_3	136.12	135.76	135.46	135.09	134.79	134.44	134.13

monary function that measures the volume of air expelled after 1 second of constant effort. Data Set FEV.DAT(on the CD-ROM given at the back of the book) contains determinations of FEV in 1980 on 654 children ages 3–19 who were in the Childhood Respiratory Disease Study(CRD Study) in East Boston, Massachusetts. These data are part of a longitudinal study to follow the change in pulmonary function over time in children. The format and variables in the file FEV.DAT are listed in Table 1.

In this population there are only three quantitative variables: FEV, Age and Height. The descriptive parameters of these three variables in the population are given in Table 2.

We consider the problem of estimation of variance of the study variable y = FEV and two auxiliary variables x_1 = Age and x_2 = Height. The population correlation coefficient between the three variables are noted as: $\rho_{yx_1} = \rho_{01} = 0.756$, $\rho_{yx_2} = \rho_{02} = 0.868$ and $\rho_{x_1x_2} = \rho_{12} = 0.792$. A three dimensional pictorial representation of the population is shown in Figure 1.

The FORTRAN code used in evaluating the proposed new methodology are given in the Appendix which in fact are used to study the relative efficiency of the five estimators d_1, d_2, Σ_1, Σ_2 and Σ_3 over the usual estimator s_0^2. Let $e_0 = s_0^2$, $e_1 = d_1$, $e_2 = d_2$, $e_3 = \Sigma_1$, $e_4 = \Sigma_2$ and $e_5 = \Sigma_3$ The percent relative efficiency(RE) of the j^{th} estimator e_j with respect the estimator e_0 is given by the formula:

$$\mathrm{RE} = \frac{\mathrm{MSE}(e_0)}{\mathrm{MSE}(e_j)} \times 100\%. \tag{4.1}$$

We consider the values of the sampling fraction $f = n/N$ in the range 0.02 to 0.1 with a step of 0.02, which means we select first phase samples of sizes 2% , 4%, 6%, 8% and 10% of the population size. From the given fist-phase sample of n units, we select second phase sample 30% to 90% with a step of 10%. The relative efficiency values obtained from such a simulation study are presented in Table 3.

Discussion of results: It is interesting to note that if the percentage $(m/n) * 100\%$ remains constant irrespective of the increase in sizes of the first phase and the second phase samples, the relative efficiency of the proposed estimators with respect to the usual estimator remains almost same. These findings are similar to those as reported in Singh *et al.* (2009) in case of non-response while estimating the ratio of two population means. The relative efficiency(RE) of the estimators d_1 and Σ_1 remains approximately 103% over the estimator s_0^2, and in this situation both the estimators d_1 and Σ_1 are found to be almost equally efficient. The relative efficiency of the estimator d_2 increases from approximately 126% to 152% as the size of the second phase sample increases from 30% to 90% of the size of first phase sample. In the same way the relative efficiency of the estimator Σ_2 increases from 152% to 157% and in contrast that of the estimator Σ_3 decreases increase from 136% to 134% in all cases as the second phase sample size increases from 30% to 90% of the size of the first phase sample. Thus, we conclude that in case of the populations similar to the considered in the present investigation, the use of the estimator Σ_2 is expected to perform better than the rest of the other estimators.

Acknowledgement

The authors are thankful to the Editor and two anonymous referees for their valuable comments on the original version of the manuscript.

Appendix:

```
!FORTRAN CODE USED IN EVALUATING THE METHODOLOGY
      USE NUMERICAL_LIBRARIES
       IMPLICIT NONE
      INTEGER ID(1000), I, M, NS,NP
       REAL AGE(1000),FEV(1000),HEIGHT(1000),SEX(1000),SK(1000)
         REAL Y(1000),X(1000),Z(1000)
          REAL ANP,SUMX,SUMY,SUMZ,XM,YM,ZM
        REAL
MU011,MU200,MU020,MU002,MU202,MU210,MU201,MU220,MU012,
      1 MU021,MU022,MU030, MU003, MU400,MU040,MU004
        REAL D202,D210,D201,D220,D012,D021,D022,D030, D003, D400,D040,
      1D004
        REAL D202S,D210S,D201S,D220S,D022S, D400S,D040S,D004S
         REAL ANS, AM, LAMBDA, VARS2,MSED1, MSED2
          REAL RHO2, RHO12, G0122, MSESIGM1
        REAL RHO2S, G0122S, MSESIGM2, MSESIGM3
        REAL C1,C2,HX1,HX2,HYX1,HYX2,HX1X2,GYX1,GYX2,G012SS,RHO2SS
```

```
      REAL RES02,RED1,RED2,RESIGM1,RESIGM2;RESIGM3,SS,RR
      CHARACTER*20 OUT_FILE
      CHARACTER*20 IN_FILE
      WRITE(*,'(A)') 'NAME OF THE OUTPUT FILE'
      READ(*,'(A20)') OUT_FILE
      OPEN(42, FILE=OUT_FILE, STATUS='UNKNOWN')
      WRITE(*,'(A)') 'NAME OF THE INPUT FILE'
      READ(*,'(A20)') IN_FILE
      WRITE(42,'(A)') IN_FILE
      OPEN(41, FILE=IN_FILE, STATUS='OLD')
      READ(41,*)NP
      DO 231 I = 1, NP
231   READ(41,*)ID(I),AGE(I),FEV(I),HEIGHT(I),SEX(I),SK(I)
      DO 245 I = 1, NP
      Y(I) = FEV(I)
      X(I) = AGE(I)
245   Z(I) = HEIGHT(I)
      ANP = NP
      SUMX = 0.0
      SUMY = 0.0
      SUMZ = 0.0
      DO 10 I=1, NP
      SUMX = SUMX + X(I)
      SUMY = SUMY + Y(I)
10    SUMZ = SUMZ + Z(I)
      YM = SUMY/ANP
      XM = SUMX/ANP
      ZM = SUMZ/ANP
      MU011 = 0.0
      MU200 = 0.0
      MU020 = 0.0
      MU002 = 0.0
      MU012 = 0.0
      MU021 = 0.0
      MU202 = 0.0
      MU210 = 0.0
      MU201 = 0.0
      MU220 = 0.0
      MU022 = 0.0
      MU030 = 0.0
      MU003 = 0.0
      MU400 = 0.0
      MU040 = 0.0
      MU004 = 0.0
      DO 11 I=1, NP
      MU011 = MU011 + (X(I)-XM)*(Z(I)-ZM)
```

```
      MU200 = MU200 + (Y(I)-YM)**2
      MU020 = MU020 + (X(I)-XM)**2
      MU002 = MU002 + (Z(I)-ZM)**2
      MU012 = MU012 + (X(I)-XM)*(Z(I)-ZM)**2
      MU202 = MU202 + (Y(I)-YM)**2*(Z(I)-ZM)**2
      MU210 = MU210 + (Y(I)-YM)**2*(X(I)-XM)
      MU021 = MU021 + (X(I)-XM)**2*(Z(I)-ZM)
      MU201 = MU201 + (Y(I)-YM)**2*(Z(I)-ZM)
      MU220 = MU220 + (Y(I)-YM)**2*(X(I)-XM)**2
      MU022 = MU022 + (X(I)-XM)**2*(Z(I)-ZM)**2
      MU030 = MU030 + (X(I)-XM)**3
      MU003 = MU003 + (Z(I)-ZM)**3
      MU400 = MU400 + (Y(I)-YM)**4
      MU040 = MU040 + (X(I)-XM)**4
11    MU004 = MU004 + (Z(I)-ZM)**4
      MU011 = MU011/ANP
      MU200 = MU200/ANP
      MU020 = MU020/ANP
      MU002 = MU002/ANP
      MU012 = MU012/ANP
      MU202 = MU202/ANP
      MU210 = MU210/ANP
      MU201 = MU201/ANP
      MU220 = MU220/ANP
      MU021 = MU021/ANP
      MU022 = MU022/ANP
      MU030 = MU030/ANP
      MU003 = MU003/ANP
      MU400 = MU400/ANP
      MU040 = MU040/ANP
      MU004 = MU004/ANP
      D202 = MU202/(MU200*MU002)
      D210 = MU210/(MU200*SQRT(MU020))
      D201 = MU201/(MU200*SQRT(MU002))
      D220 = MU220/(MU200*MU020)
      D012 = MU012/(SQRT(MU020)*MU002)
      D022 = MU022/(MU020*MU002)
      D021 = MU021/(MU020*SQRT(MU002))
      D030 = MU030/(MU020**(3/2))
      D003 = MU003/(MU002**(3/2))
      D400 = MU400/MU200**2
      D040 = MU040/MU020**2
      D004 = MU004/MU002**2
      D202S = D202-1
      D210S = D210-1
      D201S = D201-1
```

```
            D220S = D220-1
      D022S = D022-1          D400S = D400-1
          D040S = D040-1
            D004S = D004-1
      DO 21 SS = 0.02, 0.10, 0.02
          NS = SS *NP
      DO 22 RR = 0.3, 0.9, 0.1
          WRITE(42, 111)SS, RR
111   FORMAT(2X, 'Sampling Fraction=',F9.5,/2X,'Response Rate=',F9.5)
        M = RR*NS
        WRITE(42,108)NP,NS,M
108   FORMAT(2X,'NP=',I6,1X,'NS=',I5,1X,'M=',I4/)
        AM = M
          ANS = NS
            LAMBDA=(1/AM-1/ANS)
              VARS2 = D400S/AM
      MSED1 = D400S/AM-LAMBDA*D210**2-D201**2/ANS
        MSED2          =          D400S/AM-LAMBDA*(D220S**2/D040S)-
(D202S**2/D004S)/ANS
      RHO2 = D201**2/D400S
        RHO12 = MU011/SQRT(MU020*MU002)
          G0122=(D201**2+D210**2-2*RHO12*D210*D201)/(D400S*(1-
RHO12**2))
      MSESIGM1 = D400S*(1-(ANS-AM)*G0122/ANS-AM*RHO2/ANS)/AM
        RHO2S=D202S**2/(D400S*D004S)
          G0122S               =               (D004S*D220S**2-
2*D220S*D022S*D202S+D040S*D202S)/
    1     (D400S*(D040S*D004S-D022S**2))
        MSESIGM2 = D400S*(1-(ANS-AM)*G0122S/ANS-AM*RHO2S/ANS)/AM
         C1 = SQRT(MU020)/XM
          C2 = SQRT(MU002)/ZM
        HX1 = 4*C1**2-4*D030*C1+D040S
         HX2 = 4*C2**2-4*D003*C2+D004S
           HYX1 = 3*C1**2-2*C1*(D030+D210)+D220S
        HYX2=3*C2**2-3*C2*(D003+D201)+D202S
           HX1X2=4*RHO12*C1*C2-2*D012*C1-2*D021+D022S
            GYX1 = D220S-2*D210*C1
        GYX2 = D202S-2*D201*C2
          G012SS =(HX1*GYX2**2-2*HX1X2*GYX1*GYX2+HX2*GYX1**2)/
    1     (D400S*(HX1*HX2-HX1X2**2))
         RHO2SS = GYX2**2/(HX2*D400S)
          MSESIGM3          =          D400S*(1-(ANS-AM)*G012SS/ANS-
AM*RHO2SS/ANS)/AM
            RES02 = VARS2*100/VARS2
            RED1 = VARS2*100/MSED1
        RED2 = VARS2*100/MSED2
```

```
          RESIGM1 = VARS2*100/MSESIGM1
            RESIGM2 = VARS2*100/MSESIGM2
              RESIGM3 = VARS2*100/MSESIGM3
      WRITE(42,110)NS,M,RES02,RED1,RED2,RESIGM1,RESIGM2, RESIGM3
110   FORMAT(2X,'n=',I4,1X,'m=',I4,1X,'RES02=',F9.3,1X,'RED1=',F9.3,/
     1 1X,'RED2=',F9.3,1X,'RESIGM1=',F9.3,1X,'RESIGM2=',F9.3,/
     1 1X,'RESIGM3=',F9.3)
22    CONTINUE
21    CONTINUE
        STOP
        END
```

References

Chand, L. (1975). *Some Ratio-Type Estimators Based on Two or more Auxiliary Variables*, Ph. D. dissertation, Iowa State University, Ames, Iowa.

Das, A. K. and Tripathi, T. P. (1978). Use of auxiliary information in estimating the finite population variance, *Sankhyā, C*, **40**, 139–148.

Farrell, P. and Singh, S. (2010). Some contributions to Jackknifing Two-Phase sampling estimators, *Survey Methodology*, In Press.

Glesser, L. J. and Healy, J. D. (1976). Estimating the mean of a normal distribution with known coefficient of variation, *Journal of the American Statistical Association*, **71**, 977–981.

Gupta, R. K., Singh, S. and Mangat, N. S. (1992–1993). Some chain ratio type estimators for estimating finite population variance, *Aligarh Journal of Statistics*, **12&13**, 65–69.

Isaki, C. T. (1983). Variance estimation using auxiliary information, *Journal of the American Statistical Association*, **78**, 115–123.

Kiregyera, B. (1980). A chain ratio-type estimator in finite population double sampling using two auxiliary variables, *Metrika*, **27**, 217–223.

Lee, K. H. (1981). Estimation of variance of mean using known coefficient of variation, *Communications in Statistics - Theory and Methods*, **10**, 503–514.

Mukerjee, R., Rao, T. J. and Vijayan, K. (1987). Regression type estimators using multiple auxiliary information, *Australian & New Zealand Journal of Statistics*, **29**, 244–254.

Murthy, M. N. (1967). *Sampling Theory and Methods*, Calcutta Statist, Publishing Society.

Rosner, B. (2005). *Fundamentals of Biostatistics*, Thomson-Brooks/Cole, 6rd Ed.

Rueda, M., Arcos, A., Muñoz, J. F. and Singh, S. (2007). Quantile estimation in two -phase sampling, *Computational Statistics & Data Analysis*, **51**, 2559–2572.

Searls, D. T. (1964). Utilization of a known coefficient of variation in the estimation procedure, *Journal of the American Statistical Association*, **59**, 1225–1228.

Searls, D. T. and Intarapanich, P. (1990). A note on an estimator for the variance that utilizes the kurtosis, *The American Statistician*, **44**, 295–296.

Singh, H. P., Singh, S. and Kim, J.-M. (2006). General families of chain ratio type estimators of the population mean with known coefficient of variation of the second auxiliary variable in two phase sampling, *Journal of the Korean Statistical Society*, **35**, 377–395.

Singh, S. (1990). Estimation of finite population variance using double sampling, *The Aligarh Journal of Statistics*, **11**, 53–56.

Singh, S., Horn, S., Chowdhury, S. and Yu, F. (1999). Calibration of the estimators of variance, *Australian & New Zealand Journal of Statistics*, **41**, 199–212.

Singh, S., Singh, H. P., Tailor, R., Allen, J. and Kazak, M. (2009). Estimation of ratio of two finite-population means in the presence of non-response, *Communications in Statistics-Theory and Methods*, **38**, 3608–3621.

Singh, V. K., Singh, H. P. and Singh, H. P. (1994). A general class of chain estimators for ratio and product of two means of a finite population, *Communications in Statistics-Theory and Methods*, **23**, 1341–1355.

Srivastava, S. K. (1967). An estimator using auxiliary information in sample surveys, *Calcutta Statistical Association Bulletin*, **16**, 121–132.

Srivastava, S. K. (1980). A class of estimators using auxiliary information in sample surveys, *The Canadian Journal of Statistics*, **8**, 253–254.

Srivastava, S. K. and Jhajj, H. S. (1980). A class of estimators using auxiliary information for estimating finite population variance, *Sankhya C*, **42**, 87–96.

Srivastava, S. R., Srivastava, S. p. and Khare, B. B. (1989). Chain ratio type estimator for ratio of two population means using auxiliary characters, *Communications in Statistics-Theory and Methods*, **18**, 3917–3926.

Upadhyaya, L. N., Kushwaha, K. S. and Singh, H. P. (1990). A modified chain ratio-type estimator in two-phase sampling using multi-auxiliary information, *Metron*, **48**, 381–393.

Received January 2010; Accepted February 2010