J. Japan Statist. Soc.
Vol. 35 No. 2 2005 273–286

AN EFFICIENT CLASS OF CHAIN ESTIMATORS OF POPULATION VARIANCE UNDER SUB-SAMPLING SCHEME

H. S. Jhajj*, M. K. Sharma* and Lovleen Kumar Grover**

For estimating the population variance S_y^2 of study variable y, a class of chain estimators of S_y^2 has been proposed in the presence of two auxiliary variables x and z by using known information on population mean and variance of the second auxiliary variable z. In this proposed class, the second auxiliary variable z is directly highly correlated with the first auxiliary variable x, whereas the variable z is correlated with the variable y due to only the high correlation between the variables y and x. Another generalized class of estimators of S_y^2 has also been considered by using the same available information of auxiliary variable z when both the auxiliary variables x and z are directly highly correlated with the study variable y. The asymptotic expressions for the mean square errors and their optimum values have been obtained. A comparison between the two proposed classes of estimators of S_y^2 has been made empirically.

Key words and phrases: Auxiliary variable, chain estimator, consistent estimator, double sampling technique, mean square error, optimum estimator, study variable.

1. Introduction

In manufacturing industries and pharmaceutical laboratories sometimes researchers are interested in the variation of their products. To measure the variations within the values of study variable y, the problem of estimating the population variance of S_y^2 variable y also received a considerable attention of the statistician in the survey sampling. Liu (1974) gave a general class of quadratic estimators for variance and obtained a class of unbiased estimators under certain conditions. Das and Tripathi (1978) defined six estimators of population variance S_y^2 using known information on parameters of auxiliary variable. Using prior information on parameters of auxiliary variable/variables, Srivastava and Jhajj (1980, 1995), Isaki (1983), Singh and Kataria (1990), Prasad and Singh (1990, 1992), Ahmed *et al.* (2000) have defined estimators or classes of estimators of S_y^2. In a situation when prior information on parameters of auxiliary variables is not available, using double sampling technique, Singh and Singh (2001) defined a ratio-type estimator of S_y^2. Ahmed *et al.* (2003) gave some chain ratio-type as well as chain product-type estimators of S_y^2, under two-phase sampling scheme. Al-Jararha and Ahmed (2002) defined two classes of estimators of S_y^2 by using prior information on parameter of one of the two auxiliary variables under double

Received April 28, 2004. Revised July 8, 2004. Accepted February 16, 2005.
*Department of Statistics, Punjabi University, Patiala-147 002, India.
**Department of Mathematics, Guru Nanak Dev University, Amritsar-143 005, India. Email: lovleen_2@yahoo.co.in

sampling scheme.

When the population mean $\bar{X}$ and population variance S_x^2 of auxiliary variable x (highly correlated with study variable y) are known, Srivastava and Jhajj (1980) defined a class of estimators of S_y^2 as

$$\hat{V}_g = g\left(s_y^2, \frac{\bar{x}}{\bar{X}}, \frac{s_x^2}{S_x^2}\right) \tag{1.1}$$

where $g(\cdot,\cdot,\cdot)$ is parametric function satisfying certain regularity conditions; $\bar{x}$, s_x^2 and s_y^2 are sample mean of x and sample variances of x and y respectively for the sample of size n.

If $\bar{X}$ and S_x^2 are unknown then following Srivastava and Jhajj (1987), under double sampling technique, one can define a general class of estimators of population variance S_y^2 as

$$\hat{V}_{gd} = g_d\left(s_y^2, \frac{\bar{x}}{\bar{x}'}, \frac{s_x^2}{s_x'^2}\right) \tag{1.2}$$

where $g_d(\cdot,\cdot,\cdot)$ is a parametric function such that $g_d(S_y^2, 1, 1) = S_y^2$ and satisfies certain regularity conditions; $\bar{x}'$ and $s_x'^2$ are the sample mean and sample variance of variable x for the preliminary large sample of size n'; $\bar{x}$, s_x^2 and s_y^2 are sample mean of x and sample variances of x and y respectively for the sub sample of size $n(n < n')$ under the double sampling technique. In such situation, sometimes the information on population mean $\bar{Z}$ and population variance S_z^2 of another auxiliary variable z, highly correlated with study variable y, is available in advance. Following Srivastava and Jhajj (1980), one can generalize the class of estimators defined in (1.2) as

$$\hat{V}_{Hd} = H\left(s_y^2, \frac{\bar{x}}{\bar{x}'}, \frac{s_x^2}{s_x'^2}, \frac{\bar{z}}{\bar{Z}}, \frac{s_z^2}{S_z^2}\right) \tag{1.3}$$

where $\bar{z}$ and s_z^2 are the sample mean and sample variance of variable z in the second phase sample of double sampling.

In the class $\hat{V}_{Hd}$, both the auxiliary variables x and z are considered to be highly correlated with the study variable y. But sometimes in a trivariate distribution consisting study variable y and two auxiliary variables x and z, in which x is highly correlated with both variables y and z; whereas the variables y and z have no direct correlation with each other but they are just correlated with each other due to only their correlation with variable x, such as

(i) In any agricultural experiment, both the yield of crop (say y) and the labour deployed (say z) are highly correlated with the area under crop (say x). Whereas the yield of crop (y) and the labour deployed (z) are correlated with each other due to only their correlation with the area under crop (x).

(ii) In any repetitive survey, the values of a variable of interest corresponding to both the last to last year (say z) and the current year (say y) are highly correlated with the values of the same variable corresponding to the last year (say x). Whereas the values corresponding to the last to last year (z) and the values corresponding to the current year (y) are correlated with each other due to only their correlation with the values of the same variable corresponding to the last year (x).

In such situations, we propose a class of chain estimators of S_y^2. The word chain estimator means that first improve the estimators $\bar{x}'$ and $s_x'^2$ of $\bar{X}$ and S_x^2 respectively by using known values of population mean $\bar{Z}$ and variance S_z^2 of variable z. In turn, these improved estimators are used for the estimation of S_y^2 which leads to the creation of chain estimators.

The asymptotic expressions for the mean squared errors and their minimum values are obtained for the proposed class of chain estimators and the generalized class $\hat{V}_{Hd}$. It has been shown that the optimum estimator of the class of chain estimators has simpler form as compared to that of the class $\hat{V}_{Hd}$. A comparison among the different classes of estimators of S_y^2 with respect to their mean squared error is also made empirically. This comparison shows that the proposed class of chain estimators is more efficient than the generalized class $\hat{V}_{Hd}$ and hence recommended in practical applications for estimating S_y^2.

2. Notations and expectations

From the population of size N, select a first phase simple random sample of size n' and observe both the variables x and z for the selected units. A second phase simple random sample of size n ($n < n'$) is selected from the first phase sample and variables x, y and z are measured on these selected units. Let the values of variables x, y and z be denoted by X_j, Y_j and Z_j respectively on the j-th unit of the population; $j = 1, 2, \ldots, N$ and the corresponding small letters x_j, y_j and z_j denote the sample values.

We write

$$\bar{Y} = \frac{1}{N}\sum_{j=1}^{N} Y_j, \qquad \bar{X} = \frac{1}{N}\sum_{j=1}^{N} X_j, \qquad \bar{Z} = \frac{1}{N}\sum_{j=1}^{N} Z_j$$

$$S_y^2 = \frac{1}{N-1}\sum_{j=1}^{N}(Y_j - \bar{Y})^2, \qquad S_x^2 = \frac{1}{N-1}\sum_{j=1}^{N}(X_j - \bar{X})^2,$$

$$S_z^2 = \frac{1}{N-1}\sum_{j=1}^{N}(Z_j - \bar{Z})^2$$

$$\mu_{rst} = \frac{1}{N}\sum_{j=1}^{N}(Y_j - \bar{Y})^r (X_j - \bar{X})^s (Z_j - \bar{Z})^t, \qquad \lambda_{rst} = \frac{\mu_{rst}}{\mu_{200}^{r/2}\mu_{020}^{s/2}\mu_{002}^{t/2}}.$$

Obviously

$$C_0 = \frac{S_y}{\bar{Y}}, \qquad C_1 = \frac{S_x}{\bar{X}}, \qquad C_2 = \frac{S_z}{\bar{Z}}$$

$$\rho_{yx} = \rho_{01} = \lambda_{110}, \qquad \rho_{yz} = \rho_{02} = \lambda_{101}, \qquad \rho_{xz} = \rho_{12} = \lambda_{011}.$$

Let $\bar{z}'$ and $s_z'^2$ denote the sample mean and sample variance of variable z for the first phase sample of size n'. In this paper, all the sampling variances have been defined either with divisor $n'-1$ or $n-1$ depending on first phase sample or second phase sample respectively.

Letting

$$\omega = \frac{s_y^2}{S_y^2}, \qquad u_1 = \frac{\bar{x}}{\bar{x}'}, \qquad v_1 = \frac{\bar{z}}{\bar{Z}}, \qquad v_1' = \frac{\bar{z}'}{\bar{Z}},$$
$$u_2 = \frac{s_x^2}{s_x'^2}, \qquad v_2 = \frac{s_z^2}{S_z^2}, \qquad v_2' = \frac{s_z'^2}{S_z^2}.$$

For the sake of simplicity, assume that N is large enough as compared to n and n' so that all finite population correction (fpc) terms are ignored. For the given double sampling technique when both the samples drawn are simple random samples (without replacement), we have the following expectations:

$$\begin{aligned}
&E(\omega) = E(u_1) = E(v_1) = E(v_1') = E(u_2) = E(v_2) = E(v_2') = 1\\
&E(u_1-1)(v_1'-1) = E(u_1-1)(v_2'-1) = E(u_2-1)(v_1'-1)\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad = E(u_2-1)(v_2'-1) = 0\\
&E(v_1-1)^2 = \frac{1}{n}C_2^2, \qquad E(v_1'-1)^2 = \frac{1}{n'}C_2^2
\end{aligned}$$

and up to the terms of order n^{-1}, we have

$$\begin{aligned}
&E(\omega-1)^2 = \frac{1}{n}(\lambda_{400}-1), \qquad E(u_1-1)^2 = \left(\frac{1}{n}-\frac{1}{n'}\right)C_1^2\\
&E(u_2-1)^2 = \left(\frac{1}{n}-\frac{1}{n'}\right)(\lambda_{040}-1), \qquad E(v_2-1)^2 = \frac{1}{n}(\lambda_{004}-1)\\
&E(v_2'-1)^2 = \frac{1}{n'}(\lambda_{004}-1), \qquad E(\omega-1)(u_1-1) = \left(\frac{1}{n}-\frac{1}{n'}\right)\lambda_{210}C_1\\
&E(\omega-1)(v_1-1) = \frac{1}{n}\lambda_{201}C_2, \qquad E(\omega-1)(v_1'-1) = \frac{1}{n'}\lambda_{201}C_2\\
&E(\omega-1)(u_2-1) = \left(\frac{1}{n}-\frac{1}{n'}\right)(\lambda_{220}-1), \qquad E(\omega-1)(v_2-1) = \frac{1}{n}(\lambda_{202}-1)\\
&E(\omega-1)(v_2'-1) = \frac{1}{n'}(\lambda_{202}-1), \qquad E(u_1-1)(v_1-1) = \left(\frac{1}{n}-\frac{1}{n'}\right)\lambda_{011}C_1C_2\\
&E(v_1-1)(v_2-1) = \frac{1}{n}\lambda_{003}C_2, \qquad E(v_1'-1)(v_2'-1) = \frac{1}{n'}\lambda_{003}C_2\\
&E(u_1-1)(u_2-1) = \left(\frac{1}{n}-\frac{1}{n'}\right)\lambda_{030}C_1,\\
&E(u_1-1)(v_2-1) = \left(\frac{1}{n}-\frac{1}{n'}\right)\lambda_{012}C_1\\
&E(v_1-1)(u_2-1) = \left(\frac{1}{n}-\frac{1}{n'}\right)\lambda_{021}C_2,\\
&E(u_2-1)(v_2-1) = \left(\frac{1}{n}-\frac{1}{n'}\right)(\lambda_{022}-1).
\end{aligned}$$

3. Proposed class of chain estimators of S_y^2

Suppose that mean $\bar{X}$ and variance S_x^2 of first auxiliary variable x are unknown but mean $\bar{Z}$ and variance S_z^2 of second auxiliary variable z are known in advance. It is assumed that the variable z is highly correlated with variable x whereas the correlation between the variables y and z exists due to only the high correlation of variable x with the variables y and z. In such situations, we propose a class of chain estimators of S_y^2 as

$$\begin{aligned} \hat{V}_{Td} &= T\left(s_y^2, \frac{\bar{x}}{\bar{x}'}, \frac{s_x^2}{s_x'^2}, \frac{\bar{z}'}{\bar{Z}}, \frac{s_z'^2}{S_z^2}\right) \\ &= T(s_y^2, u_1, u_2, v_1', v_2') \end{aligned} \tag{3.1}$$

where $T(\cdot,\cdot,\cdot,\cdot,\cdot)$ is a parametric function of s_y^2, u_1, u_2, v_1' and v_2' such that

$$T(S_y^2, 1, 1, 1, 1) = S_y^2, \qquad \text{for all} \quad S_y^2. \tag{3.2}$$

Whatever sample is chosen, let the point $(s_y^2, u_1, u_2, v_1', v_2')$ assume values in a bounded, closed convex subset R of the five dimensional real space containing the point $(S_y^2, 1, 1, 1, 1)$. The function $T(\cdot,\cdot,\cdot,\cdot,\cdot)$ is continuous and bounded having continuous and bounded first and second order partial derivatives in R.

Since there are only a finite number of samples therefore under the above conditions, the expectation and the mean square error of the estimators of the class $\hat{V}_{Td}$ exist. On using second order Taylor's series expansion of $T(s_y^2, u_1, u_2, v_1', v_2')$ about the point $(S_y^2, 1, 1, 1, 1)$, the mean square error (MSE) of $\hat{V}_{Td}$, up to the terms of order n^{-1}, is

$$\begin{aligned} MSE(\hat{V}_{Td}) &= \frac{1}{n} S_y^4(\lambda_{400} - 1) \\ &\quad + \frac{1}{n'}\{C_2^2 T_4^2 + (\lambda_{004} - 1)T_5^2 + 2C_2 S_y^2 \lambda_{201} T_4 \\ &\qquad + 2S_y^2(\lambda_{202} - 1)T_5 + 2C_2\lambda_{003}T_4T_5\} \\ &\quad + \left(\frac{1}{n} - \frac{1}{n'}\right)\{C_1^2 T_2^2 + (\lambda_{040} - 1)T_3^2 + 2C_1 S_y^2 \lambda_{210} T_2 \\ &\qquad + 2S_y^2(\lambda_{220} - 1)T_3 + 2C_1\lambda_{030}T_2T_3\} \end{aligned} \tag{3.3}$$

where T_i; $i = 2, 3, 4, 5$ denote the first order partial derivatives of $T(s_y^2, u_1, u_2, v_1', v_2')$ with respect to u_1, u_2, v_1' and v_2' at the point $(S_y^2, 1, 1, 1, 1)$ respectively.

The MSE of $\hat{V}_{Td}$ as given in (3.3) is minimized for

$$T_2 = \frac{S_y^2}{C_1}\left(\frac{\lambda_{030}(\lambda_{220} - 1) - \lambda_{210}(\lambda_{040} - 1)}{\lambda_{040} - \lambda_{030}^2 - 1}\right) \tag{3.4}$$

$$T_3 = S_y^2\left(\frac{\lambda_{030}\lambda_{210} - \lambda_{220} + 1}{\lambda_{040} - \lambda_{030}^2 - 1}\right) \tag{3.5}$$

$$T_4 = \frac{S_y^2}{C_2}\left(\frac{\lambda_{003}\,(\lambda_{202} - 1) - \lambda_{201}\,(\lambda_{004} - 1)}{\lambda_{004} - \lambda_{003}^2 - 1}\right) \tag{3.6}$$

$$T_5 = S_y^2\left(\frac{\lambda_{003}\lambda_{201} - \lambda_{202} + 1}{\lambda_{004} - \lambda_{003}^2 - 1}\right) \tag{3.7}$$

and minimum mean square error of $\hat{V}_{Td}$, up to the terms of order n^{-1}, is

$$\begin{aligned}
(3.8)\quad & Min.MSE(\hat{V}_{Td}) \\
&= S_y^4\left[\frac{1}{n}(\lambda_{400}-1)-\frac{1}{n'}\left\{\lambda_{201}^2+\frac{(\lambda_{202}-\lambda_{201}\lambda_{003}-1)^2}{\lambda_{004}-\lambda_{003}^2-1}\right\}\right. \\
&\qquad \left.-\left(\frac{1}{n}-\frac{1}{n'}\right)\left\{\lambda_{210}^2+\frac{(\lambda_{220}-\lambda_{210}\lambda_{030}-1)^2}{\lambda_{040}-\lambda_{030}^2-1}\right\}\right] \\
(3.9)\quad &= Min.MSE(\hat{V}_{gd})-\frac{1}{n'}S_y^4\left\{\lambda_{201}^2+\frac{(\lambda_{202}-\lambda_{201}\lambda_{003}-1)^2}{\lambda_{004}-\lambda_{003}^2-1}\right\}
\end{aligned}$$

where $Min.MSE(\hat{V}_{gd})$ is the minimum asymptotic mean square error of the estimators of the class $\hat{V}_{gd}$, up to the terms of order n^{-1}, and is given by

$$\begin{aligned}
(3.10)\quad Min.MSE(\hat{V}_{gd}) = S_y^4&\left[\frac{1}{n}(\lambda_{400}-1)-\left(\frac{1}{n}-\frac{1}{n'}\right)\right. \\
&\left.\times\left\{\lambda_{210}^2+\frac{(\lambda_{220}-\lambda_{210}\lambda_{030}-1)^2}{\lambda_{040}-\lambda_{030}^2-1}\right\}\right].
\end{aligned}$$

Rewriting (3.9), we have

$$\begin{aligned}
(3.11)\quad & Min.MSE(\hat{V}_{gd})-Min.MSE(\hat{V}_{Td}) \\
&= \frac{1}{n'}S_y^4\left\{\lambda_{201}^2+\frac{(\lambda_{202}-\lambda_{201}\lambda_{003}-1)^2}{\lambda_{004}-\lambda_{003}^2-1}\right\}\geq 0.
\end{aligned}$$

In (3.11), the right hand side is the sum of two non-negative quantities, since $\lambda_{004}-\lambda_{003}^2-1\geq 0$ always. Thus we found that $Min.MSE(\hat{V}_{Td})$ is always smaller than $Min.MSE(\hat{V}_{gd})$.

4. Comparison of the class $\hat{V}_{Td}$ with the class $\hat{V}_{Hd}$

To compare the generalized class $\hat{V}_{Hd} = H(s_y^2, u_1, u_2, v_1, v_2)$ with the proposed class of chain estimators $\hat{V}_{Td} = T(s_y^2, u_1, u_2, v_1', v_2')$, we require the mean square error of $\hat{V}_{Hd}$. Proceeding in the same way as in Section 3, the asymptotic mean square error of $\hat{V}_{Hd}$ (up to the terms of order n^{-1}) is

$$\begin{aligned}
(4.1)\quad MSE(\hat{V}_{Hd}) = {}& \frac{1}{n}(\lambda_{400}-1)S_y^4 \\
&+\frac{1}{n}[C_2^2H_4^2+(\lambda_{004}-1)H_5^2+2C_2S_y^2\lambda_{201}H_4 \\
&\qquad +2S_y^2(\lambda_{202}-1)H_5+2C_2\lambda_{003}H_4H_5] \\
&+\left(\frac{1}{n}-\frac{1}{n'}\right)[C_1^2H_2^2+(\lambda_{040}-1)H_3^2+2C_1S_y^2\lambda_{210}H_2 \\
&\qquad +2S_y^2(\lambda_{220}-1)H_3+2C_1\lambda_{030}H_2H_3 \\
&\qquad +2C_1C_2\rho_{12}H_2H_4+2C_1\lambda_{012}H_2H_5 \\
&\qquad +2C_2\lambda_{021}H_3H_4+2(\lambda_{022}-1)H_3H_5]
\end{aligned}$$

where H_i; $i = 2, 3, 4, 5$ denote the first order partial derivatives of the function $H(s_y^2, u_1, u_2, v_1, v_2)$ with respect to u_1, u_2, v_1 and v_2 at the point $(S_y^2, 1, 1, 1, 1)$ respectively.

The $MSE(\hat{V}_{Hd})$ is minimized for

$$H_2 = -\frac{S_y^2}{C_1}\left\{\frac{\rho_{12}(m\lambda_{201} - \rho_{12}\lambda_{210})}{m - \rho_{12}^2} + \lambda_{210}\right\} + \frac{1}{C_1}\left\{H_5(\lambda_{012} - k\rho_{12}) - (H_3 + \theta S_y^2 H_5)\left(\lambda_{030} - \frac{L_3\rho_{12}}{m - \rho_{12}^2}\right)\right\} \tag{4.2}$$

$$H_3 = S_y^2\left[\frac{\frac{L_3}{m-\rho_{12}^2}(m\lambda_{201} - \rho_{12}\lambda_{210}) - L_1 - H_5(L_4 + L_3 k)}{L_2 - \frac{L_3^2}{m-\rho_{12}^2}}\right] \tag{4.3}$$

$$H_4 = -\frac{S_y^2}{C_2}\left(\frac{m\lambda_{201} - \rho_{12}\lambda_{210}}{m - \rho_{12}^2}\right) - \frac{1}{C_2}\left[\frac{L_3 H_3}{m - \rho_{12}^2} + H_5\left\{k - \frac{L_3\theta(1 + S_y^2)}{m - \rho_{12}^2}\right\}\right] \tag{4.4}$$

$$H_5 = S_y^4\left[\frac{\lambda_{210}\left\{\lambda_{012} - \rho_{12}k - \theta\left(\lambda_{030} - \frac{L_3\rho_{12}}{m-\rho_{12}^2}\right)\right\} + m\lambda_{201}\left(k - \frac{L_3\theta}{m-\rho_{12}^2}\right) + \theta(\lambda_{220} - 1) - m(\lambda_{202} - 1)}{m(\lambda_{004} - 1) - \lambda_{012}^2 - k(m\lambda_{003} - \rho_{12}\lambda_{012})}\right] \tag{4.5}$$

where

$$L_1 = \lambda_{220} - \lambda_{210}\lambda_{030} - 1, \qquad L_2 = \lambda_{040} - \lambda_{030}^2 - 1,$$
$$L_3 = \lambda_{021} - \rho_{12}\lambda_{030}, \qquad L_4 = \lambda_{022} - \lambda_{012}\lambda_{030} - 1$$
$$m = \frac{n'}{n' - n}, \qquad k = \frac{m\lambda_{003} - \rho_{12}\lambda_{012}}{m - \rho_{12}^2}, \qquad \theta = \frac{L_4 - L_3 k}{L_2 - \frac{L_3^2}{m-\rho_{12}^2}}$$

and the minimum mean square error of $\hat{V}_{Hd}$, up to the terms of order n^{-1}, is given by

$$\begin{aligned} \frac{Min.MSE(\hat{V}_{Hd})}{S_y^4} &= \frac{1}{n}(\lambda_{400} - 1) - \left(\frac{1}{n} - \frac{1}{n'}\right)\left(\lambda_{210}^2 + \frac{L_1^2}{L_2}\right) \\ &- \frac{1}{n}\frac{\left\{L_2\left(\lambda_{201} - \frac{\rho_{12}\lambda_{210}}{m}\right) - \frac{L_1 L_3}{m}\right\}^2}{L_2\left\{L_2\left(1 - \frac{\rho_{12}^2}{m}\right) - \frac{L_3^2}{m}\right\}} \\ &- \frac{1}{n}\frac{\left[\begin{array}{l}\left\{L_2\left(1 - \frac{\rho_{12}^2}{m}\right) - \frac{L_3^2}{m}\right\}\left(\frac{L_5\lambda_{210}}{m} - L_9\right) - \left(\lambda_{003}L_2 - \frac{L_7}{m}\right)\left(\frac{L_8\lambda_{210}}{m} - L_2\lambda_{201}\right) \\ \quad + \left\{L_4\left(1 - \frac{\rho_{12}^2}{m}\right) - L_3\left(\lambda_{003} - \frac{\rho_{12}\lambda_{012}}{m}\right)\right\}\left(\frac{L_{10}}{m}\right)\end{array}\right]^2}{L_2\left\{L_2\left(1 - \frac{\rho_{12}^2}{m}\right) - \frac{L_3^2}{m}\right\}\left[\left\{L_2\left(1 - \frac{\rho_{12}^2}{m}\right) - \frac{L_3^2}{m}\right\}\left(L_{11} - \frac{L_6}{m}\right) - \left(\lambda_{003}L_2 - \frac{L_7}{m}\right)^2\right]} \end{aligned} \tag{4.6}$$

$$= \frac{Min.MSE(\hat{V}_{gd})}{S_y^4} - \frac{1}{n} \frac{\left\{L_2\left(\lambda_{201} - \frac{\rho_{12}\lambda_{210}}{m}\right) - \frac{L_1L_3}{m}\right\}^2}{L_2\left\{L_2\left(1 - \frac{\rho_{12}^2}{m}\right) - \frac{L_3^2}{m}\right\}}$$

$$- \frac{1}{n} \frac{\left[\begin{array}{l} \left\{L_2\left(1 - \frac{\rho_{12}^2}{m}\right) - \frac{L_3^2}{m}\right\}\left(\frac{L_5\lambda_{210}}{m} - L_9\right) - \left(\lambda_{003}L_2 - \frac{L_7}{m}\right)\left(\frac{L_8\lambda_{210}}{m} - L_2\lambda_{201}\right) \\ + \left\{L_4\left(1 - \frac{\rho_{12}^2}{m}\right) - L_3\left(\lambda_{003} - \frac{\rho_{12}\lambda_{012}}{m}\right)\right\}\left(\frac{L_{10}}{m}\right) \end{array}\right]^2}{L_2\left\{L_2\left(1 - \frac{\rho_{12}^2}{m}\right) - \frac{L_3^2}{m}\right\}\left[\left\{L_2\left(1 - \frac{\rho_{12}^2}{m}\right) - \frac{L_3^2}{m}\right\}\left(L_{11} - \frac{L_6}{m}\right) - \left(\lambda_{003}L_2 - \frac{L_7}{m}\right)^2\right]}$$

where

$$L_5 = \lambda_{012} - \lambda_{030}L_4, \quad L_6 = \lambda_{012}^2 L_2 + L_4^2, \quad L_7 = \rho_{12}\lambda_{012}L_2 + L_3L_4,$$
$$L_8 = \rho_{12}L_2 - \lambda_{030}L_3, \quad L_9 = L_2(\lambda_{202} - 1), \quad L_{10} = L_2(\lambda_{220} - 1),$$
$$L_{11} = L_2(\lambda_{004} - 1).$$

The condition, under which the optimum estimator of the class $\hat{V}_{Td}$ is more efficient than that of the class $\hat{V}_{Hd}$, is obtained by using the expressions (3.9) and (4.6) in

$$Min.MSE(\hat{V}_{Td}) < Min.MSE(\hat{V}_{Hd}). \tag{4.7}$$

From (4.7), we are not able to get a concrete mathematical result about the efficiency of the class of chain estimators $\hat{V}_{Td}$ over the generalized class $\hat{V}_{Hd}$. To have an idea about the efficiency of one estimator over the other, we have considered the following specific cases:

Case 1. When (y, x, z) assumed to follow trivariate normal distribution then all odd ordered moments are vanish to zero. In this case the expressions (3.8) and (4.6) respectively reduce to

$$\frac{Min.MSE(\hat{V}_{Td})}{S_y^4} = \frac{1}{n}(\lambda_{400} - 1) - \left(\frac{1}{n} - \frac{1}{n'}\right)\frac{(\lambda_{220} - 1)^2}{\lambda_{040} - 1} - \frac{1}{n'}\frac{(\lambda_{202} - 1)^2}{\lambda_{004} - 1} \tag{4.8}$$

and

$$\frac{Min.MSE(\hat{V}_{Hd})}{S_y^4} = \frac{1}{n}(\lambda_{400} - 1) - \left(\frac{1}{n} - \frac{1}{n'}\right)\frac{(\lambda_{220} - 1)^2}{\lambda_{040} - 1} - \frac{1}{n}\frac{\left\{(\lambda_{202} - 1)(\lambda_{040} - 1) - \frac{(\lambda_{022}-1)(\lambda_{220}-1)}{m}\right\}^2}{(\lambda_{040} - 1)\left\{(\lambda_{040} - 1)(\lambda_{004} - 1) - \frac{(\lambda_{022}-1)^2}{m}\right\}}. \tag{4.9}$$

Using (4.8) and (4.9), we have

$$\frac{Min.MSE(\hat{V}_{Hd}) - Min.MSE(\hat{V}_{Td})}{S_y^4} \tag{4.10}$$
$$= \frac{1}{n'}\frac{(\lambda_{202} - 1)^2}{(\lambda_{004} - 1)} - \frac{1}{n}\frac{\left\{(\lambda_{202} - 1)(\lambda_{040} - 1) - \frac{(\lambda_{022}-1)(\lambda_{220}-1)}{m}\right\}^2}{(\lambda_{040} - 1)\left\{(\lambda_{040} - 1)(\lambda_{004} - 1) - \frac{(\lambda_{022}-1)^2}{m}\right\}}.$$

Case 2. When $n = n'$, we see that the two proposed classes $\hat{V}_{Td}$ and $\hat{V}_{Hd}$ of estimators of S_y^2 coincide with each other, that is, $\hat{V}_{Td} = \hat{V}_{Hd} = \hat{V}$ (say) which is defined as $\hat{V} \equiv \hat{V}(s_y^2, \frac{\bar{z}}{\bar{Z}}, \frac{s_z^2}{S_z^2}) = \hat{V}(s_y^2, \nu_1, \nu_2)$.

Therefore, up to the terms of order n^{-1}, the minimum mean square error of $\hat{V}$ is given by

$$\text{(4.11)} \qquad \frac{Min.MSE(\hat{V})}{S_y^4} = \frac{1}{n}\left[(\lambda_{400} - 1) - \left\{\lambda_{201}^2 + \frac{(\lambda_{202} - \lambda_{201}\lambda_{003} - 1)^2}{\lambda_{004} - \lambda_{003}^2 - 1}\right\}\right].$$

Using (3.8) and (4.11), we note that

$$\text{(4.12)} \qquad \frac{Min.MSE(\hat{V}) - Min.MSE(\hat{V}_{Td})}{S_y^4} = \left(\frac{1}{n} - \frac{1}{n'}\right)\left[(\lambda_{210}^2 - \lambda_{201}^2) + \left\{\frac{(\lambda_{220} - \lambda_{210}\lambda_{030} - 1)^2}{\lambda_{040} - \lambda_{030}^2 - 1} - \frac{(\lambda_{202} - \lambda_{201}\lambda_{003} - 1)^2}{\lambda_{004} - \lambda_{003}^2 - 1}\right\}\right].$$

For trivariate normal distribution (4.12) reduces to

$$\text{(4.13)} \qquad \frac{Min.MSE(\hat{V}) - Min.MSE(\hat{V}_{Td})}{S_y^4} = \left(\frac{1}{n} - \frac{1}{n'}\right)\left\{\frac{(\lambda_{220} - 1)^2}{\lambda_{040} - 1} - \frac{(\lambda_{202} - 1)^2}{\lambda_{004} - 1}\right\}.$$

We see that even in the specific cases considered above we could not find a concrete mathematical result showing the efficiency of one over the other. But it seems that $\hat{V}_{Td}$ must be better than $\hat{V}_{Hd}$ because $\hat{V}_{Td}$ makes full use of information on n' observations of first phase sample whereas $\hat{V}_{Hd}$ waste the information on $n' - n$ observations of the sample.

Remark 4.1. It should be noted that the efficient use of the estimators of the two proposed classes $\hat{V}_{Hd}$ and $\hat{V}_{Td}$ presumes that the optimum values of H_i and T_i; $i = 2, 3, 4, 5$ are known. But these optimum values are functions of unknown population parameters. Srivastava and Jhajj (1983) have shown that the estimators of the class with estimated values of optimum parameters obtained by their consistent estimators, attain the same minimum mean square error of estimators of the class based on optimum values, up to the first order of approximation. Although by using the same approach of Srivastava and Jhajj (1983), we can construct the estimators of the classes $\hat{V}_{Hd}$ and $\hat{V}_{Td}$. But the optimum estimator of the class $\hat{V}_{Hd}$ is very much complicated as compared to that of $\hat{V}_{Td}$. So due to such type of complexities, we should prefer the proposed class $\hat{V}_{Td}$ of the chain estimators as compared to the general class $\hat{V}_{Hd}$, in practice.

Remark 4.2. The proposed classes $\hat{V}_{Td}$ and $\hat{V}_{Hd}$ of estimators of S_y^2 are very large. Any parametric function $T(s_y^2, u_1, u_2, v_1', v_2')$ (or $H(s_y^2, u_1, u_2, v_1, v_2)$) satisfying certain regularity conditions and $T(S_y^2, 1, 1, 1, 1) = S_y^2$ (or $H(S_y^2, 1, 1, 1, 1) = S_y^2$) for all S_y^2, can generate an estimator of the class $\hat{V}_{Td}$ (or $\hat{V}_{Hd}$). For example, we have the following functions which generate some of the simple estimators of these classes $\hat{V}_{Td}$ and $\hat{V}_{Hd}$:

1. $T(s_y^2, u_1, u_2, v_1', v_2') = s_y^2 u_1^{\alpha} u_2^{\beta} v_1'^{\gamma} v_2'^{\delta}$ and
   $H(s_y^2, u_1, u_2, v_1, v_2) = s_y^2 u_1^{\alpha} u_2^{\beta} v_1^{\gamma} v_2^{\delta}$
2. $T(s_y^2, u_1, u_2, v_1', v_2') = s_y^2[\alpha(u_1 - 1) + \beta(u_2 - 1) + \gamma(v_1' - 1) + \delta(v_2' - 1)]$ and
   $H(s_y^2, u_1, u_2, v_1, v_2) = s_y^2[\alpha(u_1 - 1) + \beta(u_2 - 1) + \gamma(v_1 - 1) + \delta(v_2 - 1)]$
3. $T(s_y^2, u_1, u_2, v_1', v_2') = s_y^2[a_1 u_1^{\alpha} u_2^{\beta} + a_2 v_1'^{\gamma} v_2'^{\delta}]; \quad a_1 + a_2 = 1$ and
   $H(s_y^2, u_1, u_2, v_1, v_2) = s_y^2[a_1 u_1^{\alpha} u_2^{\beta} + a_2 v_1^{\gamma} v_2^{\delta}]; \quad a_1 + a_2 = 1$
4. $T(s_y^2, u_1, u_2, v_1', v_2') = \dfrac{s_y^4 + s_y^2\{\alpha(u_1 - 1) + \beta(u_2 - 1)\}}{s_y^2 + \gamma(v_1' - 1) + \delta(v_2' - 1)}$ and
   $H(s_y^2, u_1, u_2, v_1, v_2) = \dfrac{s_y^4 + s_y^2\{\alpha(u_1 - 1) + \beta(u_2 - 1)\}}{s_y^2 + \gamma(v_1 - 1) + \delta(v_2 - 1)}$
5. $T(s_y^2, u_1, u_2, v_1', v_2') = \dfrac{s_y^2 e^{\alpha(u_1 - 1) + \beta(u_2 - 1)}}{1 + \gamma(v_1' - 1) + \delta(v_2' - 1)}$ and
   $H(s_y^2, u_1, u_2, v_1, v_2) = \dfrac{s_y^2 e^{\alpha(u_1 - 1) + \beta(u_2 - 1)}}{1 + \gamma(v_1 - 1) + \delta(v_2 - 1)}$.

Here the optimum values of α, β, γ and δ in these estimators are so determined that they satisfy the respective normal equations and the resulting estimators should have the same minimum asymptotic mean square errors as given in (3.8) and (4.6) respectively.

5. Comparison of the proposed class $\hat{V}_{Td}$ with the existing ones

On using the information on variances alone for the two auxiliary variables x and z, Al-Jararha and Ahmed (2002) defined a class of chain estimators of S_y^2 as:

$$\hat{b}_g = f_1\left(s_y^2, \frac{s_x^2}{s_x'^2}, \frac{s_z'^2}{S_z^2}\right) \quad \text{or} \tag{5.1}$$

$$\hat{b}_g = f_1(s_y^2, u_2, \nu_2').$$

Up to the terms of order n^{-1}, the minimum mean square error of $\hat{b}_g$ is given by

$$\frac{Min.MSE(\hat{b}_g)}{S_y^4} = \frac{1}{n}(\lambda_{400} - 1) - \left(\frac{1}{n} - \frac{1}{n'}\right)\frac{(\lambda_{220} - 1)^2}{\lambda_{040} - 1} - \frac{1}{n'}\frac{(\lambda_{202} - 1)^2}{\lambda_{004} - 1}. \tag{5.2}$$

In the present paper, using information on variances of the two auxiliary variables x and z along with their means, we have proposed the two classes $\hat{V}_{Hd}$ and $\hat{V}_{Td}$ of estimators of S_y^2. Obviously our proposed class of chain estimators $\hat{V}_{Td}$ is a generalization of the class $\hat{b}_g$. On using (3.8) and (5.2), we have

$$\begin{aligned}(5.3)\qquad &\frac{Min.MSE(\hat{b}_g) - Min.MSE(\hat{V}_{Td})}{S_y^4}\\ &= \left(\frac{1}{n} - \frac{1}{n'}\right)\frac{\{\lambda_{030}(\lambda_{220}-1) - \lambda_{210}(\lambda_{040}-1)\}^2}{(\lambda_{040}-1)(\lambda_{040}-\lambda_{030}^2-1)}\\ &\quad + \frac{1}{n'}\frac{\{\lambda_{003}(\lambda_{202}-1) - \lambda_{201}(\lambda_{004}-1)\}^2}{(\lambda_{004}-1)(\lambda_{004}-\lambda_{003}^2-1)} \geq 0.\end{aligned}$$

From (5.3), we found that the proposed class of chain estimators of S_y^2 i.e. $\hat{V}_{Td}$ is always more efficient than the existing class $\hat{b}_g$. The sign of equality will hold in (5.3) if (y, x, z) follows trivariate normal distribution, that is, the two classes $\hat{V}_{Td}$ and $\hat{b}_g$ become equally efficient in trivariate normal distribution. So it is interesting to note that if (y, x, z) has trivariate normal distribution then the available information regarding means of auxiliary variables becomes useless for the estimation of population variance S_y^2.

In the same paper, Al-Jararha and Ahmed also defined another wider class of estimators of S_y^2 using the same information on variances alone for the two auxiliary variables x and z as

$$(5.4)\qquad \hat{b}_h = f_2\left(s_y^2, \frac{s_x^2}{s_x'^2}, \frac{s_z^2}{s_z'^2}, \frac{s_z'^2}{S_z^2}\right).$$

Up to the terms of order n^{-1}, the minimum mean square error of $\hat{b}_h$ is given by

$$\begin{aligned}(5.5)\qquad &\frac{Min.MSE(\hat{b}_h)}{S_y^4}\\ &= \frac{1}{n}(\lambda_{400}-1) - \left(\frac{1}{n} - \frac{1}{n'}\right)\frac{(\lambda_{220}-1)^2}{\lambda_{040}-1} - \frac{1}{n'}\frac{(\lambda_{202}-1)^2}{\lambda_{004}-1}\\ &\quad - \left(\frac{1}{n} - \frac{1}{n'}\right)\frac{\{(\lambda_{220}-1)(\lambda_{022}-1) - (\lambda_{202}-1)(\lambda_{040}-1)\}^2}{(\lambda_{040}-1)\{(\lambda_{040}-1)(\lambda_{004}-1) - (\lambda_{022}-1)^2\}}.\end{aligned}$$

Using (3.8) and (5.5), we have

$$\begin{aligned}(5.6)\qquad &\frac{Min.MSE(\hat{b}_h) - Min.MSE(\hat{V}_{Td})}{S_y^4}\\ &= \frac{1}{n'}\frac{\{(\lambda_{202}-1)\lambda_{003} - (\lambda_{004}-1)\lambda_{201}\}^2}{(\lambda_{004}-1)(\lambda_{004}-\lambda_{003}^2-1)}\\ &+ \left(\frac{1}{n} - \frac{1}{n'}\right)\end{aligned}$$

$$\times \frac{\begin{bmatrix} (\lambda_{004}-1)\{\lambda_{210}(\lambda_{040}-1)-\lambda_{030}(\lambda_{220}-1)\}^2 \\ -\{(\lambda_{220}-1)(\lambda_{022}-1)-(\lambda_{202}-1)(\lambda_{040}-1)\}^2 \\ -(\lambda_{040}-1)\{\lambda_{210}^2(\lambda_{022}-1)^2-\lambda_{030}^2(\lambda_{202}-1)^2\} \\ +2\lambda_{030}(\lambda_{220}-1)(\lambda_{022}-1)\{\lambda_{210}(\lambda_{022}-1)-\lambda_{030}(\lambda_{202}-1)\} \end{bmatrix}}{(\lambda_{040}-\lambda_{030}^2-1)\left\{(\lambda_{040}-1)(\lambda_{004}-1)-(\lambda_{022}-1)^2\right\}}.$$

From (5.6), no concrete decision regarding the efficiency of one over the other can be drawn.

6. Numerical illustration

Since in the Sections 4 and 5, we could not obtained any concrete theoretical conditions under which proposed class $\hat{V}_{Td}$ is better than $\hat{V}_{Hd}$ and $\hat{b}_h$. So to have a rough idea about the efficiencies of the proposed class of chain estimators $\hat{V}_{Td}$ and the generalized class of estimators $\hat{V}_{Hd}$, we take the two empirical populations considered in the literature. The source of population; nature of the variables y, x and z; population size N and various possible population correlation coefficients are given in Table 1. The values of requisite population parameters for the two

Table 1. Description of populations.

S.No.	Source	y	x	z	N	ρ_{yx}	ρ_{yz}	ρ_{xz}
1.	Murthy (1967) Page 288 Factories 1–40	Output	Fixed Capital	No. of Workers	40	0.950433	0.902977	0.954177
2.	Murthy (1967) Page 399 Villages 1–34	Area under Wheat in 1964	Area under Wheat in 1963	Cultivated Area in 1961	34	0.980086	0.904262	0.909276

Table 2. Parameters of two populations.

Parameters	Values of parameters	
	Population No.1	Population No.2
λ_{030}	−0.364406	0.934467
λ_{040}	2.241570	2.912271
λ_{021}	−0.119179	0.829089
λ_{012}	0.085446	0.727467
λ_{022}	1.714356	2.737891
λ_{003}	0.264201	0.654033
λ_{004}	1.776776	2.808238
λ_{210}	−0.546540	1.019591
λ_{220}	2.675041	3.104563
λ_{201}	−0.198577	0.942209
λ_{202}	1.928798	2.978977
λ_{400}	3.787926	3.725691

Table 3. Percentage efficiency of optimum estimators of S_y^2.

For population No. 1							
Sample sizes		Efficiency of estimator	Efficiencies of optimum estimators belonging to the class				
n'	n	s_y^2	$\hat{V}_{gd}$	$\hat{b}_g$	$\hat{b}_h$	$\hat{V}_{Hd}$	$\hat{V}_{Td}$
15	3	100	285.2551	367.8336	369.1357	316.624	**399.9522**
15	5	100	217.9585	305.9726	306.7226	271.2869	**343.3513**
15	7	100	176.3537	261.9233	262.3626	245.2981	**300.7845**
15	10	100	137.0987	215.4068	215.5923	222.0343	**253.6207**
15	12	100	119.3829	192.6031	192.692	211.9055	**229.6176**
For population No. 2							
Sample sizes		Efficiency of estimator	Efficiencies of optimum estimators belonging to the class				
n'	n	s_y^2	$\hat{V}_{gd}$	$\hat{b}_g$	$\hat{b}_h$	$\hat{V}_{Hd}$	$\hat{V}_{Td}$
15	3	100	312.3561	620.0806	632.5651	594.8388	**639.7961**
15	5	100	230.7038	593.0345	610.0475	577.6099	**623.5314**
15	7	100	182.894	568.2492	580.6625	568.3307	**608.0731**
15	10	100	139.5229	534.7266	541.5353	560.2955	**586.2713**
15	12	100	120.4766	514.4923	518.254	556.8582	**572.5851**

populations are given in Table 2. Table 3 gives the efficiencies of the optimum estimators of the classes $\hat{V}_{gd}$, $\hat{b}_g$, $\hat{b}_h$, $\hat{V}_{Hd}$ and $\hat{V}_{Td}$ relative to the estimator s_y^2.

Table 3 shows that the proposed class of chain estimators $\hat{V}_{Td}$ is always more efficient than the others in every case for both the populations considered.

7. Conclusions

From the Sections 4 and 5, we see that it is very cumbersome to obtain the optimum estimator of the generalized class $\hat{V}_{Hd}$ as well as its optimum mean square error, whereas it is very simple in the case of the proposed class of chain estimators $\hat{V}_{Td}$. Also from Table 3, we see that for both the empirical populations, the optimum estimator of the proposed class $\hat{V}_{Td}$ is always more efficient than that of all the classes $\hat{V}_{gd}$, $\hat{b}_g$, $\hat{b}_h$ and $\hat{V}_{Hd}$. So we conclude that the use of chain estimators belonging to class $\hat{V}_{Td}$ should be preferred over the estimators belonging to the generalized class $\hat{V}_{Hd}$ when the variable z is highly correlated with variable x instead of variable y, whereas the correlation between the variables y and z exists due to only the high correlation between the variables y and x.

Acknowledgements

The authors thank the referees for their comments which improved the quality of this paper very much.

References

Ahmed, M. S., Raman, M. S. and Hossain, M. I. (2000). Some competitive estimators of finite population variance using multivariate auxiliary information, *Information and Management Sciences*, **11**(1), 49–54.

Ahmed, M. S., Walid, A. D. and Ahmed, A. O. H. (2003). Some estimators for finite population variance under two-phase sampling, *Statistics in Transition*, **6**(1), 143–150.

Al-Jararha, J. and Ahmed, M. S. (2002). The class of chain estimators for a finite population variance using double sampling, *Information and Management Sciences*, **13**(2), 13–18.

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

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

Liu, T. P. (1974). A general unbiased estimator for the variance of a finite population, *Sankhya C*, **36**(1), 23–32.

Murthy, M. N. (1967). *Sampling Theory and Methods*, Statistical Publishing Soc., Calcutta, India.

Prasad, B. and Singh, H. P. (1990). Some improved ratio type estimators of finite population variance in sample surveys, *Commun. Statist. - Theory and Methods*, **19**(3), 1127–1139.

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

Singh, H. P. and Singh, R. (2001). Improved ratio type estimator for variance using auxiliary information, *J. Ind. Soc. Agri. Statist.*, **54**(3), 276–287.

Singh, S. and Kataria, P. (1990). An estimator of finite population variance, *J. Ind. Soc. Agri. Statist.*, **42**(2), 186–188.

Srivastava, S. K. and Jhajj, H. S. (1980). A class of estimators using auxiliary information for estimating finite population variance, *Sankhya C*, **42**(1–2), 87–96.

Srivastava, S. K. and Jhajj, H. S. (1983). A class of estimators of the population mean using multi-auxiliary information, *Cal. Stat. Assoc. Bull.*, **32**, 47–56.

Srivastava, S. K. and Jhajj, H. S. (1987). Improved estimation in two phase and successive sampling, *J. Ind. Stat. Assoc.*, **25**, 71–75.

Srivastava, S. K. and Jhajj, H. S. (1995). Classes of estimators of finite population mean and variance using auxiliary infomation, *J. Ind. Soc. Agri. Statist.*, **47**(2), 119–128.