Revista Colombiana de Estadística
Diciembre 2012, volumen 35, no. 3, pp. 509 a 521

Some Alternative Predictive Estimators of Population Variance

Algunos estimadores predictivos alternativos de la varianza poblacional

Radhakanta Nayak[1,a], Lokanath Sahoo[2,b]

[1]Department of Statistics, Khallikote College, Berhampur, India
[2]Department of Statistics, Utkal University, Bhubaneswar, India

Abstract

Using a predictive estimation procedure, an attempt has been made to develop some estimators for the finite population variance in the presence of an auxiliary variable. Analytical and simulation studies have been undertaken for understanding the performance of the suggested estimators compared to some existing ones.

***Key words*:** Auxiliary variable, Bias, Efficiency, Prediction approach.

Resumen

Mediante el uso de un procedimiento de estimación predictivo, se desarrollan algunos estimadores de la varianza poblacional en la presencia de una variable auxiliar. Estudios analíticos y de simulación son implementados para entender el desempeño de los estimadores sugeridos en comparación con otros ya existentes.

***Palabras clave*:** variable auxiliar, sesgo, eficiencia, enfoque de predicción.

1. Introduction

Let $U = \{1, 2, \ldots, i \ldots, N\}$ be a finite population, and y and x denote the study variable and the auxiliary variable taking values y_i and x_i respectively on the ith unit ($i = 1, 2, \ldots, N$). Let $\overline{Y} = \sum_{i=1}^{N} y_i/N$ and $\overline{X} = \sum_{i=1}^{N} x_i/N$ be the population means, $S_y^2 = \sum_{i=1}^{N}(y_i - \overline{Y})^2/(N-1)$ and $S_x^2 = \sum_{i=1}^{N}(x_i - \overline{X})^2/(N-1)$ be the population variances of y and x respectively. Assume that a sample s of n units is drawn from U according to simple random sampling without replacement

[a]Lecturer. E-mail: rkn2010@gmail.com
[b]Professor. E-mail: lnsahoostatuu@rediffmail.com

(SRSWOR) in order to estimate the unknown parameter S_y^2. Let $\overline{y} = \sum_{i\in s} y_i/n$ and $\overline{x} = \sum_{i\in s} x_i/n$ be the sample means, $s_y^2 = \sum_{i\in s}(y_i - \overline{y})^2/(n-1)$ and $s_x^2 = \sum_{i\in s}(x_i - \overline{x})^2/(n-1)$ the sample variances.

In certain situations, estimation of population variance S_y^2 has received considerable attention from survey statisticians. For example, in manufacturing industries and pharmaceutical laboratories, sometimes the researchers are interested in the variation of their products. Although, the literature describes a great variety of techniques for using auxiliary information by means of ratio, and product and regression methods for estimating population mean, variance estimation using auxiliary information has received scarce attention. This is perhaps due to the belief that the gain in efficiency we could obtain by involving an auxiliary variable may not be too much relevant to motivate the use of more complex estimators. However, some efforts in this direction are due to Das & Tripathi (1978), Isaki (1983), Prasad & Singh (1990)(1992), Singh & Kataria (1990), Srivastava & Jhajj (1980)(1995), Singh & Singh (2001), Ahmed, Walid & Ahmed (2003), Giiancarlo & Chiara (2004), Jhajj, Sharma & Grover (2005), Kadilar & Cingi (2006)(2007) and Grover (2007). Two notable estimators that are very much popular in the literature are due to Isaki (1983) defined by

$$\nu_1 = s_y^2 S_x^2 / s_x^2$$

and

$$\nu_2 = s_y^2 + b^*(S_x^2 - s_x^2)$$

where b^* is an estimate of the regression coefficient of s_y^2 on s_x^2 defined by $b^* = \frac{s_y^2(\widehat{\lambda}-1)}{s_x^2(\widehat{\beta}_2(x)-1)}$, such that $\widehat{\lambda} = m_{22}/m_{20}m_{02}$ and $\widehat{\beta}_2(x) = m_{40}/m_{20}^2$ with $m_{rs} = \sum_{i\in s}(x_i - \overline{x})^r (y_i - \overline{y})^s/n$ [cf., Garcia & Cebrain (1996), and Kadilar & Cingi (2006)].

During the years that followed, much emphasis has been given on the prediction of population mean or total [cf., Srivastava (1983)]. But, little interest has been shown towards the prediction of the population variance. Under this approach, the survey data at hand i.e., the sample observations are treated as fixed and unassailable. Uncertainty is then attached only to the unobserved values which need to be predicted. Bolfarine & Zacks (1992) indicated various techniques for predicting population variance. Biradar & Singh (1998), using classical estimation theory, provided some predictive estimators for S_y^2. In this paper, using auxiliary variable x, we develop some more estimators under the prediction approach of Basu (1971) with regards to a finite population setup.

2. Prediction Criterion

Let us decompose U into two mutually exclusive domains s and r of n and $N-n$ units respectively, where $r = U - s$ denotes the collection of units in U which are not included in s. Then, under the usual prediction criterion given in Bolfarine & Zacks (1992), it is possible to express

$$(N-1)S_y^2 = (n-1)s_y^2 + (N-n-1)S_{y(r)}^2 + (1-f)n(\overline{y} - \overline{Y}_r)^2, \tag{1}$$

where $f = n/N$, and $\overline{Y}_r = \sum_{i\in s} y_i/(N-n)$ and $S^2_{y(r)} = \sum_{i\in r}(y_i-\overline{Y}_r)^2/(N-n-1)$ are respectively the mean and variance of y-values belonging to r.

Notice that the first component on the right hand side of (1) is known while the second and third components are unknown. Hence, the prediction of $(N-1)S^2_y$ is possible when $S^2_{y(r)}$ and $\overline{Y}_r$ are simultaneously predicted by some means from the sample data. Using V_r and M_r as their respective predictors, a predictor of S^2_y can be provided by the equation:

$$(N-1)\widehat{S}^2_y = (n-1)s^2_y + (N-n-1)V_r + (1-f)n(\overline{y} - M_r)^2 \tag{2}$$

Most of the predictions are based either on distributional forms or an assumed model [cf., Royall (1988), Bolfarine & Zacks (1992)]. However, Sampford (1978) argued that the consideration of a model free prediction can generate a new, estimator possessing some desirable properties. Basu (1971) also encouraged the use of tools of the classical estimation theory to find out suitable predictors for $\overline{Y}$. Biradar & Singh (1998) formulated some estimators of S^2_y from (2) by considering suitable choices of the predictors V_r and M_r in terms of the auxiliary variable x under the tools of classical estimation theory. Defining $\overline{X}_r = \sum_{i\in r} x_i/(N-n)$ and $S^2_{x(r)} = \sum_{i\in r}(x_i - \overline{X}_r)^2/(N-n-1)$, we report below their estimators along with the corresponding selections of V_r and M_r:

$$\nu_3 = \left(\frac{N-2}{N-1}\right) s^2_y$$

when $V_r = s^2_y$ and $M_r = \overline{y}$,

$$\nu_4 = \frac{s^2_y}{s^2_x} S^2_x + \frac{nN(\overline{x} - \overline{X})^2}{(N-n)(N-1)} \left(\frac{\overline{y}^2}{\overline{x}^2} - \frac{s^2_y}{s^2_x}\right)$$

when $V_r = s^2_y S^2_{x(r)}/s^2_x$ and $M_r = \overline{y}\overline{X}_r/\overline{x}$, and

$$\nu_5 = \frac{s^2_y}{s^2_x} S^2_x + \frac{nN(\overline{x} - \overline{X})^2}{(N-n)(N-1)} \left(b^2_{yx} - \frac{s^2_y}{s^2_x}\right)$$

when $V_r = s^2_y S^2_{x(r)}/s^2_x$ and $M_r = \overline{y} + b_{yx}(\overline{X}_r - \overline{x})$, where $b_{yx} = s_{yx}/s^2_x$.

Biradar & Singh (1998) also identified Isaki's (1983) estimator ν_1 as a special case of (2) for $V_r = s^2_y S^2_{x(r)}/s^2_x$ and $M_r = \overline{y} + s_y(\overline{X}_r - \overline{x})/s_x$. This shows that the estimator possesses a predictive character.

3. Some New Predictive Estimators of S^2_y

In the following discussions, we introduce some alternative approaches in order to develop a few more predictive estimators of S^2_y.

1. Consider the following alternative but equivalent representation of S_y^2:

$$(N-1)S_y^2 = (n-1)s_y^2 + (N-n)[\sigma^2_{y(r)} + f(\overline{y} - \overline{Y}_r)^2] \tag{3}$$

where $\sigma^2_{y(r)} = \sum_{i \in r}(y_i - \overline{Y}_r)^2/(N-n)$. Denoting V_r^* as a predictor of $\sigma^2_{y(r)}$ and M_r, as the predictor of $\overline{Y}_r$, the following alternative predictive equation can be considered:

$$(N-1)S_y^2 = (n-1)s_y^2 + (N-n)[V_r^* + f(\overline{y} - M_r)^2] \tag{4}$$

Then, for $V_r^* = \left(\frac{n-1}{n}\right) s_y^2$ and $M_r = \overline{y}$ in (4) we get an estimator of S_y^2 defined by

$$\nu_6 = \left(\frac{n-1}{n}\right)\left(\frac{N}{N-1}\right) s_y^2$$

2. Biradar & Singh (1998) developed the estimator ν_5 from (2) with $V_r = s_y^2 S^2_{x(r)}/s_x^2$ and $M_r = \overline{y} + b_{yx}(\overline{X}_r - \overline{x})$. See that in such an attempt V_r has been assumed a ratio version of the variance estimator while the connected mean estimator is a regression estimator. Hence as a matter of curiosity, we may also think in the light of Isaki (1983) to use a regression version of the variance estimator i.e., $V_r = s_y^2 + b^*(S^2_{x(r)} - s_x^2)$ along with the mean estimator $M_r = \overline{y} + b_{yx}(\overline{X}_r - \overline{x})$ in the predictive equation (2) to predict S_y^2. This operation, after a considerable simplification, leads to produce the following estimator:

$$\nu_7 = \frac{N-2}{N-1}\left[s_y^2 + b^*\left(\frac{N-1}{N-2}S_x^2 - s_x^2\right)\right]$$

3. Srivastava (1983) considered the predictive equation:

$$\hat{\overline{Y}} = f\overline{y} + (1-f)M_r \tag{5}$$

where M_r is the implied predictor of $\overline{Y}_r$, for predicting $\overline{Y}$ and shown that when $M_r = \overline{y}\,\overline{X}_r/\overline{x}, \hat{\overline{Y}} = \overline{y}_R = \overline{y}\,\overline{X}/\overline{x}$, the classical ratio estimator of $\overline{Y}$, and when $M_r = \overline{y} + b_{yx}(\overline{X}_r - \overline{x}), \hat{\overline{Y}} = \overline{y}_L = \overline{y} + b_{yx}(\overline{X} - \overline{x})$, the classical regression estimator of $\overline{Y}$. Thus, both the ratio and regression estimators ($\overline{y}_R$ and $\overline{y}_L$) of the mean possess a predictive character, the origin of which actually lies in predicting y_i's, $i \in r$, by $y_i = \overline{y}x_i/\overline{x}$ and $y_i = \overline{y} + b_{yx}(x_i - \overline{x})$ in that order. In view of this, we designate these two estimators as basic estimators of the population mean. Notice that the predictive estimators ν_1, ν_4, ν_5 and ν_7 suggested so far have been obtained by using either $V_r = s_y^2 S^2_{x(r)}/s_x^2$ or $V_r = s_y^2 + b^*(S^2_{x(r)} - s_x^2)$ as the case may be. This means that the unknown quantity $S^2_{y(r)}$ is estimated as a whole with the same principle as that applied to estimate $\overline{Y}_r$. But, such a choice of V_r seems to be arbitrary by nature. Rather, we feel that it is more appropriate if the variance is established by predicting individual y_i's, $i \in r$, for which we need to express S_y^2 in the following form:

$$(N-1)S_y^2 = (n-1)s_y^2 + \sum_{i \in r} y_i^2 - (N-n)\overline{Y}_r^2 + (1-f)n(\overline{y} - \overline{Y}_r)^2 \qquad (6)$$

A number of new estimators can be easily generated from this equation on the basis how $\sum_{i \in r} y_i^2$ is predicted. But, for simplicity, here we consider the prediction of $y_i, i \in r$, either by $y_i = \overline{y}x_i/\overline{x} = \overline{y} + \overline{y}(x_i - \overline{x})/\overline{x}$ or by $y_i = \overline{y} + b_{yx}(x_i - \overline{x})$ and prediction of $\overline{Y}_r$ by $\overline{y}\overline{X}_r/\overline{x}$.

Then, accordingly after a considerable simplification, we obtain the following two new estimators:

$$\nu_8 = \left(\frac{n-1}{N-1}\right)\left[s_y^2 + \left(\frac{\overline{y}}{\overline{x}}\right)^2 \left(\frac{N-1}{n-1}S_x^2 - s_x^2\right)\right]$$

$$\nu_9 = \left(\frac{n-1}{N-1}\right)\left[s_y^2 + b_{yx}^2 \left(\frac{N-1}{n-1}S_x^2 - s_x^2\right)\right]$$

4. Performance of the Proposed Estimators

Out of the nine estimators considered or proposed in the preceding sections, the estimators ν_3 and ν_6 were achieved without using any auxiliary information whereas others were achieved through the use of information on the auxiliary variable x. A desirable goal here is to study the performance of the proposed estimators ν_6 to ν_9 compared to ν_1 to ν_5 at least in respect of bias and mean square error (MSE) i.e., efficiency, where bias and MSE of an estimator ν_i of S_y^2 are defined respectively by $B(\nu_i) = E(\nu_i) - S_y^2$ and $M(\nu_i) = E(\nu_i - S_y^2)^2 (i = 1, 2, \ldots, 9)$. But, we see that some of the estimators are so complex that it is not possible to derive exact expressions for their bias and MSE. Biradar & Singh (1998) presented asymptotic expressions for these performance measures for the estimators ν_1 to ν_5. On the other hand, Nayak (2009) derived these expressions in favor of ν_1 to ν_9. But, the sufficient conditions for superiority of one estimator over other derived by the authors using asymptotic expressions are so complicated that it is not conducive to compare different estimators meaningfully. However, to facilitate our comparison, these expressions are considered under the following widely used linear regression model:

$$y_i = \boldsymbol{\beta} x_i + e_i,\ i = 1, 2, \ldots, N \qquad (7)$$

where $\boldsymbol{\beta}(> 0)$ is the model parameter and e_i is the error component such that $E(e_i/x_i) = 0, E(e_i^2/x_i) = \delta x^g (\delta > 0, 0 \le g \le 1)$, and $E(e_i e_j/x_i, x_j) = 0$ for $i \ne j$. Further, we also assume that $E(e_i^4/x_i) = \xi x^g$ and $E(e_i^3/x_i) = E(e_i^3 e_j/x_i, x_j) = E(e_i e_j^3/x_i, x_j) = 0, (i \ne j)$. It may be pointed out here that the asymptotic expressions for bias and MSE of different estimators under this assumed model are derived through the Taylor linearization method.

4.1. Comparison of Bias

After some algebraic manipulations (suppressed to save space), we get the following model-based results in respect of the bias of different estimators up to $O(n^{-1})$

$$B(\nu_1) = \mathcal{C}\delta E(x^g) \tag{8}$$
$$B(\nu_2) = 0 \tag{9}$$
$$B(\nu_3) = -\frac{1}{N-1}[\beta^2 S_x^2 + \delta E(x^g)] \tag{10}$$
$$B(\nu_4) = -(\mathcal{B} - \mathcal{C})\delta E(x^g) \tag{11}$$
$$B(\nu_5) = -(\mathcal{K} - \mathcal{C})\delta E(x^g) \tag{12}$$
$$B(\nu_6) = -\frac{N-n}{N-1}[\beta^2 S_x^2 + \delta E(x^g)] \tag{13}$$
$$B(\nu_7) = -\frac{1}{N-1}\left(\frac{n-2}{n-1}\right)\delta E(x^g) \tag{14}$$
$$B(\nu_8) = -(N-n)\mathcal{B}\delta E(x^g) \tag{15}$$
$$B(\nu_9) = -\left(\frac{N-n}{N-1}\right)\left(\frac{n-2}{n-1}\right)\delta E(x^g) \tag{16}$$

where $\mathcal{B} = \frac{1}{N-1}\left(1 - \frac{C_x^2}{n}\right)$, $\mathcal{C} = \frac{1}{n}(\beta_2(x) - 2)$ and $\mathcal{K} = \left(\frac{n}{n-1}\right)\left(\frac{1}{N-1}\right)$, such that C_x and $\beta_{2(x)}$ are respectively the coefficient of variation and β_2- coefficient of the auxiliary variable x.

In the light of the expressions (8) to (16), we state the following comments on the bias of the estimators:

(i) The regression estimator ν_2 is model-unbiased, ν_1 is positively biased and the rest seven estimators are negatively biased.

(ii) $|B(\nu_3)| < |B(\nu_6)|$. This indicates that the bias of ν_6 is always greater than that of ν_3.

(iii) $|B(\nu_8)| < |B(\nu_7)|$i.e., ν_8 is less biased than ν_7.

(iv) $|B(\nu_7)| < |B(\nu_9)|$i.e., ν_7 is less biased than ν_9.

(v) $|B(\nu_9)| \lesseqgtr |B(\nu_8)|$ according as $C_x^2 \lesseqgtr \frac{n}{n-1}$.

(vi) $|B(\nu_4)| < |B(\nu_7)|$, when $|\mathcal{B} - \mathcal{C}| < \frac{1}{N-1}\left(\frac{n-2}{n-1}\right)$.

(vii) $|B(\nu_5)| < |B(\nu_7)|$, when $|\mathcal{K} - \mathcal{C}| < \frac{1}{N-1}\left(\frac{n-2}{n-1}\right)$.

(viii) $|B(\nu_7)| < |B(\nu_1)|$, when $\mathcal{C} > \mathcal{K}$ and $n > 2$.

In view of (iii) and (iv), although we can conclude that ν_8 is less biased than ν_7 and ν_9, we fail to obtain a clear-cut idea on the magnitude of bias of ν_8 compared to ν_1,ν_4 and ν_5. Because, comparison of (15) with (8) or (11) or (12) does not lead to any meaningful conditions.

4.2. Comparison of Efficiency

We present below model-based asymptotic expressions of the MSEs of different estimators up to $O(n^{-1})$ together with the exact expression for the variance of the traditional unbiased estimator s_y^2.

$$V(s_y^2) = V(\nu_2) + \mathcal{C}\beta^4 S_x^4 \tag{17}$$

$$M(\nu_1) = M(\nu_2) + \mathcal{C}\delta^2 E^2(x^g) \tag{18}$$

$$M(\nu_2) = \xi(x^g) + 4\boldsymbol{\beta}^2 S_x^2 \frac{\delta E(x^g)}{n-1} - \frac{n-3}{n(n-1)}\delta^2 E^2(x^g) \tag{19}$$

$$M(\nu_3) = \left(\frac{N-2}{N-1}\right)^2 V(s_y^2) \cong V(s_y^2) \tag{20}$$

$$M(\nu_4) = M(\nu_2) + \mathcal{C}\delta^2 E^2(x^g) + \frac{2}{N-1}\delta^2 E^2(x^g) \tag{21}$$

$$M(\nu_5) = M(\nu_2) + \mathcal{C}\delta^2 E^2(x^g) + \frac{2}{N-1}\left(\frac{n}{n-1}\right)\delta^2 E^2(x^g) \tag{22}$$

$$M(\nu_6) = \left(\frac{n-1}{n}\right)^2 \left(\frac{N}{N-1}\right)^2 V(s_y^2) \cong \left\{1 - 2\left(\frac{1}{n} + \frac{1}{N-1}\right)\right\} V(s_y^2) \tag{23}$$

$$M(\nu_7) = M(\nu_2) \tag{24}$$

$$M(\nu_8) = M(\nu_2) - 4\left(\frac{N-n}{N-1}\right)^2 \boldsymbol{\beta}^2 S_x^2 \frac{\delta E(x^g)}{n-1} + 2\left(\frac{N-n}{N-1}\right)^2 \frac{C_x^2}{n}(2\boldsymbol{\beta}^2 S_x^2 - 1)\delta E(x^g) \tag{25}$$

$$M(\nu_9) = M(\nu_2) + 2\left(1 - 2\frac{N-n}{N-1}\right)\delta^2 \frac{E^2(x^g)}{n-1} + \left(\frac{N-n}{N-1}\right)^2 \delta^2 E^2(x^g). \tag{26}$$

From these expressions, as ν_2 appears to be more efficient than $s_y^2, \nu_1, \nu_3, \nu_4$ and ν_5, we present the following results concerning efficiencies of the suggested estimators:

(ix) $M(\nu_6) < M(\nu_3) < V(s_y^2)$. This indicates that ν_6 is more efficient than both s_y^2 and ν_3.

(x) $M(\nu_7) = M(\nu_2)$ i.e., ν_7 and ν_2 are equally efficient even though they are configurationally different.

(xi) ν_8 is more efficient than ν_2 when $\boldsymbol{\beta}^2 S_x^2 < \frac{1}{2}$ which is very often satisfied in practice. This means that there is a scope to improve upon the Isaki's regression estimator ν_2 through ν_8.

(xii) The estimator ν_9 is less efficient than ν_2 when $n < \dfrac{N+1}{2}$.

(xiii) $M(\nu_8) < M(\nu_2) = M(\nu_7) < M(\nu_9)$, when $n < \dfrac{N+1}{2}$ and $\boldsymbol{\beta}^2 S_x^2 < \frac{1}{2}$. This shows that ν_8 is preferred to ν_2, ν_7 and ν_9 when the stated conditions are satisfied. The first condition is not a serious one. The second condition is easily satisfied for characters being measured in smaller magnitudes. We can also reduce the mean square error by considering transformations on the auxiliary variable and making the second condition more feasible.

4.3. Some Remarks

From the previous model-based comparisons, we see that the proposed estimator ν_8 turns out to be more efficient than others. But no meaningful conclusion could be drawn in favor of the four proposed estimators $\nu_i, i = 6, 7, 8, 9$ in respect of bias. This negative finding may be discouraging but not very decisive as our comparisons are based on the asymptotic expressions derived through Taylor linearization. However, as a counterpart to these analytical comparisons, we do carry out a simulation study in the next section with an objective to examine the overall performance of the different variance estimators. The performance measures of an estimator ν_i taken into consideration in this study are (i) *Absolute Relative Bias* $(ARB) = |B(\nu_i)|/S_y^2$, and (ii) *Percentage Relative Efficiency* $(PRE) = 100 \times V(s_y^2)/M(\nu_i), (i = 1, 2, \ldots, 9)$

5. Description of the Simulation Study

Our simulation study involves repeated draws of simple random (without replacement) samples from 20 natural populations described in Table 1. 2,000 independent samples, for $n = 6, 8$ and 10, were selected from a population and for each sample several estimators were calculated. Then, considering 2,000 such combinations, simulated values of the performance measures were calculated and displayed in Tables 2 and 3. To save space, the numerical values of the performance measures for $n = 8$ and 10 are not shown, but the results based on these values are only reported. Major findings of the study are discussed in subsections 5.1 and 5.2.

5.1. Results Based on the ARB

The numerical values on the ARB reveal that there is no definite pattern in the performances of different estimators. The estimator ν_1 possesses the least ARB in 7 populations for $n = 6$ and in 6 populations for $n = 8$ and 10. ν_8 is found to have least ARB in 8, 10 and 11 populations for $n = 6, 8$ and 10 respectively. This clearly indicates that the overall performance of ν_8 improves with the increase in sample size. Searching for an estimator as the third choice is difficult owing to very erratic results in favor of the estimators (except ν_1 and ν_8).

Table 1: Description of the populations.

Pop	Source	N	y	x
1	Cochran (1977) p. 152	49	no of inhabitants in 1930	no. of inhabitants in 1920
2	Sukhatme & Sukhatme (1977) p. 185	34	area under wheat in 1937	area under wheat in1936
3	Sukhatme & Sukhatme (1977) p. 185	34	area under wheat in 1937	area under wheat in1931
4	Sampford (1962) p. 61	35	acreage under oats in 1957	acreage of crops and grass in 1947
5	Wetherill (1981) p. 104	32	yield of petroleum sprit	petroleum fraction end point
6	Murthy (1967) p. 398	43	no of absentees	no of workers
7	Murthy (1967) p. 399	34	area under wheat in 1964	cultivated area in 1961
8	Murthy (1967) p. 399	34	area under wheat in 1964	area under wheat in 1963
9	Steel & Torrie (1960) p. 282	30	leaf burn in secs.	percentage of potassium
10	Shukla (1966)	50	fiber yield	height of plant
11	Shukla (1966)	50	fiber yield	base diameter
12	Murthy (1967) p. 178	108	area under winter paddy	geographical area
13	Dobson (1990) p. 83	30	cholesterol	age in years
14	Dobson (1990) p. 83	30	cholesterol	body mass
15	Yates (1960) p. 159	25	measured volume of timber	eye estimated volume of timber
16	Yates (1960) p. 159	43	no. of absentees	total no. of persons
17	Panse & Sukhatme (1985) p. 118	25	progeny mean	parental plant value
18	Panse & Sukhatme (1985) p. 118	25	progeny mean	parental plot mean
19	Dobson (1990) p. 69	20	total calories from carbohydrate	calories as protein
20	Horvitz & Thompson (1952)	20	actual no. of households	eye estimated number of households

5.2. Results Based on the PRE

Results on the PRE of the competing estimators show that the estimator ν_8 is decidedly more efficient than the rest of the estimators in all populations for $n = 6$ and in 18 populations (except populations 1 and 17) for $n = 8$ and 10. Also the efficiency gain due to this estimator is noticeably high. The estimator ν_9 is found to be the second best estimator being more efficient than others (except ν_8) in 12 populations for $n = 6$ and in 10 populations for $n = 8$ and 10.

Further, it is observed that both ν_3 and ν_6 i.e., the estimators exploiting no auxiliary information, perform satisfactorily with ν_6 being better than ν_3 in all populations. It may also be noted here that for $n = 6, \nu_8$ is the only estimator using auxiliary variable x that is better than s_y^2 in all populations. However, this situation slightly changes with the increase in the sample size as it is worse than s_y^2 in one population for $n = 8$ and in two populations for $n = 10$. The estimators ν_1 ,ν_2, ν_4 and ν_5 do not fare well in most of the cases.

Table 2: ARB of the estimators for $n = 6$.

Pop No	ν_1	ν_2	ν_3	ν_4	ν_5	ν_6	ν_7	ν_8	ν_9
1	10.24	12.23	4.10	10.24	10.29	9.55	12.27	1.85	13.75
2	18.81	18.45	8.20	18.63	18.69	18.72	18.33	5.68	15.13
3	1.19	4.53	8.20	2.03	1.39	18.72	4.84	36.58	13.16
4	57.23	18.51	39.30	49.81	50.48	46.35	18.35	13.50	13.99
5	24.85	26.94	26.39	32.36	27.66	34.57	27.62	79.72	44.42
6	31.04	45.37	44.37	41.58	41.73	51.83	45.88	33.46	64.22
7	0.57	4.13	7.01	1.57	0.36	17.76	4.46	39.28	13.35
8	1.19	0.56	7.01	0.92	1.14	17.76	1.49	0.35	1.47
9	32.96	13.67	22.47	24.87	30.23	30.78	16.04	81.23	70.68
10	19.36	24.10	35.40	20.17	19.76	43.93	24.67	78.83	49.51
11	62.42	3.47	35.40	57.74	58.13	43.93	4.08	73.58	30.10
12	25.10	11.15	51.06	23.71	22.81	58.44	11.19	8.64	15.69
13	61.77	14.72	35.31	62.93	60.25	42.24	13.68	7.95	10.23
14	27.91	34.55	35.31	28.71	28.75	42.24	36.14	72.76	72.55
15	7.04	3.13	3.08	3.73	10.98	12.21	4.61	2.02	31.25
16	43.05	46.28	44.62	44.10	54.08	51.59	46.77	67.22	63.75
17	33.62	29.05	19.07	25.71	36.39	26.70	30.55	46.47	57.58
18	40.92	18.98	19.07	21.61	21.92	26.70	11.23	8.13	51.70
19	33.30	5.06	25.42	24.22	27.79	30.95	2.80	4.32	26.57
20	0.74	2.34	16.31	1.27	1.34	22.51	2.97	15.91	11.19

6. Conclusions

Our model-assisted analytical and simulated studies lead to an overall conclusion that the estimator ν_8 is preferable to others on the ground of efficiency. Although the analytical comparison fails to conclude which estimator is decidedly better than others on the ground of bias, the simulation study gives an indication that on this ground ν_8 is the better performer than other estimators. In view of these findings, if computational difficulty is not a matter of great concern, the variance estimator ν_8 may be considered as the most suitable estimator. Of course, these findings are only indicative and are no able to reveal essential features of the comparable estimators in a straightforward manner. Further investigations in this direction may be made for arriving at the conclusions.

Table 3: PRE of the estimators for $n = 6$.

Pop No	ν_1	ν_2	ν_3	ν_4	ν_5	ν_6	ν_7	ν_8	ν_9
1	285	223	104	226	283	138	228	395	246
2	419	444	106	427	416	135	444	531	403
3	270	261	106	281	271	135	262	425	301
4	21	68	106	29	28	135	68	313	68
5	122	127	106	154	129	135	129	646	176
6	191	197	104	196	195	137	201	816	368
7	410	370	106	437	414	135	372	804	440
8	958	908	106	985	960	136	911	1037	989
9	16	78	107	17	17	135	82	206	202
10	61	83	104	63	67	138	84	400	194
11	11	45	104	12	12	138	45	663	45
12	15	65	108	17	16	141	65	398	66
13	13	19	107	14	13	135	19	475	206
14	72	104	107	74	73	135	109	665	435
15	146	146	109	153	152	133	149	196	139
16	211	211	105	218	215	138	215	478	393
17	208	169	109	226	239	133	179	615	406
18	6	52	108	11	10	133	54	190	70
19	9	27	111	12	11	130	27	841	23
20	121	124	111	130	122	131	126	817	155

Acknowledgement

The authors are grateful to the referees and the Editor-in-Chief for providing some useful comments on an earlier draft of the paper.

[Recibido: diciembre de 2011 — Aceptado: noviembre de 2012]

References

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

Basu, D. (1971), An essay on the logical foundations of survey sampling, *in* V. P. Godambe & D. A. Sprott, eds, 'Foundations of Statistical Inference', Vol. I, Holt, Rinehart and Wintson, Toronto, Canada, pp. 203–242.

Biradar, R. S. & Singh, H. P. (1998), 'Predictive estimation of finite population variance', *Calcutta Statistical Association Bulletin* **48**, 229–235.

Bolfarine, H. & Zacks, S. (1992), *Prediction Theory for Finite Populations*, Springer-Verlag.

Cochran, W. (1977), *Sampling Techniques*, Wiley Eastern Limited.

Das, A. & Tripathi, T. P. (1978), 'Use of auxiliary information in estimating the finite population variance', *Sankhyā* **C40**, 139–148.

Dobson, A. J. (1990), *An Introduction to Generalized Linear Models*, Chapman and Hall, New York.

Garcia, M. R. & Cebrain, A. A. (1996), 'Repeated substitution method: The ratio estimator for the population variance', *Metrika* **43**, 101–105.

Giiancarlo, D. & Chiara, T. (2004), 'Estimation for finite population variance in double sampling', *Metron* **62**, 223–232.

Grover, L. K. (2007), 'A wide and efficient class of estimators of population variance under sub-sampling scheme', *Model Assisted Statistics and Applications* **2**, 41–51.

Horvitz, D. G. & Thompson, D. J. (1952), 'A generalization of sampling without replacement from a finite universe', *Journal of the American Statistical Association* **47**, 663–685.

Isaki, C. T. (1983), 'Variance estimation using auxiliary information', *Journal of the American Statistical Association* **78**(117-123).

Jhajj, H. S., Sharma, M. K. & Grover, L. K. (2005), 'An efficient class of chain estimators of population variance under sub-sampling scheme', *Journal of the Japan Statistical Society* **35**, 273–286.

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

Kadilar, C. & Cingi, H. (2007), 'Improvement in variance estimation in simple random sampling', *Communications in Statistics-Theory and Methods* **36**, 2075–2081.

Murthy, M. N. (1967), *Sampling Theory and Methods*, Statistical Publishing Society, Kolkata.

Nayak, R. K. (2009), Some Estimation Strategies in Finite Population Survey Sampling Using Auxiliary Information, PhD., Utkal University, India.

Panse, V. G. & Sukhatme, P. V. (1985), *Statistical Methods for Agricultural Workers*, Indian Council for Agricultural Research, New Delhi.

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

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

Royall, R. M. (1988), The prediction approach to sampling theory, *in* P. R. Krishnaish & C. R. Rao, eds, 'Handbook of Statistics', Vol. 6, North Holland, pp. 351–358.

Sampford, M. R. (1962), *An Introduction to Sampling Theory*, Oliver and Boyd, Edinburg.

Sampford, M. R. (1978), Predictive estimation and internal congruency, *in* H. A. David, ed., 'Contribution to Survey Sampling and Applied Statistics', Academic Press, New York, pp. 29–39.

Shukla, G. K. (1966), 'An alternative multivariate ratio estimate for finite population', *Calcutta Statistical Association Bulletin* **15**, 127–134.

Singh, H. P. & Singh, R. (2001), 'Improved ratio-type estimators for variance using auxiliary information', *Journal of the Indian Society of Agricultural Statistics* **54**, 276–287.

Singh, S. & Kataria, P. (1990), 'An estimator of the finite population variance', *Journal of the Indian Society of Agricultural Statistics* **42**, 186–188.

Srivastava, S. K. (1983), 'Predictive estimation of finite population using product estimators', *Metrika* **30**, 93–99.

Srivastava, S. K. & Jhajj, H. S. (1980), 'A class of estimators using auxiliary information for estimating finite population variance', *Sankhyā* **C42**, 87–96.

Srivastava, S. K. & Jhajj, H. S. (1995), 'Classes of estimators of finite population mean and variance using auxiliary information', *Journal of the Indian Society of Agricultural Statistics* **47**, 119–128.

Steel, R. G. D. & Torrie, J. H. (1960), *Principles and Procedures of Statistics*, Mc. Graw Hill Book Company.

Sukhatme, P. V. & Sukhatme, B. V. (1977), *Sampling Theory of Surveys with Applications*, Asia Publishing House, New Delhi.

Wetherill, G. B. (1981), *Intermediate Statistical Methods*, Chapman and Hall, London.

Yates, F. (1960), *Sampling Methods for Censuses and Surveys*, Charls and Griffin, London.