Nonlinear Analysis: Modelling and Control, 2010, Vol. 15, No. 2, 233–253

A family of estimators of population mean using multi-auxiliary variate and post-stratification

Gajendra K. Vishwakarma[1], **Housila P. Singh**[1], **Sarjinder Singh**[2]

[1]School of Studies in Statistics, Vikram University
Ujjain-456010, M.P., India
vishwagk@rediffmail.com; hpsujn@rediffmail.com

[2]Department of Mathematics, Texas A&M University-Kingsville
TX 78363, USA
sarjinder@yahoo.com

Received: 2009-09-15 **Revised:** 2010-04-07 **Published online:** 2010-06-01

Abstract. This paper suggests a family of estimators of population mean using multi-auxiliary variate based on post-stratified sampling and its properties are studied under large sample approximation. Asymptotically optimum estimator in the class is identified alongwith its approximate variance formulae. The proposed class of estimators is also compared with corresponding unstratified class of estimators based on estimated optimum value. At the end, an empirical study has been carried out to support the proposed methodology.

Keywords: multi-auxiliary variate, study variate, bias, variance, post-stratified sampling.

1 Introduction

Stratification is one of the most widely used techniques in sample survey design serving the dual purpose of providing samples that are representative of major sub-groups of the population and improving the precision of estimators [1]. Stratified sampling presupposes the knowledge of strata size as well as the availability of a frame for drawing a sample in each stratum [2]. However application of this technique presupposes the knowledge of strata size and the availability of sampling frames within strata. In many socio-economic and agricultural surveys where it is necessary to partition the finite population under consideration, due to its heterogeneity, into different sub-populations (strata), the sampling frame within strata may not be available. However frame for entire population may be available and percentage of population units falling into different strata may be known. Under such circumstances usual stratified sampling can not be used and thus an effort is made to get over the problem through post-stratification which consists in selecting a sample from the whole population by the procedure of simple random sampling without replacement followed by the classification of the selected sample units by strata and then treating it as if it were stratified sample, for instance, see [1, 3–11].

It is further noted that in sample surveys, the information on an auxiliary variate correlated with the principal (study) variate under study is either readily available or may be made available by diverting a part of the survey resources. This information may be utilized to increase the precision of estimators of population mean $\overline{Y}$ of the study variate y. Such an information is the known population mean $\overline{X}$ of the auxiliary variate x. For illustration, the average farm size in a local government area or district may be known while the problem is to estimate the average area under a particular crop per farm. The strata may be formed according to farm size, the percentage of farms falling into different size groups may be known but the identity of farms within a size group may not be known, see [12].

We assume that the population comprises N units, which can be uniquely partitioned into L strata of size $N_1, N_2, \ldots, N_L$ such that $\sum_{h=1}^{L} N_h = N$. The strata weights $W_h = N_h/N$ $(h = 1, 2, \ldots, L)$ are assumed known. Let (y_{hi}, x_{hi}) $(i = 1, 2, \ldots, N_h)$ denote the values of variates (y, x) respectively for i-th unit in h-th stratum and $\overline{Y}_h$ and $\overline{X}_h$ denote strata means. A simple random sample of size n is drawn without replacement from the population which results into the configuration $\underline{n} = (n_1, n_2, \ldots, n_L)$, n_h denoting the number of units in the sample falling in stratum h, $\sum_{h=1}^{L} n_h = n$. Assume that n is large enough so that the probability of n_h being zero is small (i.e. $Pr(n_h = 0) = 0$). Based on the foregoing procedure which is known as post-stratification, the usual unbiased post-stratified estimators for population means $\overline{Y} = \sum_{h=1}^{L} W_h \overline{Y}_h$ and $\overline{X} = \sum_{h=1}^{L} W_h \overline{X}_h$ of the study variate y and the auxiliary variate x are $\overline{y}_{PS} = \sum_{h=1}^{L} W_h \overline{y}_h$ and $\overline{x}_{PS} = \sum_{h=1}^{L} W_h \overline{x}_h$, where $\overline{y}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} y_{hi}$ and $\overline{x}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} x_{hi}$ are the means of the n_h sample units that fall into the h-th stratum whose size N_h is assumed to be known.

For given configuration of sample $\underline{n} = (n_1, n_2, \ldots, n_L)$ we have

$$\mathrm{Var}(\overline{y}_{PS}|\underline{n}) = E\big((\overline{y}_{PS} - \overline{Y})^2|\underline{n}\big) = \sum_{h=1}^{L} W_h^2 \left(\frac{1-f_h}{n_h}\right) S_{hy}^2,$$

$$\mathrm{Var}(\overline{x}_{PS}|\underline{n}) = E\big((\overline{x}_{PS} - \overline{X})^2|\underline{n}\big) = \sum_{h=1}^{L} W_h^2 \left(\frac{1-f_h}{n_h}\right) S_{hx}^2,$$

$$\mathrm{Cov}\{(\overline{y}_{PS}, \overline{x}_{PS})|\underline{n}\}$$

$$= E\big\{\big((\overline{y}_{PS} - \overline{Y})(\overline{x}_{PS} - \overline{X})\big)|\underline{n}\big\} = \sum_{h=1}^{L} W_h^2 \left(\frac{1-f_h}{n_h}\right) S_{hxy},$$

see [1], where

$$f_h = \frac{n_h}{N_h}, \quad S_{hy}^2 = \frac{1}{N_h - 1}\sum_{i=1}^{N_h} (y_{hi} - \overline{Y}_h)^2,$$

$$S_{hx}^2 = \frac{1}{N_h - 1}\sum_{i=1}^{N_h} (x_{hi} - \overline{X}_h)^2, \quad S_{hxy} = \frac{1}{N_h - 1}\sum_{i=1}^{N_h} (x_{hi} - \overline{X}_h)(y_{hi} - \overline{Y}_h).$$

Using the results from [13] for $E(n_h^{-1})$, to the terms of order n^{-2}, we have

$$\mathrm{Var}(\overline{y}_{PS}) = \left(\frac{1-f}{n}\right)\sum_{h=1}^{L} W_h S_{hy}^2 + \left(\frac{N-n}{N-1}\right)\frac{1}{n^2}\sum_{h=1}^{L}(1-W_h)S_{hy}^2,$$

$$\mathrm{Var}(\overline{x}_{PS}) = \left(\frac{1-f}{n}\right)\sum_{h=1}^{L} W_h S_{hx}^2 + \left(\frac{N-n}{N-1}\right)\frac{1}{n^2}\sum_{h=1}^{L}(1-W_h)S_{hx}^2,$$

$$\mathrm{Cov}(\overline{y}_{PS}, \overline{x}_{PS}) = \left(\frac{1-f}{n}\right)\sum_{h=1}^{L} W_h S_{hxy} + \left(\frac{N-n}{N-1}\right)\frac{1}{n^2}\sum_{h=1}^{L}(1-W_h)S_{hxy},$$

where $f = n/N$ is over all sampling fraction.

It is known that when the auxiliary information is used at the estimation stage, the ratio estimator is the best among a wide class of estimators when the relation between y and x, the variate under study and the auxiliary variate respectively, is a straight line through the origin and the variance of y about this line is proportional to x, see [14]. In such a situation the ratio estimator is as good as regression estimator. In many practical situations, the regression line does not pass through the origin. In these situations, the ratio estimator does not perform equally well as that of regression estimator. Keeping this fact in view and also due to the stronger intuitive appeal statisticians are more inclined towards the use of the ratio and the product estimators and hence a large amount of work has been carried out towards the modification of ratio and product estimators, for instance, see [11, 15–17] etc. These authors have proposed various estimators under simple random sampling without replacement (SRSWOR) and stratified random technique which under some realistic conditions is more efficient than the mean per unit estimator, the ratio and the product estimator are efficient as the linear regression estimator in optimum case. It is to be mentioned that the problem of estimation of population mean $\overline{Y}$ of the study variate y based on post-stratification and auxiliary information has not attracted much attention of survey statisticians, for instance, [12] and [18].

In this paper, following approaches developed by [19] and [20], we have suggested a family of estimators of population mean $\bar{Y}$ of the study variate y based on post-stratification using multi-auxiliary variate and its properties are studied.

When information on p-auxiliary variates $x_1, x_2, \ldots, x_p$ is available. Let W_h ($h = 1, 2, \ldots, L$) and $\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_p$ be the known strata weights and the known population means of the auxiliary variates $\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_p$ respectively. Suppose the observations (y_{hi}, x_{khi}), $i = 1, 2, \ldots, n_h$, $h = 1, 2, \ldots, L$ and $k = 1, 2, \ldots, p$ are available. We denote

$$\overline{x}_{kPS} = \sum_{h=1}^{L} W_h \overline{x}_{kh}, \quad \overline{x}_{kh} = \frac{1}{n_h}\sum_{i=1}^{n_h} x_{khi},$$

$$\overline{X}_k = \sum_{h=1}^{L} W_h \overline{X}_{kh}, \quad \overline{X}_{kh} = \frac{1}{N_h}\sum_{i=1}^{N_h} x_{khi}.$$

Let $\underline{\overline{x}}_{PS}$ denote the column vector of p-elements $\overline{x}_{1PS}, \overline{x}_{2PS}, \ldots, \overline{x}_{pPS}$. Superfix T over a column vector denotes the corresponding row vector.

Defining $\varepsilon_0 = (\overline{y}_{PS} - \overline{Y})$, $\varepsilon_k = (\overline{x}_{kPS} - \overline{X}_k)$ and $\varepsilon^T = (\varepsilon_1, \varepsilon_2, \ldots \varepsilon_p)$, we have for a given configuration of $\underline{n} = (n_1, n_2, \ldots, n_L)$, the values of the conditional expectations:

$$E(\varepsilon_0|\underline{n}) = 0 = E(\varepsilon_k|\underline{n})$$

and if n_h is large, to terms of order n_h^{-1}, the conditional expected values are

$$\begin{aligned}
E(\varepsilon_0^2|\underline{n}) &= \sum_{h=1}^{L} W_h^2 \left(\frac{1-f_h}{n_h}\right) S_{h0}^2, \\
E(\varepsilon_k^2|\underline{n}) &= \sum_{h=1}^{L} W_h^2 \left(\frac{1-f_h}{n_h}\right) S_{hk}^2, \\
E(\varepsilon_0\varepsilon_k|\underline{n}) &= \sum_{h=1}^{L} W_h^2 \left(\frac{1-f_h}{n_h}\right) S_{h0k}, \\
E(\varepsilon_k\varepsilon_l|\underline{n}) &= \sum_{h=1}^{L} W_h^2 \left(\frac{1-f_h}{n_h}\right) S_{hkl},
\end{aligned} \tag{1}$$

where

$$S_{h0k} = \rho_{h0k} S_{h0} S_{hk} \Rightarrow \rho_{h0k} = \frac{S_{h0k}}{S_{h0} S_{hk}}, \quad S_{hkl} = \rho_{hkl} S_{hk} S_{hl} \Rightarrow \rho_{hkl} = \frac{S_{hkl}}{S_{hk} S_{hl}},$$

$$S_{h0}^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} \left(y_{hi} - \overline{Y}_h\right)^2, \quad S_{hk}^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} \left(x_{hki} - \overline{X}_{hk}\right)^2,$$

$$S_{h0k} = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} \left(y_{hi} - \overline{Y}_h\right)\left(x_{hki} - \overline{X}_{hk}\right),$$

$$S_{hkl} = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} \left(x_{hki} - \overline{X}_{hk}\right)\left(x_{hli} - \overline{X}_{hl}\right).$$

Putting the above results in matrix notations, we have

$$E(\varepsilon|\underline{n}) = 0, \quad E\left(\varepsilon\varepsilon^T|\underline{n}\right) = D, \quad E(\varepsilon_0\varepsilon|\underline{n}) = A, \tag{2}$$

where

$$A^T = (a_1, a_2, \ldots, a_p), \quad a_k = \sum_{h=1}^{L} W_h^2 \left(\frac{1-f_h}{n_h}\right) S_{h0k},$$

$$D = [d_{kl}]_{p\times p}, \quad d_{kl} = \sum_{h=1}^{L} W_h^2 \left(\frac{1-f_h}{n_h}\right) S_{hkl}.$$

The unconditional expectations are:

$$E(\varepsilon_0) = E(\varepsilon_k) = 0, \quad \text{for all } \; k = 1, 2, \ldots, p$$

and for large n, to terms of order n^{-1}, the unconditional expected values are given by

$$\begin{aligned} E(\varepsilon_0^2) &= \frac{1-f}{n}\sum_{h=1}^{L} W_h S_{h0}^2, \\ E(\varepsilon_k^2) &= \frac{1-f}{n}\sum_{h=1}^{L} W_h S_{hk}^2, \\ E(\varepsilon_0\varepsilon_k) &= \frac{1-f}{n}\sum_{h=1}^{L} W_h S_{h0k}, \\ E(\varepsilon_k\varepsilon_l) &= \frac{1-f}{n}\sum_{h=1}^{L} W_h S_{hkl}. \end{aligned} \tag{3}$$

Putting the above results in matrix notation, we have

$$E(\varepsilon) = 0, \quad E(\varepsilon\varepsilon^T) = A^*, \quad E(\varepsilon_0\varepsilon) = D^*, \tag{4}$$

where

$$A^{*T} = (a_1^*, a_2^*, \ldots, a_p^*), \quad a_k^* = \frac{1-f}{n}\sum_{h=1}^{L} W_h S_{h0k},$$

$$D^* = [d_{kl}^*]_{p\times p}, \quad \text{and} \quad d_{kl}^* = \frac{1-f}{n}\sum_{h=1}^{L} W_h S_{hkl}.$$

2 The suggested family of estimators

Let $\underline{\overline{X}}^T = (\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_p)$ denote the row vector of p elements $\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_p$. Whatever be the sample chosen let $(\overline{y}_{PS}, \underline{\overline{x}}_{PS}^T)$ assume values in a closed convex subset, Q, of the $(p+1)$ dimensional real space containing the point $(\overline{Y}, \underline{\overline{X}}^T)$. We suggest a family of post-stratified estimators for the population mean using multi-auxiliary variable as:

$$\widehat{\overline{Y}}_G = G(\overline{y}_{PS}, \overline{x}_{1PS}, \overline{x}_{2PS}, \ldots, \overline{x}_{pPS}) = G\Big(\overline{y}_{PS}, \underline{\overline{x}}_{PS}^T\Big), \tag{5}$$

where $G(\overline{y}_{PS}, \underline{\overline{x}}_{PS}^T)$ is a function of $\overline{y}_{PS}, \overline{x}_{1PS}, \overline{x}_{2PS}, \ldots, \overline{x}_{pPS}$ such that

$$G\Big(\overline{Y}, \underline{\overline{X}}^T\Big) = \overline{Y}, \quad \text{for all } \; \overline{Y} \tag{6}$$

and such that it satisfies the following conditions:

1. The function $G(\overline{y}_{PS}, \underline{\overline{x}}_{PS}^T)$ is continuous and bounded in Q,
2. The first and second order partial derivatives of the function $G(\overline{y}_{PS}, \underline{\overline{x}}_{PS}^T)$ exist and are continuous and bounded in Q.

Expanding the function $G(\overline{y}_{PS}, \underline{\overline{x}}_{PS}^T)$ about the point $(\overline{Y}, \underline{\overline{X}}^T)$ in a second order Taylor's series, we obtain:

$$\begin{aligned}\widehat{\overline{Y}}_G = G\big(\overline{Y}, \underline{\overline{X}}^T\big) &+ \big(\overline{y}_{PS} - \overline{Y}\big)\frac{\partial G(\cdot)}{\partial \overline{y}_{PS}}\bigg|_{\big(\overline{Y}, \underline{\overline{X}}^T\big)} + \big(\underline{\overline{x}}_{PS} - \underline{\overline{X}}\big)^T G^{(1)}\big(\overline{Y}, \underline{\overline{X}}^T\big) \\ &+ \frac{1}{2}\Bigg\{\big(\overline{y}_{PS} - \overline{Y}\big)^2 \frac{\partial^2 G(\cdot)}{\partial \overline{y}_{PS}^2}\bigg|_{\big(\overline{y}^*_{PS}, \underline{\overline{x}}^{*T}_{PS}\big)} \\ &\quad + 2\big(\overline{y}_{PS} - \overline{Y}\big)\big(\underline{\overline{x}}_{PS} - \underline{\overline{X}}\big)^T \frac{\partial G^{(1)}(\cdot)}{\partial \overline{y}_{PS}}\bigg|_{\big(\overline{y}^*_{PS}, \underline{\overline{x}}^{*T}_{PS}\big)} \\ &\quad + \big(\underline{\overline{x}}_{PS} - \underline{\overline{X}}\big)^T G^{(2)}\big(\overline{y}^*_{PS}, \underline{\overline{x}}^{*T}_{PS}\big)\big(\underline{\overline{x}}_{PS} - \underline{\overline{X}}\big)\Bigg\}, \end{aligned} \tag{7}$$

where $\overline{y}^*_{PS} = \overline{Y} + \xi(\overline{y}_{PS} - \overline{Y})$, $\underline{\overline{x}}^*_{PS} = \underline{\overline{X}} + \xi(\underline{\overline{x}}_{PS} - \underline{\overline{X}})$, $0 < \xi < 1$, $G^{(1)}$ denotes the p elements column vector of first partial derivatives of $G(\cdot)$ i.e. $G^{(1)T} = (G_1^{(1)}, G_2^{(1)}, \ldots, G_p^{(1)})$ with $G_k^{(1)} = (\partial G(\overline{y}_{PS}, \underline{\overline{x}}_{PS})/\partial \overline{x}_{kPS})|_{(\overline{Y}, \overline{X}_k)}$ and $G^{(2)}$ denotes the $p \times p$ matrix of the second partial derivatives of $G(\cdot)$ with respect to $\underline{\overline{x}}_{PS}$ about the point $(\overline{Y}, \underline{\overline{X}}^T)$. Expressing (7) in terms of ε's and noting that $G(\overline{Y}, \underline{\overline{X}}^T) = \overline{Y}^T$, we have

$$\begin{aligned}\widehat{\overline{Y}}_G = \overline{Y} &+ \frac{\partial G(\cdot)}{\partial \overline{y}_{PS}}\bigg|_{\big(\overline{Y}, \underline{\overline{X}}^T\big)} + \varepsilon^T G^{(1)}(\overline{Y}, \underline{\overline{X}}^T) \\ &+ \frac{1}{2}\Bigg\{\varepsilon_0^2 \frac{\partial^2 G(\cdot)}{\partial \overline{y}_{PS}^2}\bigg|_{\big(\overline{y}^*_{PS}, \underline{\overline{x}}^{*T}_{PS}\big)} + 2\varepsilon_0 \varepsilon^T \frac{\partial G^{(1)}(\cdot)}{\partial \overline{y}_{PS}}\bigg|_{\big(\overline{y}^*_{PS}, \underline{\overline{x}}^{*T}_{PS}\big)} \\ &\quad + \varepsilon^T G^{(2)}\big(\overline{Y}^*_{PS}, \underline{\overline{x}}^{*T}_{PS}\big)\varepsilon\Bigg\} \end{aligned} \tag{8}$$

Taking conditional expectation in (8) and noting that second derivatives are bounded. Thus we arrived at the following theorem:

Theorem 1.

$$E\big(\widehat{\overline{Y}}_G \big| \underline{n}\big) = \overline{Y} + o\big(n_h^{-1}\big).$$

From Theorem 1, it follows that the bias of the estimator $\widehat{\overline{Y}}_G$ is of the order n_h^{-1}, and hence its contribution to the mean squared error of $\widehat{\overline{Y}}_G$ will be of the order of n_h^{-2}.

Now we prove the following result:

Theorem 2. *Up to terms of order n_h^{-1}, the conditional variance of $\widehat{\overline{Y}}_G$ is minimized for*

$$G^{(1)}\big(\overline{Y}, \underline{\overline{X}}^T\big) = -D^{-1}A \tag{9}$$

and the conditional minimum variance is given by

$$\mathrm{Var}\big(\widehat{\overline{Y}}_G|\underline{n}\big) = \big(1-R^2\big)S_0^{*2}. \tag{10}$$

Proof. From (8), we have upto terms of order n_h^{-1},

$$\begin{aligned}
\mathrm{Var}\big(\widehat{\overline{Y}}_G|\underline{n}\big) &= E\big\{\big(\widehat{\overline{Y}}_G - \overline{Y}\big)^2|\underline{n}\big\} \\
&= E\Bigg\{\bigg(\varepsilon_0 \frac{\partial G(\cdot)}{\partial \overline{y}_{PS}}\bigg|_{\left(\overline{Y},\underline{\overline{X}}^T\right)} + \varepsilon^T G^{(1)}\big(\overline{Y}, \underline{\overline{X}}^T\big)\bigg)^2\Big|\underline{n}\Bigg\} \\
&= E\Big\{\big(\varepsilon_0 + \varepsilon^T G^{(1)}\big(\overline{Y}, \underline{\overline{X}}^T\big)\big)^2|\underline{n}\Big\},
\end{aligned}$$

from (6) which implies that $\frac{\partial G(\cdot)}{\partial \overline{y}_{PS}}|_{(\overline{Y},\underline{\overline{X}}^T)} = 1$,

$$\begin{aligned}
\mathrm{Var}\big(\widehat{\overline{Y}}_G|\underline{n}\big) &= \Big[E\big(\varepsilon_0^2|\underline{n}\big) + 2E\big(\varepsilon_0\varepsilon^T|\underline{n}\big)G^{(1)}\big(\overline{Y}, \underline{\overline{X}}^T\big) \\
&\qquad + \big(G^{(1)}\big(\overline{Y}, \underline{\overline{X}}^T\big)\big)^T E\big(\varepsilon\varepsilon^T|\underline{n}\big)\big(G^{(1)}\big(\overline{Y}, \underline{\overline{X}}^T\big)\big)\Big] \\
&= S_0^{*2} + 2A^T G^{(1)}\big(\overline{Y}, \underline{\overline{X}}^T\big) + \big(G^{(1)}\big(\overline{Y}, \underline{\overline{X}}^T\big)\big)^T D\big(G^{(1)}\big(\overline{Y}, \underline{\overline{X}}^T\big)\big)
\end{aligned} \tag{11}$$

which is minimized for

$$G_{opt}^{(1)} = -D^{-1}A = \delta_0 \quad \text{(say)}, \tag{12}$$

where $S_0^{*2} = \sum_{h=1}^{L} W_h^2 (\frac{1-f_h}{n_h}) S_{h0}^2$.

Thus the resulting conditional variance is given by

$$\min.\mathrm{Var}\big(\widehat{\overline{Y}}_G|\underline{n}\big) = \big(1-R^2\big)S_0^{*2}, \tag{13}$$

where $R^2 = \frac{A^T D^{-1} A}{S_0^{*2}}$ and R is the multiple correlation coefficient between $\overline{y}_{PS}$ and the vector $\underline{\overline{x}}_{PS}$. Hence proved the Theorem 2. □

The conditional variance of any estimator of the class (5) can be obtained from (11). From (11) the conditional minimum variance (i.e. $\min.\mathrm{Var}(\widehat{\overline{Y}}_G|\underline{n})$) is not larger than the conditional variance of the unbiased estimator $\overline{y}_{PS}$, since $A^T D^{-1} A > 0$.

Taking unconditional expectation in (8) and noting that second derivatives are bounded, we have:

Theorem 3.

$$E\big(\widehat{\overline{Y}}_G\big) = \overline{Y} + o\big(n^{-1}\big).$$

Theorem 3 shows that the bias of the estimator $\widehat{\overline{Y}}_G$ is of the order n^{-1}, and hence its contribution to the mean square error (MSE) of $\widehat{\overline{Y}}_G$ will be of the order n^{-2}. Thus, to the first order of approximation, the unconditional variance of $\widehat{\overline{Y}}_G$ will be the same.

Theorem 4. *Upto terms of order n^{-1}, the unconditional $\mathrm{Var}(\widehat{\overline{Y}}_G)$ is minimized for*

$$G^{(1)}\big(\overline{Y}, \underline{\overline{X}}^T\big) = -D^{*-1}A^* \tag{14}$$

and the unconditional minimum variance of $\widehat{\overline{Y}}_G$ is given by

$$\min.\mathrm{Var}\big(\widehat{\overline{Y}}_G\big) = \big(1 - R^{*2}\big)S_0^{**2}, \tag{15}$$

where $S_0^{**2} = \frac{1-f}{n}\sum_{h=1}^{L} W_h S_{h0}^2$.

Proof. From (8), we have upto terms of order n^{-1},

$$\begin{aligned}\mathrm{Var}\big(\widehat{\overline{Y}}_G\big) &= E\big(\widehat{\overline{Y}}_G - \overline{Y}\big)^2 = E\left(\varepsilon_0 \frac{\partial G(\cdot)}{\partial \overline{y}_{PS}}\bigg|_{\left(\overline{Y}, \underline{\overline{X}}^T\right)} + \varepsilon^T G^{(1)}\big(\overline{Y}, \underline{\overline{X}}^T\big)\right)^2 \\ &= E\big(\varepsilon_0 + \varepsilon^T G^{(1)}\big(\overline{Y}, \underline{\overline{X}}^T\big)\big)^2,\end{aligned}$$

from (6) which implies that $\frac{\partial G(\cdot)}{\partial \overline{y}_{PS}}|_{(\overline{Y}, \underline{\overline{X}}^T)} = 1$,

$$\begin{aligned}\mathrm{Var}\big(\widehat{\overline{Y}}_G\big) = \Big[&E\big(\varepsilon_0^2\big) + 2E\big(\varepsilon_0 \varepsilon^T\big) G^{(1)}\big(\overline{Y}, \underline{\overline{X}}^T\big) \\ &+ \big(G^{(1)}\big(\overline{Y}, \underline{\overline{X}}^T\big)\big)^T E\big(\varepsilon\varepsilon^T\big)\big(G^{(1)}\big(\overline{Y}, \underline{\overline{X}}^T\big)\big)\Big].\end{aligned}$$

Using the results (3) and (4) in the above expression we get the unconditional variance over all possible distribution, for large n, to the terms of order n^{-1}, as:

$$\begin{aligned}\mathrm{Var}\big(\widehat{\overline{Y}}_G\big) = &S_0^{**2} + 2A^{*T} G^{(1)}\big(\overline{Y}, \underline{\overline{X}}^T\big) \\ &+ \big(\big(G^{(1)}\big(\overline{Y}, \underline{\overline{X}}^T\big)\big)^T D^* \big(G^{(1)}\big(\overline{Y}, \underline{\overline{X}}^T\big)\big)\end{aligned} \tag{16}$$

which is minimized for

$$G_{opt}^{(1)} = -D^{*-1}A^* = G_0^{(1)} \quad \text{(say)}, \tag{17}$$

where $S_0^{**2} = \frac{1-f}{n_h}\sum_{h=1}^{L} W_h S_{h0}^2$.

Thus the resulting unconditional variance of $\widehat{\overline{Y}}_G$ is given by

$$\min.\mathrm{Var}\big(\widehat{\overline{Y}}_G\big) = \big(1 - R^{*2}\big)S_0^{**2}, \tag{18}$$

where $R^{*2} = \frac{A^{*T} D^{*-1} A^*}{S_0^{**2}}$ and R^* is the multiple correlation coefficient between $\overline{y}_{PS}$ and the vector $\underline{\overline{x}}_{PS}$. Hence proved the Theorem 4. □

The unconditional variance of any estimator of the class (5) can be obtained from (16). From (16), the $\min\mathrm{Var}(\widehat{\overline{Y}}_G)$ is not large than the unconditional variance of the unbiased estimator y_{PS}, since $A^{*T}D^{*-1}A^* > 0$.

Let $G^{(1)}\big(\overline{Y}, \underline{\overline{X}}^T\big) = -\alpha G_0^{(1)} = -\alpha D^{*-1}A^*$, is a departure from the optimum value ($\alpha > 0$ is a constant), we have

$$\begin{aligned}\mathrm{Var}\big(\widehat{\overline{Y}}_G\big) &= \big[S_0^{**2} - 2\alpha A^{*T}D^{*-1}A^* + \alpha^2 A^{*T}D^{*-1}A^*\big] \\ &= \big[S_0^{**2} - \alpha(2-\alpha)A^{*T}D^{*-1}A^*\big]. \end{aligned} \tag{19}$$

It is well known that the unconditional variance of the usual unbiased estimator $\overline{y}_{PS}$ is

$$\mathrm{Var}\big(\widehat{\overline{Y}}_G\big) = \frac{1-f}{n}\sum_{h=1}^{L} W_h S_{h0}^2 = S_0^{**2}. \tag{20}$$

Thus for any $G^{(1)}(\overline{Y}, \underline{\overline{X}}^T)$, we find from (18) and (19) that

$$\mathrm{Var}(\overline{y}_{PS}) - \mathrm{Var}\big(\widehat{\overline{Y}}_G\big) = \alpha(2-\alpha)A^{*T}D^{*-1}A^* \tag{21}$$

which shows that the proposed class of estimators $\widehat{\overline{Y}}_G$ would be better than usual unbiased estimator $\overline{y}_{PS}$ as for as $0 < \alpha < 2$.

Remark 1. It is to be mentioned that optimum estimators in the class are not unique but all of them have the same variance given either by (13) or (18). We also note that in practice the value of $\delta_0 = -D^{-1}A$ at (12) or $G_0^{(1)} = -D^{*-1}A^* = \delta_0^*$ at (17) may not be known. However, they can be estimated by

$$\widehat{\delta}_0 = \widehat{\delta}_0^* - \widehat{D}^{-1}\widehat{A}, \tag{22}$$

where

$$\widehat{D} = \big[\widehat{d}_{kl}\big]_{p\times p}, \quad \widehat{d}_{kl} = \sum_{h=1}^{L} W_h^2\left(\frac{1-f_h}{n_h}\right)s_{hkl},$$

$$\widehat{A}^T = \big(\widehat{a}_1, \widehat{a}_2, \ldots, \widehat{a}_p\big), \quad \widehat{a}_k = \sum_{h=1}^{L} W_h^2\left(\frac{1-f_h}{n_h}\right)s_{h0k},$$

$$s_{h0k} = \frac{1}{n_h - 1}\sum_{i=1}^{n_h}(y_{hi} - \overline{y}_h)(x_{hki} - \overline{x}_{hk}),$$

$$s_{hkl} = \frac{1}{n_h - 1}\sum_{i=1}^{n_h}(x_{hki} - \overline{x}_{hk})(x_{hli} - \overline{x}_{hl}), \quad \overline{y}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} y_{hi},$$

$$\overline{x}_{kh} = \frac{1}{n_h}\sum_{i=1}^{n_h} x_{khi}.$$

In such a case we may define a class of estimators based on estimated optimum value $\widehat{\delta}_0$ as:

$$\widehat{\overline{Y}}^{*}_{G} = G^{*}\big(\overline{y}_{PS}, \underline{\overline{x}}^{T}_{PS}, \widehat{\delta}^{T}_{0}\big), \tag{23}$$

where $G^{*}\big(\overline{y}_{PS}, \underline{\overline{x}}^{T}_{PS}, \widehat{\delta}^{T}_{0}\big)$ is a function of $(\overline{y}_{PS}, \underline{\overline{x}}^{T}_{PS}, \widehat{\delta}^{T}_{0})$ such that:

$$\begin{aligned}
&G^{*}\big(\overline{Y}, \underline{\overline{X}}^{T}, \delta^{T}\big) = \overline{Y},\\
&\left.\frac{\partial G^{*}(\cdot)}{\partial \overline{y}_{PS}}\right|_{\left(\overline{Y}, \underline{\overline{X}}^{T}, \delta^{T}\right)} = 1,\\
&\left.\frac{\partial G^{*}(\cdot)}{\partial \underline{\overline{x}}_{PS}}\right|_{\left(\overline{Y}, \underline{\overline{X}}^{T}, \delta^{T}\right)} = \delta_0 = -D^{-1}A,\\
&\left.\frac{\partial G^{*}(\cdot)}{\partial \widehat{\delta}_0}\right|_{\left(\overline{Y}, \underline{\overline{X}}^{T}, \delta^{T}\right)} = 0.
\end{aligned} \tag{24}$$

Under (24) the class of estimators $\widehat{\overline{Y}}^{*}_{G}$ at (23) is expected to have, to the first order of approximation, the conditional and unconditional variances respectively as

$$\mathrm{Var}\Big(\widehat{\overline{Y}}^{*}_{G}|\underline{n}\Big) = \min.\mathrm{Var}\big(\widehat{\overline{Y}}^{*}_{G}|\underline{n}\big) = \big(1 - R^2\big)S^{*2}_{0} \tag{25}$$

and

$$\mathrm{Var}\Big(\widehat{\overline{Y}}^{*}_{G}\Big) = \min.\mathrm{Var}\big(\widehat{\overline{Y}}^{*}_{G}\big) = \big(1 - R^{*2}\big)S^{**2}_{0}. \tag{26}$$

3 Comparison with corresponding unstratified multivariate estimators

We assume that information on p auxiliary variates $x_1, x_2, \ldots, x_p$ is available. A simple random sample of size n is drawn from the given finite population of size N. Let y_i and x_i denote the values of the variates y and x_k of the i-th unit of the sample, $k = 1, 2, \ldots, p$; $i = 1, 2, \ldots, n$. Defining:

$$\begin{aligned}
&\overline{y} = \frac{1}{n}\sum_{i=1}^{n}\overline{y}_i, \quad \overline{x}_k = \frac{1}{n}\sum_{i=1}^{n}x_{ki}, \quad \overline{Y} = \sum_{i=1}^{N}y_i, \quad \overline{X}_k = \frac{1}{N}\sum_{i=1}^{N}x_{ki},\\
&S^2_0 = \frac{1}{N-1}\sum_{i=1}^{N}\big(y_i - \overline{Y}\big)^2, \quad S^2_k = \frac{1}{N-1}\sum_{i=1}^{N}\big(x_{ki} - \overline{X}_k\big)^2,\\
&S_{0k} = \frac{1}{N-1}\sum_{i=1}^{N}\big(y_i - \overline{Y}\big)\big(x_{ki} - \overline{X}_k\big).
\end{aligned}$$

Further ρ_{0k} and ρ_{kl} denote the correlation coefficients between the variates y and x_k and between the x_k and x_l.

Define $\underline{\overline{x}}^T = (\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_k)$, $\varepsilon_0^* = (\overline{y} - \overline{Y})$ and $\varepsilon_k^* = (\overline{x}_k - \overline{X}_k)$ such that:

$$E(\varepsilon_0^*) = 0, \quad E(\varepsilon_k^*) = 0 \quad \text{for all} \quad k = 1, 2, \ldots, p,$$
$$E(\varepsilon_0^{*2}) = \frac{1-f}{n} S_0^2, \quad E(\varepsilon_k^2) = \frac{1-f}{n} S_k^2,$$
$$E(\varepsilon_0^* \varepsilon_k^*) = \frac{1-f}{n} b_k, \quad E(\varepsilon_k^* \varepsilon_l^*) = \frac{1-f}{n} q_{kl},$$

where $(kl) = 1, 2, \ldots, p$, $b_k = S_{0k} = \rho_{0k} S_0 S_k$, $q_{kl} = \rho_{kl} S_k S_l$.

Putting the above results in matrix notations, we have

$$E(\varepsilon_0^*) = 0, \quad E(\varepsilon_0^* \varepsilon^*) = \frac{1-f}{n} b, \quad E(\varepsilon_0^* \varepsilon^{*T}) = \frac{1-f}{n} Q,$$

where, $b^T = (b_1, b_2, \ldots, b_p)$, $Q = [q_{kl}]_{p \times p}$.

Let $\underline{\overline{X}}^T = (\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_p)$ denote the row vector of p elements $\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_p$.

Whatever be the sample chosen, let $(\overline{y}, \underline{\overline{x}}^T)$ assume values in a closed convex subset, W, of the $(p+1)$ dimensional real space containing the point $(\overline{Y}, \underline{\overline{X}}^T)$. Following [21] one may define a class of estimator of population mean $\overline{Y}$ as

$$\widehat{\overline{Y}}_G^{(1)} = G\big(\overline{y}, \underline{\overline{x}}^T\big) = G\big(\overline{y}, \overline{x}_1, \overline{x}_2, \ldots, \overline{x}_p\big), \tag{27}$$

where $G(\overline{y}, \underline{\overline{x}}^T)$ is a function of $\overline{y}, \overline{x}_1, \overline{x}_2, \ldots, \overline{x}_p$ such that

$$G\big(\overline{Y}, \underline{\overline{X}}^T\big) = \overline{Y}, \quad \text{for all } \overline{Y}$$

and such that it satisfies the following conditions:

1. The function $G(\overline{y}, \underline{\overline{x}}^T)$ is continuous and bounded in W.
2. The first and second order partial derivatives of the function $G(\overline{y}, \underline{\overline{x}}^T)$ exist and are continuous and bounded in W.

To the first degree of approximation, the variance of $\widehat{\overline{Y}}_G^{(1)}$ is given by

$$\mathrm{Var}\big(\widehat{\overline{Y}}_G^{(1)}\big) = \frac{1-f}{n}\Big[S_0^2 + 2b^T G^{(1)}\big(\overline{Y}, \underline{\overline{X}}^T\big) + \big(G^{(1)}\big(\overline{Y}, \underline{\overline{X}}^T\big)\big)^T Q \big(G^{(1)}\big(\overline{Y}, \underline{\overline{X}}^T\big)\big)\Big] \tag{28}$$

which is minimized when

$$G^{(1)}\big(\overline{Y}, \underline{\overline{X}}^T\big) = -Q^{-1} b = \eta_0 \quad \text{(say)}, \tag{29}$$

where $G^{(1)}(\overline{Y}, \underline{\overline{X}}^T)$ denotes the p elements column vector of the partial derivatives of $G(\overline{y}, \underline{\overline{x}}^T)$ with respect to $\underline{\overline{x}}^T$ about the point $(\overline{Y}, \underline{\overline{X}}^T)$.

Thus the resulting minimum variance of $\widehat{\overline{Y}}_G^{(1)}$ is given by

$$\min.\mathrm{Var}\big(\widehat{\overline{Y}}_G^{(1)}\big) = \frac{1-f}{n}\big(1 - R^{**2}\big)S_0^2, \tag{30}$$

where $R^{**2} = (b^T Q^{-1} b)/S_0^2$ and R^{**} is the multiple correlation coefficient between y and $(x_1, x_2, \ldots, x_p)$.

Following [18], and from (18) and (23) it can be shown that the proposed class of estimators $\widehat{\overline{Y}}_G$ in post-stratified sampling in unconditionally more efficient than the corresponding unstratified class of estimators $\widehat{\overline{Y}}_G^{(1)}$.

Remark 2. In practice, the exact optimum value η_0 of $G^{(1)}(\overline{Y}, \underline{\overline{X}}^T)$ at (29) is not known, it is available to replace it by its consistent estimate of η_0 from the sample data at hand. Thus following the procedure outlined in [21], we define a class of estimators for population mean $\overline{Y}$ (based on estimated optimum values) as:

$$\widehat{\widehat{\overline{Y}}}_G^{(1)} = G\big(\overline{y}, \underline{\overline{x}}^T, \widehat{\eta}_0^T\big), \tag{31}$$

where

$$\widehat{\eta}_0 = -\widehat{Q}^{-1}\widehat{b} \text{ with } \widehat{Q} = [\widehat{q}_{kl}]_{p\times p},\ \widehat{q}_{kl} = s_{kl},\ \widehat{b} = (\widehat{b}_1, \widehat{b}_2, \ldots, \widehat{b}_k),\ \widehat{b}_k = s_{0k},$$
$$s_{0k} = \frac{1}{n-1}\sum_{i=1}^{n}\big(y_i - \overline{y}\big)\big(x_{ki} - \overline{x}_k\big),\ \ s_{kl} = \frac{1}{n-1}\sum_{i=1}^{n}\big(x_{ki} - \overline{x}_k\big)\big(x_{li} - \overline{x}_l\big).$$

Now $G(\overline{y}, \underline{\overline{x}}^T, \widehat{\eta}_0^T)$ is a function of $(\overline{y}, \underline{\overline{x}}^T, \widehat{\eta}_0^T)$ such that:

$$\begin{aligned}
&G\big(\overline{Y}, \underline{\overline{X}}^T, \eta_0^T\big) = \overline{Y},\\
&\left.\frac{\partial G(\cdot)}{\partial \overline{y}}\right|_{\left(\overline{Y}, \underline{\overline{X}}^T, \eta_0^T\right)} = 1,\\
&\left.\frac{\partial G(\cdot)}{\partial \underline{\overline{x}}_{PS}}\right|_{\left(\overline{Y}, \underline{\overline{X}}^T, \eta_0^T\right)} = \eta_0,\\
&\left.\frac{\partial G(\cdot)}{\partial \widehat{\eta}_0}\right|_{\left(\overline{Y}, \underline{\overline{X}}^T, \eta_0^T\right)} = 0.
\end{aligned} \tag{32}$$

Under the condition (32), it can be shown to the first degree of approximation that the variance of $\widehat{\widehat{\overline{Y}}}_G^{(1)}$ is

$$\mathrm{Var}\Big(\widehat{\widehat{\overline{Y}}}_G^{(1)}\Big) = \min.\mathrm{Var}\big(\widehat{\overline{Y}}_G^{(1)}\big) = \frac{1-f}{n}\big(1 - R^{**2}\big)S_0^2. \tag{33}$$

From (26) and (33), we

$$\mathrm{Var}\Big(\widehat{\overline{\widehat{Y}}}^{*}_{G}\Big) < \mathrm{Var}\Big(\widehat{\overline{\widehat{Y}}}^{(1)}_{G}\Big),$$

if

$$\big(1-R^{*2}\big)S_0^{**2} = \frac{1-f}{n}\big(1-R^{**2}\big)S_0^2. \tag{34}$$

Thus the proposed class of estimators $\widehat{\overline{\widehat{Y}}}^{*}_{G}$ based on estimated optimum values in post-stratified sampling would be better than the corresponding non-stratified class of estimators $\widehat{\overline{\widehat{Y}}}^{(1)}_{G}$ based on estimated optimum values in simple random sampling without replacement (SRSWOR), if the condition (34) holds true.

4 Empirical study

In the empirical study, we consider the relative efficiency of the post-stratified sampling estimator $\widehat{\overline{Y}}_G$ ($=\widehat{\theta}_1$, say) and non-stratified estimator $\widehat{\overline{Y}}^{(1)}_{G}$ ($=\widehat{\theta}_2$, say) with respect to the simple sample mean estimator without using an auxiliary information $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ ($=\widehat{\theta}_0$, say). The percent relative efficiency of the estimator $\widehat{\theta}_j$, $j = 1, 2$ with respect to the estimator $\widehat{\theta}_0$ is computed as:

$$RE\big(\widehat{\theta}_0, \widehat{\theta}_j\big) = \frac{V(\widehat{\theta}_0)}{V(\widehat{\theta}_j)} \times 100\,\% = RE(0, j), \quad \text{(say)}. \tag{35}$$

We consider the problem of estimation of Forced Expiratory Volume (FEV) of the Pulmonary Disease persons, based on a dataset of 654 persons available on the CD with the book by [22], by using their age and height at the estimation stage, and using the other variables gender and smoking status as post-stratification variables. Thus the population of 654 persons has been divided into four post-strata. Post-stratum 1 consists of non-smoking females $(0, 0)$, post-stratum 2 consists of non-smoking males $(0, 1)$, post-stratum 3 consists of smoking females $(1, 0)$, and post-stratum 4 consists of smoking males $(1, 1)$. The descriptive parameters of the three variables: FEV, Age and $Height$ (HT) in the four post-strata are given in Table 1. The population correlation coefficients between the three variables in the four post-strata are given in Table 2.

In order to have a closer look at the data structure in four different post-strata, we have also devoted Fig. 1 to display the three variables.

To investigate various situations, we apply power transformations on the study variable in all the four strata as $Y_i = (FEV)^T$ for different choice of values of T in the range of 0.1 to 2.5 with a step of 0.1. The other two variables: $X_{1i} = (Age)$ and $X_{2i} = (Height)$ were used at the estimation stage. We decided to select a sample of size being 10 % of the total population size, and later we post-stratified the sample based on gender and smoking status into four different homogeneous groups. A total of

sample size $n = 65.4$ (can be rounded to 65) was selected from the entire population of $N = 654$ persons. Out of 65.4 persons, 27.9 persons were found to be from stratum-1, 3.9 persons were from stratum-2, 31.0 persons were from stratum-3 and 2.6 persons were from stratum-4. We used R-code given in the Appendix to produce the results shown in Table 3.

Table 1. Descriptive parameters of FEV, Age and HT.

Stratum	N_h	Mean	St. Dev	Minimum	Q_1	Median	Q_3	Maximum	Skewness	Kurtosis
					FEV					
1	279	2.3792	0.6393	0.7910	1.8770	2.4170	2.8660	3.8160	-0.07	-0.69
2	39	2.9659	0.4229	2.1980	2.6770	3.0740	3.2080	3.8350	-0.25	-0.39
3	310	2.7344	0.9741	0.7960	1.9565	2.5475	3.3578	5.7930	0.70	-0.13
4	26	3.7430	0.8890	1.6940	3.3420	3.878	4.4300	4.8720	-0.89	0.13
					Age					
1	279	9.366	2.693	3.00	8.00	9.00	11.00	18.00	0.42	0.48
2	39	13.256	2.245	10.00	11.00	13.00	15.00	19.00	0.65	0.42
3	310	9.687	2.778	3.00	8.00	10.00	11.00	19.00	0.43	0.33
4	26	13.923	2.465	9.00	12.00	14.00	16.00	18.00	-0.13	-0.78
					HT					
1	279	59.605	4.739	46.00	57.00	60.50	63.00	71.00	-0.60	-0.10
2	39	64.551	2.291	60.00	63.00	65.00	66.00	69.50	0.09	-0.68
3	310	61.519	6.268	47.00	57.00	61.00	67.00	74.00	-0.11	-0.90
4	26	68.058	3.232	58.00	67.00	68.00	69.75	72.00	-1.63	3.86

Table 2. Pearson correlation coefficient values for four strata.

Correlations	Stratum-1 $(0,0)$		Correlations	Stratum-2 $(0,1)$	
	FEV	Age	FEV	Age	
Age	0.767	—	Age	-0.047	—
HT	0.843	0.776	HT	0.251	-0.092
	Stratum-3 $(1,0)$			Stratum-4 $(1,1)$	
	0.822	—	Age	0.394	—
HT	0.883	0.842	HT	0.750	0.352

Table 3. Relative efficiency of the post-stratified and non-stratified estimators with respect to the sample mean estimator.

T	Correlations	Stratum-1	Stratum-2	Stratum-3	Stratum-4	$RE(0,1)$	$RE(0,2)$
0.2	ρ_{yx_1}	0.76839	-0.05719	0.82536	0.42073		
	ρ_{yx_2}	0.85847	0.24024	0.90837	0.78591	558.13	514.11
	$\rho_{x_1x_2}$	0.77642	-0.09219	0.84230	0.35209		

T	Correlations	Stratum-1	Stratum-2	Stratum-3	Stratum-4	$RE(0,1)$	$RE(0,2)$
0.3	ρ_{yx_1}	0.76675	−0.05889	0.82367	0.42475		
	ρ_{yx_2}	0.85921	0.23834	0.91010	0.79138	557.63	518.16
	$\rho_{x_1x_2}$	0.77642	−0.09219	0.84230	0.35209		
0.4	ρ_{yx_1}	0.76623	−0.05933	0.82311	0.42574		
	ρ_{yx_2}	0.85923	0.23785	0.91040	0.79273	556.72	518.43
	$\rho_{x_1x_2}$	0.77642	−0.09219	0.84230	0.35209		
0.5	ρ_{yx_1}	0.76604	−0.05947	0.82291	0.42607		
	ρ_{yx_2}	0.85931	0.23768	0.91048	0.79317	556.34	518.44
	$\rho_{x_1x_2}$	0.77642	−0.09219	0.84230	0.35209		
0.6	ρ_{yx_1}	0.76597	−0.05953	0.82283	0.42620		
	ρ_{yx_2}	0.85931	0.23762	0.91051	0.79335	556.17	518.44
	$\rho_{x_1x_2}$	0.77642	−0.09219	0.84230	0.35209		
0.7	ρ_{yx_1}	0.76593	−0.05956	0.82279	0.42626		
	ρ_{yx_2}	0.85931	0.23759	0.91053	0.79343	556.10	518.44
	$\rho_{x_1x_2}$	0.77642	−0.09219	0.84230	0.35209		
0.8	ρ_{yx_1}	0.76592	−0.05957	0.82278	0.42628		
	ρ_{yx_2}	0.85931	0.23757	0.91054	0.79347	556.06	518.43
	$\rho_{x_1x_2}$	0.77642	−0.09219	0.84230	0.35209		
0.9	ρ_{yx_1}	0.76591	−0.05957	0.82277	0.42630		
	ρ_{yx_2}	0.85931	0.23757	0.91054	0.79348	556.05	518.43
	$\rho_{x_1x_2}$	0.77642	−0.09219	0.84230	0.35209		
1.0	ρ_{yx_1}	0.76591	−0.05957	0.82277	0.42630		
	ρ_{yx_2}	0.85931	0.23757	0.91054	0.79348	556.05	518.43
	$\rho_{x_1x_2}$	0.77642	−0.09219	0.84230	0.35209		
1.1	ρ_{yx_1}	0.76592	−0.05957	0.82277	0.42629		
	ρ_{yx_2}	0.85931	0.23757	0.91053	0.79347	556.06	518.43
	$\rho_{x_1x_2}$	0.77642	−0.09219	0.84230	0.35209		
1.2	ρ_{yx_1}	0.76593	−0.05956	0.82279	0.42626		
	ρ_{yx_2}	0.85931	0.23758	0.91053	0.79344	556.09	518.44
	$\rho_{x_1x_2}$	0.77642	−0.09219	0.84230	0.35209		
1.3	ρ_{yx_1}	0.76598	−0.05954	0.82281	0.42622		
	ρ_{yx_2}	0.85931	0.23760	0.91052	0.79339	556.14	518.44
	$\rho_{x_1x_2}$	0.77642	−0.09219	0.84230	0.35209		
1.4	ρ_{yx_1}	0.76599	−0.05951	0.82285	0.42616		
	ρ_{yx_2}	0.85931	0.23764	0.91050	0.79330	556.23	518.44
	$\rho_{x_1x_2}$	0.77642	−0.09219	0.84230	0.35209		
1.5	ρ_{yx_1}	0.76606	−0.05946	0.82293	0.42604		
	ρ_{yx_2}	0.85930	0.23770	0.91047	0.79313	556.37	518.44
	$\rho_{x_1x_2}$	0.77642	−0.09219	0.84230	0.35209		
1.6	ρ_{yx_1}	0.76618	−0.05936	0.82306	0.42582		
	ρ_{yx_2}	0.85929	0.23780	0.91042	0.79284	556.63	518.44
	$\rho_{x_1x_2}$	0.77642	−0.09219	0.84230	0.35209		

T	Correlations	Stratum-1	Stratum-2	Stratum-3	Stratum-4	$RE(0,1)$	$RE(0,2)$
1.7	ρ_{yx_1}	0.76640	−0.05919	0.82329	0.42542		
	ρ_{yx_2}	0.85927	0.23800	0.91031	0.79230	557.05	518.38
	$\rho_{x_1x_2}$	0.77642	−0.09219	0.84230	0.35209		
1.8	ρ_{yx_1}	0.76681	−0.05884	0.82373	0.42463		
	ρ_{yx_2}	0.85920	0.23839	0.91007	0.79123	557.73	518.11
	$\rho_{x_1x_2}$	0.77642	−0.09219	0.84230	0.35209		
1.9	ρ_{yx_1}	0.76755	−0.05814	0.82451	0.42301		
	ρ_{yx_2}	0.85897	0.23918	0.90946	0.78903	558.47	516.96
	$\rho_{x_1x_2}$	0.77642	−0.09219	0.84230	0.35209		
2.0	ρ_{yx_1}	0.76878	−0.05666	0.82574	0.41945		
	ρ_{yx_2}	0.85810	0.24082	0.90766	0.78420	557.30	511.93
	$\rho_{x_1x_2}$	0.77642	−0.09219	0.84230	0.35209		
2.1	ρ_{yx_1}	0.76992	−0.05338	0.82653	0.41105		
	ρ_{yx_2}	0.85451	0.24439	0.90143	0.77282	542.98	490.27
	$\rho_{x_1x_2}$	0.77642	−0.09219	0.84230	0.35209		
2.2	ρ_{yx_1}	0.76560	−0.04576	0.81962	0.38930		
	ρ_{yx_2}	0.83905	0.25223	0.87710	0.74336	471.43	410.63
	$\rho_{x_1x_2}$	0.77642	−0.09219	0.84230	0.35209		
2.3	ρ_{yx_1}	0.72779	−0.02781	0.76664	0.32980		
	ρ_{yx_2}	0.77699	0.26831	0.78138	0.66352	296.36	247.49
	$\rho_{x_1x_2}$	0.77642	−0.09219	0.84230	0.35209		
2.4	ρ_{yx_1}	0.57491	0.00564	0.55799	0.20157		
	ρ_{yx_2}	0.59403	0.28686	0.50344	0.50005	148.39	130.84
	$\rho_{x_1x_2}$	0.77642	−0.09219	0.84230	0.35209		
2.5	ρ_{yx_1}	0.31371	−0.00472	0.28826	0.06937		
	ρ_{yx_2}	0.33364	0.26138	0.20492	0.34442	105.93	103.92
	$\rho_{x_1x_2}$	0.77642	−0.09219	0.84230	0.35209		

For $T = 0.2$, the values of the population correlation coefficients between FEV and Age are 0.76839, -0.05719, 0.82536 and 0.42073 in the first, second, third and fourth post-stratum, respectively. The values of the population correlation coefficients between FEV and $Height$ are 0.85847, 0.24024, 0.90837 and 0.78591 in the first, second, third and fourth stratum, respectively. In the same way, the values of the populations correlation coefficients between Age and $Height$ are 0.77642, -0.09219, 0.84230 and 0.35209 in the first, second, third and fourth stratum respectively. In this particular situation, the percent relative efficiency of the post-stratified sampling estimator $\widehat{\theta}_1$ with respect to the simple sample mean estimator $\widehat{\theta}_0$ remains 558.13 % and that of the non-stratified estimator $\widehat{\theta}_2$ remains 514.11 %. In the same way, the results in Table 3 are readable for other values of T. It is to be noted that so long as the value of T is less than or equal to 2.0, the percent relative efficiency of the post-stratified estimator remains around 557 % and that of the non-stratified estimator remains around 511 %. As soon as the value of T becomes 2.3, the relative efficiency of the post-stratified estimator drastically reduces

to 296.36 % and that of non-stratified estimator reduces to 247.49 %. For higher value of T equal to 2.5, the relative efficiency of the post-stratified sampling estimator reduces to 105.93 % and that of non-stratified sampling estimator reduces to 103.92 %. Thus, we conclude that the proposed post stratified sampling estimator can be used to estimate population mean of a study variable in the presence of multi-auxiliary variables more efficiently than a non-stratified sampling estimator.

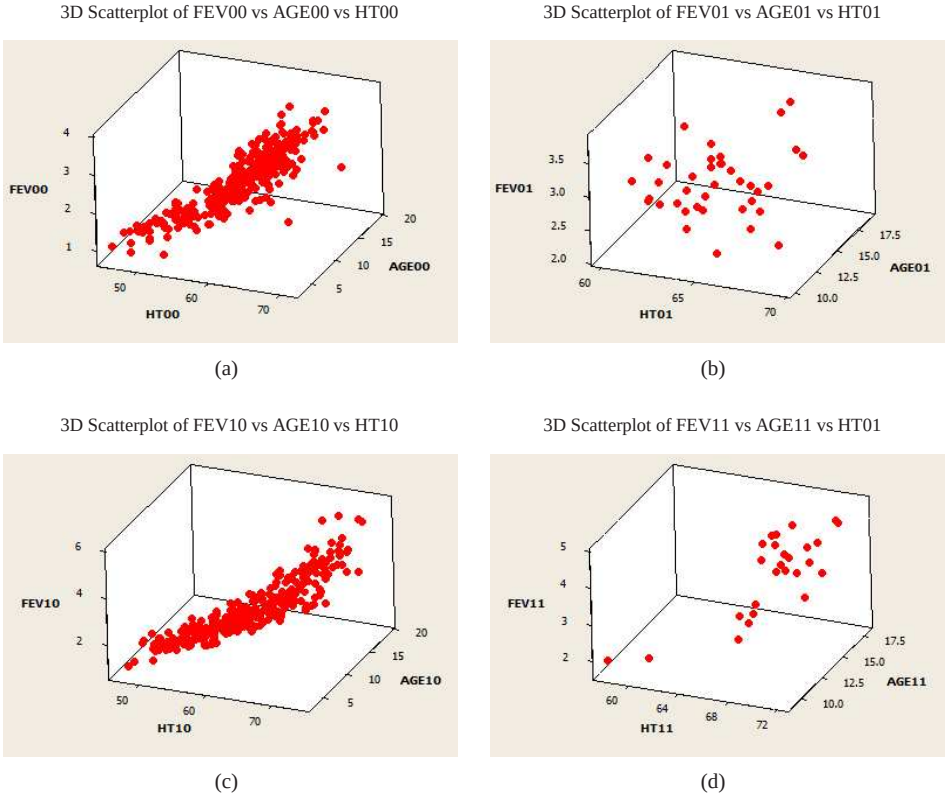

Fig. 1. Pictorial representation of four post-strata: (a) stratum-1 (Female $=0$, Smoking $=0$); (b) stratum-2 (Female $=0$, Smoking $=1$); (c) stratum-3 (Male $=1$, Smoking $=0$); (d) stratum-4 (Male $=1$, Smoking $=1$).

Appendix

```
#R-Code used in the simulation study (File Name: post2.r)
names<-c(0,0,0,0,0);
inp11<-read.fwf("c:\\rc\\out00",c(14,10,8,5,5),header=FALSE,
sep="\t", as.is=FALSE,skip=0,col.names=names);
```

```
inp10<-read.fwf("c:\\rc\\out01",c(14,10,8,5,5),header=FALSE,
sep="\t",as.is=FALSE,skip=0,col.names=names);
inp01<-read.fwf("c:\\rc\\out10",c(14,10,8,5,5),header=FALSE,
sep="\t",as.is=FALSE,skip=0,col.names=names);
inp00<-read.fwf("c:\\rc\\out11",c(14,10,8,5,5),header=FALSE,
sep="\t",as.is=FALSE,skip=0,col.names=names);
y1<-c(inp11[[1]])
x11<-c(inp11[[2]])
x12<-c(inp11[[3]])
sex11<-c(inp11[[4]])
sk11<-c(inp11[[5]])
y2<-c(inp10[[1]])
x21<-c(inp10[[2]])
x22<-c(inp10[[3]])
sex10<-c(inp10[[4]])
sk10<-c(inp10[[5]])
y3<-c(inp01[[1]])
x31<-c(inp01[[2]])
x32<-c(inp01[[3]])
sex01<-c(inp01[[4]])
sk01<-c(inp01[[5]])
y4<-c(inp00[[1]])
x41<-c(inp00[[2]])
x42<-c(inp00[[3]])
sex00<-c(inp00[[4]])
sk00<-c(inp00[[5]])
np1<-length(y1)
np2<-length(y2)
np3<-length(y3)
np4<-length(y4)
np<-np1+np2+np3+np4
print(c(np1,np2,np3,np4,np))
w1<-np1/np
w2<-np2/np
w3<-np3/np
w4<-np4/np
ns<-0.10*np
ns1<-ns*np1/np
ns2<-ns*np2/np
ns3<-ns*np3/np
ns4<-ns*np4/np
print(c(ns1,ns2,ns3,ns4,ns))
f1<-(1-ns1/np1)/ns1
f2<-(1-ns2/np2)/ns2
f3<-(1-ns3/np3)/ns3
f4<-(1-ns3/np3)/ns3
t<-0.1
```

```
for (i in 1:25) {
t<-t+0.1
print(c('t=',t))
y1<-(y1)^t
x11<-(x11)
x12<-(x12)
ry1x11<-cov(y1,x11)/sqrt(var(y1)*var(x11))
ry1x12<-cov(y1,x12)/sqrt(var(y1)*var(x12))
rx11x12<-cov(x11,x12)/sqrt(var(x11)*var(x12))
print(c('ry1x11=',ry1x11,'ry1x12=',ry1x12,'rx11x12=',rx11x12))
y2<-(y2)^t
x21<-(x21)
x22<-(x22)
ry2x21<-cov(y2,x21)/sqrt(var(y2)*var(x21))
ry2x22<-cov(y2,x22)/sqrt(var(y2)*var(x22))
rx21x22<-cov(x21,x22)/sqrt(var(x21)*var(x22))
print(c('ry2x21=',ry2x21,'ry2x22=',ry2x22,'rx21x22=',rx21x22))
y3<-(y3)^t
x31<-(x31)
x32<-(x32)
ry3x31<-cov(y3,x31)/sqrt(var(y3)*var(x31))
ry3x32<-cov(y3,x32)/sqrt(var(y3)*var(x32))
rx31x32<-cov(x31,x32)/sqrt(var(x31)*var(x32))
print(c('ry3x31=',ry3x31,'ry3x32=',ry3x32,'rx31x32=',rx31x32))
y4<-(y4)^t
x41<-(x41)
x42<-(x42)
ry4x41<-cov(y4,x41)/sqrt(var(y4)*var(x41))
ry4x42<-cov(y4,x42)/sqrt(var(y4)*var(x42))
rx41x42<-cov(x41,x42)/sqrt(var(x41)*var(x42))
print(c('ry4x41=',ry4x41,'ry4x42=',ry4x42,'rx41x42=',rx41x42))
s02<-f1*w1^2*var(y1)+f2*w2^2*var(y2)+f3*w3^2*var(y3)+
f4*w4^2*var(y4)
a<-matrix(0,1,2)
d<-matrix(0,2,2)
a[1]<-f1*w1^2*cov(y1,x11)+f2*w2^2*cov(y2,x21)+
f3*w3^2*cov(y3,x31)+ f4*w4^2*cov(y4,x41)
a[2]<-f1*w1^2*cov(y1,x12)+f2*w2^2*cov(y2,x22)+
f3*w3^2*cov(y3,x32)+ f4*w4^2*cov(y4,x42)
d[1,1]<-f1*w1^2*cov(x11,x11)+f2*w2^2*cov(x21,x21)+
f3*w3^2*cov(x31,x31)+ f4*w4^2*cov(x41,x41)
d[1,2]<-f1*w1^2*cov(x11,x12)+f2*w2^2*cov(x21,x22)+
f3*w3^2*cov(x31,x32)+ f4*w4^2*cov(x41,x42)
d[2,1]<-d[1,2]
d[2,2]<-f1*w1^2*cov(x12,x12)+f2*w2^2*cov(x22,x22)+
f3*w3^2*cov(x32,x32)+ f4*w4^2*cov(x42,x42)
#print(d)
```

```
#print(a)
invd<-solve(d)
#print(invd)
out<-a%*%invd%*%t(a)
rsq<-out/s02
y<-c(y1,y2,y3,y4)
x1<-c(x11,x21,x31,x41)
x2<-c(x12,x22,x32,x42)
vary<-(1-ns/np)*var(y)/ns
re1<-vary*100/(s02*(1-rsq))
print(c("re (post stattification)=",re1))
b<-matrix(0,1,2)
q<-matrix(0,2,2)
b[1]<-cov(y,x1)
b[2]<-cov(y,x2)
q[1,1]<-cov(x1,x1)
q[1,2]<-cov(x1,x2)
q[2,1]<-q[1,2]
q[2,2]<-cov(x2,x2)
#print(q)
#print(b)
invq<-solve(q)
#print(invq)
out1<-b%*%invq%*%t(b)
rsq1<-out1/var(y)
re2<-100/(1-rsq1)
print(c("re (no stratification)=",re2))
}
```

Acknowledgements

The authors are thankful to the Journal Secretary Dr. Romas Baronas and a referee for fruitful comments on the original version of the manuscript.

References

1. D. Holt, T.M.F. Smith, Post-stratification, *J. Roy. Stat. Soc. A Sta.*, **142**, pp. 33–46, 1979.

2. P.V. Sukhatme, B.V. Sukhatme, S. Sukhatme, C. Ashok, *Sampling Theory of Surveys with Applications*, Iowa State University Press, Ames, Iowa, 1984.

3. W. Fuller, Estimation employing post-strata, *J. Am. Stat. Assoc.*, **61**, pp. 1172–1183, 1966.

4. D.C. Doss, H.O. Hartly, G.R. Somayajulu, An exact small sample theory for post-stratification *J. Stat. Plan. Infer.*, **3**, pp. 235–249, 1979.

5. P. Jagers, A. Oden, L. Trulsson, Post-stratification and ratio estimation: Usages of auxiliary information in survey sampling and opinion polls, *Int. Stat. Rev.*, **53**, pp. 221–238, 1985.

6. P. Jagers, Post-stratification against bias in sampling, *Int. Stat. Rev.*, **55**, pp. 159–167, 1986.

7. M.C. Agrawal, K.B. Panda, An efficient estimator in post-stratification, *Metron*, **51**, pp. 179–188, 1993.

8. M.C. Agrawal, K.B. Panda, On efficient estimation in post-stratification, *Metron*, **53**, pp. 107–115, 1995.

9. M. Ruiz Espejo, D. Pineda, On variance estimation for post-stratification: a review, *Metron*, **55**, pp. 209–220, 1997.

10. H.P. Singh, M. Ruiz Espejo, Improved post-stratified estimation, *B. Int. Statist. Inst.*, **60**, pp. 341–342, 2003.

11. H.P. Singh, G.K. Vishwakarma, An efficient variant of the product and ratio estimators in stratified random sampling, *Statistics in Transition*, **7**(6), pp. 1311–1325, 2006.

12. A.F. Ige, T.P. Tripathi, Estimation of population mean using post-stratification and auxiliary information, *Abacus*, **18**(2), pp. 265–276, 1989.

13. F. Stephan, The expected value and variance of the reciprocal and other negative powers of a positive Bernoullian variate, *Ann. Math. Stat.*, **16**, pp. 50–61, 1945.

14. W.G. Cochran, *Sampling Techniques*, 3rd ed., Wiley Eastern Ltd., New York, 1977.

15. H.P. Singh, A generalized class of estimators of ratio, product and mean using supplementary information on an auxiliary character in PPSWR sampling scheme, *Gujarat Stat. Rev.*, **13**(2), pp. 1–30, 1986.

16. H.P. Singh, M. Ruiz Espejo, On linear regression and ratio-product estimation of a finite population mean, *Statistician*, **52**(1), pp. 59–67, 2003.

17. H.P. Singh, G.K. Vishwakarma, A family of estimators of population mean using auxiliary information in stratified sampling, *Commun. Stat.-Theor. M.*, **37**(7), pp. 1038–1050, 2008.

18. R.K. Tuteja, S. Bahl, T.P. Tripathi, Sampling strategies for population mean based on post-stratification and multivariate auxiliary information, in: *Proceedings-III, International Symposium on Optimization and Statistics, Dec. 19–21, held at Aligarh Muslim University, India*, pp. 56–61, 1995.

19. S.K. Srivastava, A class of estimators using auxiliary information in sample surveys, *Can. J. Stat.*, **8**, pp. 253–254, 1980.

20. H.P. Singh, G.K. Vishwakarma, Estimation of mean using auxiliary information and post-stratification, *Commun. Stat.-Theor. M.*, 2009 (accepted).

21. S.K. Srivastava, H.S. Jhajj, A class of estimators of the population mean using multi-auxiliary information, *Calcutta Stat. Assoc.*, **32**, pp. 47–56, 1983.

22. B. Rosner, *Fundamentals of Biostatistics*, Thomson-Brooks/Cole, 6rd ed., 2006.