Journal of the Korean Statistical Society (2006), **35**: *4, pp 441–452*

EFFICIENT ESTIMATION OF POPULATION MEAN IN STRATIFIED SAMPLING USING REGRESSION TYPE ESTIMATOR

LOVLEEN KUMAR GROVER[1]

ABSTRACT

Here an efficient regression type estimator for a stratified population mean is proposed under the two-phase sampling scheme. While constructing the proposed estimator, it is assumed that the first auxiliary variable x is directly and highly correlated with the study variable y, and the second auxiliary variable z is directly and highly correlated with the first auxiliary variable x. However the variable z is not directly correlated with the variable y, but they are just correlated with each other only due to their direct and high correlation with the variable x. The proposed regression type estimator is found to be always more efficient than the existing estimators defined under the same situation.

AMS 2000 subject classifications. Primary 62D05; Secondary 94A20.
Keywords. Auxiliary variable, regression type estimator, simple random sampling without replacement, study variable, two-phase sampling scheme, variance.

1. INTRODUCTION OF USE OF REGRESSION TYPE ESTIMATORS IN STRATIFIED SAMPLING

In practice, the populations often consist of heterogeneous units with respect to the variable under study. So in such situations, it may be necessary to divide the whole population into several non-overlapping sub-populations (called strata), which are approximately homogenous in nature with respect to the variable under study. In order to estimate the population parameter of interest, samples of predetermined sizes are drawn independently from each stratum by using simple random sampling with or without replacement. Since a stratified sample consists of units selected separately from each stratum therefore such a

Received March 2006; accepted August 2006.
[1]Department of Mathematics, Guru Nanak Dev University, Amritsar-143 005, Punjab, India (e-mail: lovleen_2@yahoo.co.in)

sample is expected to be a better representation of the whole population than a simple random sample selected from the whole universe or population. Let the population consisting N units be divided into k strata. Consider a stratified random sampling with N_h and n_h as population stratum size and sample size respectively for the h^{th} stratum $(h = 1, 2, \ldots, k)$ so that $\sum_{h=1}^{k} N_h = N$ and $\sum_{h=1}^{k} n_h = n$. Thus the stratified sample, consisting n units, is a sample from a population of size N. Let $\overline{Y}_h$ be the population mean of study variable y in the h^{th} stratum. The population mean of the study variable y for the entire population is $\overline{Y} = 1/N \sum_{h=1}^{k} N_h \overline{Y}_h$. Obviously to get the estimate of $\overline{Y}$, we have to estimate $\overline{Y}_h$, for all h. The ordinary unbiased estimate of the population mean $\overline{Y}$, under the stratified random sampling, is given by

$$\widehat{\overline{y}}_{st} = \sum_{h=1}^{k} W_h \overline{y}_h, \tag{1.1}$$

where $W_h = N_h/N$ and $\overline{y}_h$ is the sample mean of variable y in the h^{th} stratum such that $E(\overline{y}_h) = \overline{Y}_h$, $h = 1, 2, \ldots, k$.

Let there exists an auxiliary variable x, which is directly and highly correlated with the study variable y. Since Sukhatme and Sukhatme (1970, p. 214) pointed out that separate difference estimator is more efficient than combined difference estimator, so our next effort will concentrate only on the separate difference estimator. Now if the population mean $\overline{X}_h$ of auxiliary variable x for the h^{th} stratum is known in advance for all $h = 1, 2, \ldots, k$ then the usual unbiased separate difference estimator of $\overline{Y}$ is

$$\widehat{\overline{y}}_{DS} = \sum_{h=1}^{k} W_h \widehat{\overline{y}}_{hDS}, \tag{1.2}$$

where in the h^{th} stratum, $\widehat{\overline{y}}_{hDS} = \overline{y}_h + B_{yx}^{(h)}(\overline{X}_h - \overline{x}_h)$ is the usual difference estimator of $\overline{Y}_h$, $\overline{y}_h$ and $\overline{x}_h$ are the sample means of variables y and x respectively, and $B_{yx}^{(h)}$ is the population regression coefficient of y on x whose value is assumed to be known.

If the values of $B_{yx}^{(h)}$'s are not known then these can be suitably estimated which results in the following usual separate regression estimator of $\overline{Y}$:

$$\widehat{\overline{y}}_{rS} = \sum_{h=1}^{k} W_h \widehat{\overline{y}}_{hrS}, \tag{1.3}$$

where in the h^{th} stratum, $\widehat{\overline{y}}_{hrS} = \overline{y}_h + b_{yx}^{(h)}(\overline{X}_h - \overline{x}_h)$ is the ordinary regression estimator of $\overline{Y}_h$, and $b_{yx}^{(h)}$ is the sample regression coefficient of y on x that gives a consistent estimator of $B_{yx}^{(h)}$.

When the prior information regarding $\overline{X}_h$'s is not available then one can use two phase sampling scheme to estimate first of all $\overline{X}_h$ and then ultimately estimate $\overline{Y}_h$. The usual separate regression estimator of $\overline{Y}$, under the two phase sampling scheme, is defined as

$$\widehat{\overline{y}}_{rdS} = \sum_{h=1}^{k} W_h \widehat{\overline{y}}_{hrdS}, \tag{1.4}$$

where in the h^{th} stratum, $\widehat{\overline{y}}_{hrdS} = \overline{y}_h + b_{yx}^{(h)}(\overline{x}'_h - \overline{x}_h)$ is the ordinary regression estimator of $\overline{Y}_h$ under two phase sampling scheme, $\overline{x}'_h$ is the sample mean of auxiliary variable x based on first phase sample drawn from the h^{th} stratum, and $\overline{y}_h$, $\overline{x}_h$ and $b_{yx}^{(h)}$ are the means of variables y and x, and the regression coefficient of y on x respectively based on the second phase sample of the h^{th} stratum.

If the prior information regarding $\overline{X}_h$'s is not available but the stratum population means $\overline{Z}_h, h = 1, 2, \ldots, k$ of another auxiliary variable z are known in advance. Here this second auxiliary variable z is strongly correlated with the first auxiliary variable x, whereas the correlation between the variables z and y is remote in certain sense. For example,

(i) In any agricultural experiment, 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 only due to 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 remotely correlated with each other.

Under such situations, Sahoo and Bala (2000) suggested the following regression type estimator of $\overline{Y}$ by using two phase sampling scheme:

$$\widehat{\overline{y}}_{rdzS} = \sum_{h=1}^{k} W_h \widehat{\overline{y}}_{hrdzS}, \tag{1.5}$$

where in the h^{th} stratum, $\widehat{\overline{y}}_{hrdzS} = \overline{y}_h + b_{yx}^{(h)}(\overline{x}'_h - \overline{x}_h) + b_{yz}^{(h)}(\overline{Z}_h - \overline{z}'_h)$ which is similar to the estimator proposed by Sahoo *et al.* (1993), $\overline{z}'_h$ is the sample mean of auxiliary variable z based on first phase sample drawn from the h^{th} stratum; and $b_{yz}^{(h)}$ is the regression coefficient of y on z based on the second phase sample of the h^{th} stratum.

Sahoo and Bala (2000) have shown that their suggested estimator $\widehat{\overline{y}}_{rdzS}$ is always more efficient than both the usual stratified regression estimator under the two phase sampling scheme *i.e.* $\widehat{\overline{y}}_{rdS}$ and the usual unbiased estimator in the absence of any auxiliary variable *i.e.* $\widehat{\overline{y}}_{st}$.

For such situations, in the present paper, we propose a new stratified regression type estimator of $\overline{Y}$ under the two-phase sampling scheme by using two auxiliary variables. It is found that the proposed stratified regression type estimator is always more efficient than all the existing stratified estimators of population mean $\overline{Y}$. The gain in efficiency of the proposed estimator is illustrated by taking an empirical population considered in the literature.

2. PROPOSED DIFFERENCE TYPE ESTIMATOR OF POPULATION MEAN

Consider a population consisting N units, which is divided into k strata so that N_h is the size of the h^{th} stratum, $h = 1, 2, \ldots, k$, Let y, x and z be the study variable, and the first and the second auxiliary variables respectively. It is assumed that the first auxiliary variable x is highly and directly correlated with both the study variable y and the second auxiliary variable z. Whereas the correlation between variables y and z exists only due to their high correlation with variable x.

Let Y_{hi}, X_{hi} and Z_{hi}, $h = 1, 2, \ldots, k$ and $i = 1, 2, \ldots, N_h$ be the values of the variables y, x and z respectively on the i^{th} population unit of the h^{th} stratum. The corresponding small letters will denote the values of the respective variables in the samples. Taking population stratum means of various variables as follows:

$$\overline{Y}_h = \frac{1}{N_h}\sum_{i=1}^{N_h} Y_{hi}, \qquad \overline{X}_h = \frac{1}{N_h}\sum_{i=1}^{N_h} X_{hi}, \qquad \overline{Z}_h = \frac{1}{N_h}\sum_{i=1}^{N_h} Z_{hi}.$$

Here the prior information regarding the values of $\overline{Z}_h$'s is assumed to be available for all strata. Consider the following two-phase sampling scheme in the h^{th} stratum:

(a) A simple random sample without replacement, say s'_h, of fixed size $n'_h(n'_h <$

N_h) is drawn from the h^{th} stratum. On this first phase sample, we take the measurements of variables x and z.

(b) A simple random sample without replacement, say s_h, of fixed size $n_h(n_h < n'_h)$ is drawn from the first phase sample s'_h. On this second phase sample, we take the measurements of variables y and x.

It should be noted that for the stratified sampling all the first phase samples are drawn independently from their respective stratum.

Let $\overline{x}'_h$ and $\overline{z}'_h$ be the sample means of variables x and z respectively based on the first phase sample in the h^{th} stratum. Let again $\overline{y}_h$, $\overline{x}_h$ and $\overline{z}_h$ be the sample means of variables y, x and z respectively based on the second phase sample in the h^{th} stratum. Since the second phase sample is the sub-sample of the first phase sample so the information regarding z variable is also available in the second phase sample and thus we can obtain the value of $\overline{z}_h$.

The different regression or ratio type estimators of $\overline{Y}$, under two-phase sampling scheme, have been suggested by a number of authors such as Mukerjee *et al.* (1987, 2000), Tripathi and Ahmed (1995), Ahmed (1998), Prasad *et al.* (2002), and Roy (2003). All these authors have used the prior information regarding population mean of second auxiliary variable z. By getting motivation from Roy (2003), we propose here the following unbiased chain type difference estimator of $\overline{Y}_h$:

$$\widehat{\overline{y}}_{hef} = \overline{y}_h + k_{1h}[\overline{x}'_h + k_{2h}(\overline{Z}_h - \overline{z}'_h) - \{\overline{x}_h + k_{3h}(\overline{Z}_h - \overline{z}_h)\}], \tag{2.1}$$

where k_{1h}, k_{2h} and k_{3h} are constants. In turn, the proposed unbiased difference type estimator of $\overline{Y}$ is

$$\widehat{\overline{y}}_{efst} = \sum_{h=1}^{k} W_h \widehat{\overline{y}}_{hef}. \tag{2.2}$$

So the proposed estimator $\widehat{\overline{y}}_{efst}$ is simply the extension of the difference type estimator of Roy (2003) to the stratified sampling. It should be noted that the construction of the estimator of the type (2.1) requires the assumption that the variable x is highly correlated with both variables y and z whereas the correlation between y and z is remote in certain sense. Now the next aim is to find some suitable values of constants k_{1h}, k_{2h} and k_{3h}.

3. PROPOSED REGRESSION TYPE ESTIMATOR AND ITS VARIANCE

Since the various samples are drawn independently from the respective stratum therefore the variance of the proposed difference type estimator $\widehat{\overline{y}}_{efst}$ is given by

$$Var(\widehat{\overline{y}}_{efst}) = \sum_{h=1}^{k} W_h^2 Var(\widehat{\overline{y}}_{hef}), \tag{3.1}$$

where $Var(\widehat{\overline{y}}_{hef})$ is the variance of estimator $\overline{y}_{hef}$ and it is given as under:

$$\begin{aligned} Var(\widehat{\overline{y}}_{hef}) = \Big(\frac{1}{n_h} - \frac{1}{N_h}\Big)\Big\{ & S_{yh}^2 + k_{1h}^2 S_{xh}^2 + k_{1h}^2 k_{3h}^2 S_{zh}^2 - 2k_{1h} S_{yxh} \\ & + 2k_{1h}k_{3h}S_{yzh} - 2k_{1h}^2 k_{3h} S_{xzh}\Big\} \\ + \Big(\frac{1}{n_h'} - \frac{1}{N_h}\Big)\Big\{ & k_{1h}^2 k_{2h}^2 S_{zh}^2 - k_{1h}^2 S_{xh}^2 + 2k_{1h} S_{yxh} - 2k_{1h}k_{2h}S_{yzh} \\ & + 2k_{1h}^2 k_{3h} S_{xzh} - 2k_{1h}^2 k_{2h} k_{3h} S_{zh}^2 \Big\} \end{aligned} \tag{3.2}$$

with

$$S_{uh}^2 = \frac{1}{N_h - 1}\sum_{i=1}^{N_h}(U_{hi} - \overline{U}_h)^2, \quad u = y, x, z,$$

$$S_{uvh} = \frac{1}{N_h - 1}\sum_{i=1}^{N_h}(U_{hi} - \overline{U}_h)(V_{hi} - \overline{V}_h), \quad u = y, x, z, \quad v = y, x, z \text{ and } u \neq v,$$

(U_{hi}, V_{hi}) being the value of bivariate (u, v) on the i^{th} population unit of the h^{th} stratum with population mean $(\overline{U}_h, \overline{V}_h)$.

Define $\rho_{uv}^{(h)} = S_{uvh}/S_{uh}S_{vh}$, which is the population correlation coefficient between variables u and v in the h^{th} stratum; $u = y, x, z$, $v = y, x, z$ and $u \neq v$. The values of constants k_{1h}, k_{2h} and k_{3h} are so chosen that the variance $Var(\widehat{\overline{y}}_{efst})$ become minimum or equivalently the variance $Var(\widehat{\overline{y}}_{hef})$ become minimum. The optimum values of these constants which minimizes $Var(\widehat{\overline{y}}_{hef})$ are obtained as

$$k_{1h}^{(opt)} = B_{yx.z}^{(h)}, \tag{3.3}$$

$$k_{2h}^{(opt)} = B_{xz}^{(h)}, \tag{3.4}$$

$$k_{3h}^{(opt)} = B_{xz}^{(h)} - \frac{B_{yz}^{(h)}}{B_{yx.z}^{(h)}}, \tag{3.5}$$

where

$B_{xz}^{(h)} = S_{xzh}/S_{zh}^2$, population regression coefficient of x on z in the h^{th} stratum,

$B_{yz}^{(h)} = S_{yzh}/S_{zh}^2$, population regression coefficient of y on z in the h^{th} stratum,

$B_{yx.z}^{(h)} = \{(\rho_{yx}^{(h)} - \rho_{yz}^{(h)}\rho_{xz}^{(h)})/(1 - \rho_{xz}^{(h)^2})\} \times (S_{yh}/S_{xh})$, population partial regression coefficient of y on x keeping z fixed in the h^{th} stratum.

On substituting the optimum values of constants k_{1h}, k_{2h} and k_{3h}, as obtained in (3.3) to (3.5), in the expression (3.2), we can obtain the following minimum value of $\text{Var}(\widehat{\overline{y}}_{hef})$:

$$\text{Var}(\widehat{\overline{y}}_{hef}) = S_{yh}^2\Big\{\Big(\frac{1}{n_h} - \frac{1}{N_h}\Big)\Big(1 - \rho_{y.xz}^{(h)^2}\Big) + \Big(\frac{1}{n'_h} - \frac{1}{N_h}\Big)\Big(1 - \rho_{yz}^{(h)^2}\Big)\rho_{yx.z}^{(h)^2}\Big\}, \quad (3.6)$$

where

$$\rho_{y.xz}^{(h)} = \sqrt{\frac{\rho_{yx}^{(h)^2} + \rho_{yz}^{(h)^2} - 2\rho_{yx}^{(h)}\rho_{yz}^{(h)}\rho_{xz}^{(h)}}{1 - \rho_{xz}^{(h)^2}}}$$

denotes population multiple correlation coefficient between y and (x, z) and

$$\rho_{yx.z}^{(h)} = \frac{\rho_{yx}^{(h)} - \rho_{yz}^{(h)}\rho_{xz}^{(h)}}{\sqrt{\Big(1 - \rho_{yz}^{(h)^2}\Big)\Big(1 - \rho_{xz}^{(h)^2}\Big)}}$$

denotes population partial correlation coefficient between variables y and x keeping z fixed. Ultimately the minimum value of $\text{Var}(\widehat{\overline{y}}_{efst})$ is given by

$$\begin{aligned}\text{minVar}(\widehat{\overline{y}}_{efst}) &= \sum_{h=1}^{k} W_h^2 \text{minVar}(\widehat{\overline{y}}_{hef}) \\ &= \sum_{h=1}^{k} W_h^2 S_{yh}^2\Big\{\Big(\frac{1}{n_h} - \frac{1}{N_h}\Big)\Big(1 - \rho_{y.xz}^{(h)^2}\Big) \\ &\quad + \Big(\frac{1}{n'_h} - \frac{1}{N_h}\Big)\Big(1 - \rho_{yz}^{(h)^2}\Big)\rho_{yx.z}^{(h)^2}\Big\}. \quad (3.7)\end{aligned}$$

Since the optimum values $k_{1h}^{(opt)}, k_{2h}^{(opt)}$ and $k_{3h}^{(opt)}$ as obtained in (3.3), (3.4) and (3.5) respectively are the functions of unknown population quantities, in general. So, we can estimate these optimum values just by replacing unknown population quantities with their respective analogous consistent estimators based on the second phase sample as given below:

$$\left.\begin{aligned} \widehat{k}_{1h}^{(opt)} &= b_{yx.z}^{(h)} \\ \widehat{k}_{2h}^{(opt)} &= b_{xz}^{(h)} \\ \widehat{k}_{3h}^{(opt)} &= b_{xz}^{(h)} - \frac{b_{yz}^{(h)}}{b_{yx.z}^{(h)}} \end{aligned}\right\}, \tag{3.8}$$

where $b_{yz}^{(h)}$, $b_{xz}^{(h)}$ and $b_{yx.z}^{(h)}$ are the analogous consistent estimators based on second phase sample for $B_{yz}^{(h)}$, $B_{xz}^{(h)}$ and $B_{yx.z}^{(h)}$ respectively.

On substituting the corresponding consistent estimators of optimum values *i.e.* $\widehat{k}_{1h}^{(opt)}$, $\widehat{k}_{2h}^{(opt)}$ and $\widehat{k}_{3h}^{(opt)}$ in (2.1) and then in (2.2), the proposed chain regression type estimator of $\overline{Y}$ in stratified sampling situation will take the following form:

$$\widehat{\overline{y}}_{efst}^{(reg)} = \sum_{h=1}^{k} W_h\Big\{\overline{y}_h + b_{yx.z}^{(h)}\Big(\overline{x}'_h - \overline{x}_h\Big) + b_{yx.z}^{(h)} b_{xz}^{(h)}\Big(\overline{z}_h - \overline{z}'_h\Big) + b_{yz}^{(h)}\Big(\overline{Z}_h - \overline{z}_h\Big)\Big\}. \tag{3.9}$$

We can see that he proposed chain regression type estimator $\widehat{\overline{y}}_{efst}^{(reg)}$ is simply the extension of regression type estimator of Roy (2003). Srivastava and Jhajj (1983) have shown that on replacing the unknown population quantities in the optimum values of constants of an estimator of interest with their respective consistent estimators, the variance of the estimator of interest remain the same up to first order of approximation. Obviously our proposed stratified regression type estimator $\widehat{\overline{y}}_{efst}^{(reg)}$, as obtained in (3.9), has the same variance as given in (3.7), up to first order of approximation. Thus the approximate variance of the proposed regression type estimator $\widehat{\overline{y}}_{efst}^{(reg)}$ is

$$\mathrm{Var}(\widehat{\overline{y}}_{efst}^{(reg)}) \cong \sum_{h=1}^{k} W_h^2 S_{yh}^2\Big\{\Big(\frac{1}{n_h} - \frac{1}{N_h}\Big)\Big(1 - \rho_{y.xz}^{(h)^2}\Big) + \Big(\frac{1}{n'_h} - \frac{1}{N_h}\Big)\Big(1 - \rho_{yz}^{(h)^2}\Big)\rho_{yx.z}^{(h)^2}\Big\}. \tag{3.10}$$

REMARK 3.1. By getting motivation from Mukerjee *et al.* (1987) and Tripathi and Ahmed (1995), one can propose the following unbiased difference type estimator of $\overline{Y}_h$:

$$\overline{\overline{y}}_{hd} = \overline{y}_h + \tau_{1h}(\overline{x}_h - \overline{x}'_h) + \tau_{2h}(\overline{z}_h - \overline{z}'_h) + t_h(\overline{Z}_h - \overline{z}_h), \tag{3.11}$$

where τ_{1h}, τ_{2h} and t_h are unknown constants. This is the simple difference type estimator for the case of multi-auxiliary variable. While constructing the difference type estimator of the form (3.11), we make the assumption that both the auxiliary variables x and z are directly correlated with the study variable y. Consequently, the unbiased difference type estimator of $\overline{Y}$ is

$$\overline{\overline{y}}_d = \sum_{h=1}^{k} W_h \overline{\overline{y}}_{hd}. \tag{3.12}$$

Now, we can obtain the regression type estimator of $\overline{Y}$ corresponding to the above difference type estimator $\overline{\overline{y}}_d$ by applying the usual practice of minimizing the variance of $\overline{\overline{y}}_d$ with respect to the unknown constants τ_{1h}, τ_{2h} and t_h. It is interesting to note that the resultant regression type estimator of $\overline{Y}$, so obtained, is exactly the same as obtained earlier in (3.9) and hence its approximate variance is same as obtained in (3.10).

REMARK 3.2. Fuller (2002) gave a broad overview of historical and recent developments in the use of regression estimation. So a comprehensive study about the usage of the regression estimation in survey sampling can be found in his paper.

4. A COMPARATIVE STUDY

To compare the efficiency of the proposed stratified regression type estimator with that of the existing ones, we first of all require the variances (or variances up to first order of approximation) of different existing estimators and therefore these are given as under:

$$\mathrm{Var}(\widehat{\overline{y}}_{st}) = \sum_{h=1}^{k} W_h^2 S_{yh}^2 \Big(\frac{1}{n_h} - \frac{1}{N_h}\Big), \tag{4.1}$$

$$\mathrm{Var}(\widehat{\overline{y}}_{rdS}) \cong \sum_{h=1}^{k} W_h^2 S_{yh}^2 \Big\{\Big(\frac{1}{n_h} - \frac{1}{N_h}\Big) - \Big(\frac{1}{n_h} - \frac{1}{n'_h}\Big)\rho_{yx}^{(h)^2}\Big\}, \tag{4.2}$$

$$\mathrm{Var}(\widehat{\overline{y}}_{rdzS}) \cong \sum_{h=1}^{k} W_h^2 S_{yh}^2 \Big\{\Big(\frac{1}{n_h} - \frac{1}{N_h}\Big) - \Big(\frac{1}{n_h} - \frac{1}{n'_h}\Big)\rho_{yx}^{(h)^2} - \Big(\frac{1}{n'_h} - \frac{1}{N_h}\Big)\rho_{yz}^{(h)^2}\Big\}. \tag{4.3}$$

The variance of the proposed stratified regression type estimator as obtained in *(3.10)* can be rewritten as

$$\operatorname{Var}\left(\widehat{\overline{y}}_{efst}^{(reg)}\right) \cong \sum_{h=1}^{k} W_h^2 S_{yh}^2 \left\{ \left(\frac{1}{n_h} - \frac{1}{N_h}\right) - \left(\frac{1}{n_h} - \frac{1}{n'_h}\right)\rho_{yx}^{(h)^2} - \left(\frac{1}{n'_h} - \frac{1}{N_h}\right)\rho_{yz}^{(h)^2} \right.$$
$$\left. -\left(\frac{1}{n_h} - \frac{1}{n'_h}\right)\frac{\left(\rho_{yz}^{(h)} - \rho_{yx}^{(h)}\rho_{xz}^{(h)}\right)^2}{1-\rho_{xz}^{(h)^2}} \right\}. \tag{4.4}$$

Using (4.1) to (4.4), we can obtain the following results:

$$\operatorname{Var}\left(\widehat{\overline{y}}_{st}\right) - \operatorname{Var}\left(\widehat{\overline{y}}_{efst}^{(reg)}\right) = \sum_{h=1}^{k} W_h^2 S_{yh}^2 \left\{ \left(\frac{1}{n_h} - \frac{1}{n'_h}\right)\rho_{yx}^{(h)^2} + \left(\frac{1}{n'_h} - \frac{1}{N_h}\right)\rho_{yz}^{(h)^2} \right.$$
$$\left. +\left(\frac{1}{n_h} - \frac{1}{n'_h}\right)\frac{\left(\rho_{yz}^{(h)} - \rho_{yx}^{(h)}\rho_{xz}^{(h)}\right)^2}{1-\rho_{xz}^{(h)^2}} \right\} > 0, \tag{4.5}$$

$$\operatorname{Var}\left(\widehat{\overline{y}}_{rdS}\right) - \operatorname{Var}\left(\widehat{\overline{y}}_{efst}^{(reg)}\right) = \sum_{h=1}^{k} W_h^2 S_{yh}^2 \left\{ \left(\frac{1}{n'_h} - \frac{1}{N_h}\right)\rho_{yz}^{(h)^2} \right.$$
$$\left. +\left(\frac{1}{n_h} - \frac{1}{n'_h}\right)\frac{\left(\rho_{yz}^{(h)} - \rho_{yx}^{(h)}\rho_{xz}^{(h)}\right)^2}{1-\rho_{xz}^{(h)^2}} \right\} > 0, \tag{4.6}$$

$$\operatorname{Var}\left(\widehat{\overline{y}}_{rdzS}\right) - \operatorname{Var}\left(\widehat{\overline{y}}_{efst}^{(reg)}\right) = \sum_{h=1}^{k} W_h^2 S_{yh}^2 \left(\frac{1}{n_h} - \frac{1}{n'_h}\right)\frac{\left(\rho_{yz}^{(h)} - \rho_{yx}^{(h)}\rho_{xz}^{(h)}\right)^2}{1-\rho_{xz}^{(h)^2}} > 0. \tag{4.7}$$

REMARK 4.1. From the expressions (4.1) to (4.4), we can see that in addition to the above results obtained, the following inequality always hold good:

$$\operatorname{Var}\left(\widehat{\overline{y}}_{efst}^{(reg)}\right) < \operatorname{Var}\left(\widehat{\overline{y}}_{rdzS}\right) < \operatorname{Var}\left(\widehat{\overline{y}}_{rdS}\right) < \operatorname{Var}\left(\widehat{\overline{y}}_{st}\right). \tag{4.8}$$

REMARK 4.2. From the results obtained (4.5) to (4.8), we can conclude that the proposed separate stratified regression type estimator $\widehat{\overline{y}}_{efst}^{(reg)}$ is always more efficient than the existing separate stratified estimators of population mean $\overline{Y}$.

REMARK 4.3. Moreover from the results (4.7) and (4.8), the proposed stratified regression type estimator $\widehat{\overline{y}}_{efst}^{(reg)}$ become always more efficient than the stratified estimator suggested by Sahoo and Bala (2000) inspite of usage of the same prior information about the second auxiliary variable z in terms of its population mean.

5. NUMERICAL ILLUSTRATION

To have a rough idea about the gain in efficiency for the proposed stratified regression type estimator, we take the same population as considered by Sahoo and Bala (2000). So for this purpose, we consider the data of Murthy (1967, p. 228) about 80 factories in a region. Here, we define the following variables:

y : out put of factory in thousands of rupees,
x : number of workers in the factory, and
z : fixed capital of factory in thousands of rupees.

In general, the information regarding z variable is available in advance so the population is arbitrarily divided in to four different strata with respect to the values of z variable. The first, second, third and fourth strata consist population units for which $z \leq 500$, $500 < z \leq 1,000$, $1,000 < z \leq 2,000$ and $z > 2,000$ respectively. The stratum size, requisite parameters and sample sizes for the four strata are given in the Table 5.1. The amount of gain in efficiency is given in the Table 5.2.

TABLE 5.1 *Parameters of population*

Stratum number	*Stratum size*	*Parameters of stratum*				*Sample sizes*	
h	N_h	S_{yh}^2	$\rho_{yx}^{(h)}$	$\rho_{yz}^{(h)}$	$\rho_{xz}^{(h)}$	n'_h	n_h
1	19	573183.5	0.813815	0.936386	0.904402	11	5
2	32	447730.5	0.888286	0.925965	0.845634	17	8
3	14	173150.2	0.929553	0.983494	0.936638	8	3
4	15	416912.8	0.978722	0.969232	0.945385	9	4

TABLE 5.2 *Relative efficiencies of estimators*

Estimator	*Relative efficiency of an estimator*
$\widehat{\overline{y}}_{st}$	100.00
$\widehat{\overline{y}}_{rdS}$	236.59
$\widehat{\overline{y}}_{rdzS}$	534.61
$\widehat{\overline{y}}_{efst}^{(reg)}$	**1056.83**

From the Table 5.2, we can see that the proposed stratified regression type estimator has a big gain in efficiency as compared to that of the existing ones.

Thus, in practice, we recommend the use of the proposed stratified regression type estimator for estimating the population mean.

ACKNOWLEDGEMENTS

The author is thankful to the referees whose constructive suggestions have improved the quality of the paper.

REFERENCES

AHMED, M. S. (1998). "A note on regression-type estimators using multiple auxiliary information", *Australian & New Zealand Journal of Statistics*, **40**, 373–376.

FULLER, W. A. (2002). "Regression estimation for survey samples", *Survey Methodology*, **28**, 5–23.

MUKERJEE, R., RAO, T. J. AND VIJAYAN, K. (1987). "Regression type estimators using multiple auxiliary information", *The Australian Journal of Statistics*, **29**, 244–254.

MUKERJEE, R., RAO, T. J. AND VIJAYAN, K. (2000). Rejoinder to M. S. Ahmed : "A note on regression-type estimators using multiple auxiliary information", *Australian & New Zealand Journal of Statistics*, **42**, 245.

MURTY, M. N. (1967). *Sampling Theory and Methods*, Statistical Publishing Society, Calcutta.

PRASAD, B., SINGH, R. S. AND SINGH H. P. (2002). "Modified chain ratio estimators for finite population mean using two auxiliary variables in double sampling", *Statistics in Transition*, **5**, 1051–1066.

ROY, D. C. (2003). "A regression-type estimator in two-phase sampling using two auxiliary variables", *Pakistan Journal of Statistics*, **19**, 281–290.

SAHOO, J. AND BALA, M. K. (2000). "A note on the estimation of population mean in stratified random sampling using two auxiliary variables", *Biometrical Journal*, **42**, 87–92.

SAHOO, J., SAHOO, L. N. AND MOHANTY, S. (1993). "A regression approach to estimation in two-phase sampling using two auxiliary variables", *Current Science*, **65**, 73–75.

SRIVASTAVA, S. K. AND JHAJJ, H. S. (1983). "A class of estimators of the population mean using multi-auxiliary information", *Calcutta Statistical Association Bulletin*, **32**, 47–56.

SUKHATME, P. V. AND SUKHATME, B. V. (1970). *Sampling Theory of Surveys with Applications*, Asia Publishing House, Bombay.

TRIPATHI, T. P. AND AHMED, M. S. (1995). "A class of estimators for a finite population mean based on multivariate information and general two phase sampling", *Calcutta Statistical Association Bulletin*, **45**, 203–218.