Journal of Applied Statistics
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713428038

On improvement in estimating population mean in stratified random sampling

Nursel Koyuncu [a];Cem Kadilar [a]
[a] Department of Statistics, Hacettepe University, Ankara, Turkey

Online publication date: 11 May 2010

To cite this Article Koyuncu, Nursel andKadilar, Cem(2010) 'On improvement in estimating population mean in stratified random sampling', Journal of Applied Statistics, 37: 6, 999 — 1013

To link to this Article: DOI: 10.1080/02664760903002675

URL: http://dx.doi.org/10.1080/02664760903002675

Journal of Applied Statistics
Vol. 37, No. 6, June 2010, 999–1013

On improvement in estimating population mean in stratified random sampling

Nursel Koyuncu* and Cem Kadilar

Hacettepe University, Department of Statistics, Beytepe, Ankara, Turkey

(*Received 30 June 2008; final version received 27 April 2009*)

Gupta and Shabbir [2] have suggested an alternative form of ratio-type estimators for estimating the population mean. In this paper, we obtained a corrected version for the mean square error (MSE) of the Gupta–Shabbir estimator, up to first order of approximation, and the optimum case is discussed. We expand this estimator to the stratified random sampling and propose general classes for combined and separate estimators. Also an empirical study is carried out to show the properties of the proposed estimators.

Keywords: ratio estimator; auxiliary information; mean square error; efficiency; stratified random sampling

Mathematics Subject Classification: Primary: 62D05

1. Introduction

A ratio estimator is commonly used when the study variable Y is highly correlated with the auxiliary variable X. When the population mean $\bar{X}$ of the auxiliary variable is known, a number of modified versions of ratio estimators have been suggested by various authors. Further, many authors used some population parameters of the auxiliary variable to improve the precision of ratio estimators such as Sisodia and Dwivedi [11], Upadhyaya and Singh [13], Singh and Tailor [7] and others. Gupta and Shabbir [2] have suggested a general class of ratio estimators when the population parameters of the auxiliary variable are known. In addition to these studies, Kadilar and Cingi [3], Shabbir and Gupta [6], Singh *et al*. [8], Singh and Vishwakarma [10], Koyuncu and Kadilar [5] extended the estimators suggested in a simple random sampling to the stratified random sampling. In this study, we derive the correct expression of the MSE in Gupta and Shabbir [2] and suggest similar estimators in the stratified random sampling.

*Corresponding author. Email: nkoyuncu@hacettepe.edu.tr

ISSN 0266-4763 print/ISSN 1360-0532 online

DOI: 10.1080/02664760903002675
http://www.informaworld.com

Consider a finite population of size N from which a sample s of size n is drawn according to the simple random sampling without replacement. Let y_i and x_i be the values of the study and the auxiliary variables on the ith unit ($i = 1, 2, \ldots, N$). Further, let $\bar{y}$ and $\bar{x}$ be the sample means of the study and auxiliary variables, respectively.

To obtain the bias and MSE, let us define $\delta_y = (\bar{y} - \bar{Y})/\bar{Y}$ and $\delta_x = (\bar{x} - \bar{X})/\bar{X}$. Using these notations, we have

$$E(\delta_y) = E(\delta_x) = 0, \quad E(\delta_y^2) = \gamma C_y^2, \quad E(\delta_x^2) = \gamma C_x^2, \quad E(\delta_y \delta_x) = \gamma C_{yx} = \gamma \rho C_y C_x,$$

where

$$C_y^2 = \frac{S_y^2}{\bar{Y}^2}, \quad C_x^2 = \frac{S_x^2}{\bar{X}^2}, \quad C_{yx} = \frac{S_{yx}}{\overline{YX}}, \quad S_y^2 = \frac{\sum_{i=1}^{N}(y_i - \bar{Y})^2}{N-1}, \quad S_x^2 = \frac{\sum_{i=1}^{N}(x_i - \bar{X})^2}{N-1},$$

$$S_{yx} = \frac{\sum_{i=1}^{N}(y_i - \bar{Y})(x_i - \bar{Y})}{N-1} \quad \text{and} \quad \gamma = \frac{N-n}{Nn}.$$

2. Estimators in the simple random sampling

For estimating the population mean $\bar{Y}$, a regression type estimator can be given in the following way when w_1 and w_2 are constants that have no restriction:

$$\bar{y}_{p(0)} = w_1 \bar{y} + w_2 \left(\bar{X} - \bar{x}\right). \tag{1}$$

The bias and MSE of the estimator in Equation (1) are respectively given by

$$B\left(\bar{y}_{p(0)}\right) = (w_1 - 1)\,\bar{Y}, \tag{2}$$

$$\text{MSE}\left(\bar{y}_{p(0)}\right) = (w_1 - 1)^2\,\bar{Y}^2 + w_1^2 \bar{Y}^2 \gamma C_y^2 + w_2^2 \bar{X}^2 \gamma C_x^2 - 2 w_1 w_2 \overline{XY} \gamma C_{yx}. \tag{3}$$

Optimum values of w_1 and w_2 are, respectively, given by

$$w_1^* = \frac{C_x^2}{C_x^2 + \gamma C_x^2 C_y^2 - \gamma C_{yx}^2}, \quad w_2^* = \frac{\bar{Y} C_{yx}}{\bar{X} C_x^2\left(1 + \gamma C_y^2\right) - \bar{X} \gamma C_{yx}^2}. \tag{4}$$

Substituting these optimum values in Equation (3), the minimum MSE of $\bar{y}_{p(0)}$ is given by

$$\text{MSE}\left(\bar{y}_{p(0)}\right)_{\min} = \bar{Y}^2 \frac{\gamma\, C_y^2\left(1 - \rho^2\right)}{1 + \gamma\, C_y^2\left(1 - \rho^2\right)} = \frac{\text{MSE}\left(\bar{y}_{\text{reg}}\right)}{1 + \left(\text{MSE}\left(\bar{y}_{\text{reg}}\right)/\bar{Y}^2\right)}. \tag{5}$$

When w_1 and w_2 are any constants and η and λ are either constants or functions of known parameters, the ratio type estimator suggested by Gupta and Shabbir [2] is given by

$$\bar{y}_p = [w_1 \bar{y} + w_2(\bar{X} - \bar{x})]\left(\frac{\eta\,\bar{X} + \lambda}{\eta\,\bar{x} + \lambda}\right). \tag{6}$$

It is worth mentioning that when values of w_1, w_2, η and λ are conveniently chosen, many common estimators can be obtained such as the classical ratio estimator $\bar{y}_0$, the regression type estimator $\bar{y}_{p(0)}$, the estimators suggested by Singh and Tailor [7], Sisodia and Dwivedi [11], Upadhyaya and Singh [13] etc. In addition to these estimators, some new estimators, which are also generated from Equation (6), are given in the Table 1.

Expressing Equation (6) in terms of δ_i, we have

$$\bar{y}_p = \left[w_1\bar{Y} + w_1\bar{Y}\delta_y - w_2\bar{X}\delta_x\right](1+\tau\delta_x)^{-1}, \tag{7}$$

where $\tau = \eta\bar{X}/(\eta\bar{X}+\lambda)$. We expand the terms of Equation (7) up to the first order of approximation and we have

$$B\left(\bar{y}_p\right) = (w_1 - 1)\bar{Y} + \gamma\left[w_1\bar{Y}\left(\tau^2C_x^2 - \tau C_{yx}\right) + w_2\bar{X}\tau C_x^2\right], \tag{8}$$

$$\begin{aligned}\text{MSE}\left(\bar{y}_p\right) &= (w_1-1)^2\bar{Y}^2 + w_1^2\bar{Y}^2\gamma\left(C_y^2 - 4\tau C_{yx} + 3\tau^2C_x^2\right) + w_2^2\bar{X}^2\gamma C_x^2 \\ &\quad - 2w_1\bar{Y}^2\gamma\left(\tau^2C_x^2 - \tau C_{yx}\right) - 2\overline{YX}w_2\tau\gamma C_x^2 - 2\overline{YX}w_2w_1\gamma\left(C_{yx} - 2\tau C_x^2\right)\end{aligned} \tag{9}$$

and optimum values of w_1 and w_2 are, respectively, found as

$$w_1^* = \frac{1-\gamma\tau^2C_x^2}{1+\gamma C_y^2 - \gamma\rho^2C_y^2 - \gamma\tau^2C_x^2}, \quad w_2^* = \frac{\bar{Y}}{\bar{X}}\left(\tau + \frac{\left(1-\gamma\tau^2C_x^2\right)\left(C_{yx} - 2\tau C_x^2\right)}{C_x^2 + \gamma C_x^2C_y^2 - \gamma C_{yx}^2 - \gamma\tau^2C_x^4}\right). \tag{10}$$

Substituting these optimum values in Equation (9), the minimum MSE of $\bar{y}_p$ can be written as follows:

$$\text{MSE}\left(\bar{y}_p\right)_{\min} = \frac{\left(1 - C_x^2\tau^2\gamma\right)\text{MSE}\left(\bar{y}_{\text{reg}}\right)}{\left(1 - C_x^2\tau^2\gamma\right) + \left(\text{MSE}\left(\bar{y}_{\text{reg}}\right)/\bar{Y}^2\right)}. \tag{11}$$

From Equation (11), it is clear that the values of η and λ affect the minimum MSE of $\bar{y}_p$. Note that the optimum choice of the constants w_1 and w_2 involves unknown parameters. These quantities can be guessed quite accurately through a pilot sample survey or sample data or experience gathered in due course of time.

3. Efficiency comparisons in the simple random sampling

In this section, we will compare the Gupta and Shabbir [2] estimator with the classical estimators by using the corrected MSE in Equation (11).

$$\begin{gathered}\text{MSE}\left(\bar{y}_{p(0)}\right)_{\min} - \text{MSE}\left(\bar{y}_p\right)_{\min} > 0, \\ \frac{1}{\bar{Y}^2 + \text{MSE}\left(\bar{y}_{\text{reg}}\right)} - \frac{\left(1 - C_x^2\tau^2\gamma\right)}{\left(1-\gamma\tau^2C_x^2\right)\bar{Y}^2 + \text{MSE}\left(\bar{y}_{\text{reg}}\right)} > 0.\end{gathered} \tag{12}$$

If the condition (12) is satisfied, $\bar{y}_p$ is more efficient than $\bar{y}_{p(0)}$.

$$\begin{gathered}\text{MSE}\left(\bar{y}_{\text{reg}}\right) - \text{MSE}\left(\bar{y}_p\right)_{\min} > 0, \\ 1 - \frac{\left(1-\gamma\tau^2C_x^2\right)}{\left(1-\gamma\tau^2C_x^2\right) + \left(\text{MSE}\left(\bar{y}_{\text{reg}}\right)/\bar{Y}^2\right)} > 0.\end{gathered} \tag{13}$$

If the condition (13) is satisfied, $\bar{y}_p$ is more efficient than $\bar{y}_{\text{reg}}$. Note that $\bar{y}_{p(0)}$ is a member of $\bar{y}_p$, and condition (13) is always satisfied for $\bar{y}_{p(0)}$. So we can say that this estimator is always more efficient than $\bar{y}_{\text{reg}}$.

4. Estimators in the stratified random sampling

4.1 *Notations*

Let the population of size N be stratified into L strata with hth stratum containing N_h units, where $h = 1, 2, \ldots, L$ such that $\sum_{h=1}^{L} N_h = N$. A simple random sample of size n_h is drawn without replacement from the hth stratum such that $\sum_{h=1}^{L} n_h = n$. Let (y_{hi}, x_{hi}) denote the observed values of Y and X on the ith unit of the hth stratum, where $i = 1, 2, \ldots, N_h$ and $h = 1, 2, \ldots, L$. Moreover, assume that $\bar{y}_h = \sum_{i=1}^{n_h} y_{hi}/n_h$, $\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h$, and $\bar{Y}_h = \sum_{i=1}^{N_h} Y_{hi}/N_h$, $\bar{Y} = \sum_{h=1}^{L} W_h \bar{Y}_h$ be the sample and population means of Y, respectively, where $W_h = N_h/N$ is the stratum weight. Similar expressions for X can also be defined.

4.2 *General class of estimators*

Following Srivastava [12], a general combined class in the stratified random sampling is defined by

$$t_c = g\left(\bar{y}_{st}, u_{st}\right), \tag{14}$$

where $u_{st} = \bar{x}_{st}/\bar{X}$, and $g(\bar{y}_{st}, u_{st})$ is a function of $\bar{y}_{st}$ and u_{st}. To study the properties of t_c, we assume following regularity conditions:

(1) The point $(\bar{y}_{st}, u_{st})$ assumes the value in a closed convex subset R_2 of a two-dimensional real space containing the point $(\bar{Y}, 1)$,
(2) The function $g(\bar{y}_{st}, u_{st})$ is continuous and bounded in R_2,
(3) $g(\bar{Y}, 1) = \bar{Y}$ and $g_0(\bar{Y}, 1) = 1$, where $g_0(\bar{Y}, 1)$ denotes the first-order partial derivative of g with respect to $\bar{y}_{st}$,
(4) The first- and second-order partial derivatives of $g(\bar{y}_{st}, u_{st})$ exist and are continuous and bounded in R_2.

Expanding $g(\bar{y}_{st}, u_{st})$ about the point $(\bar{Y}, 1)$ in a second-order Taylor series and using the above regularity conditions, we have

$$t_c = \bar{y}_{st} + (u_{st} - 1)\, g_1 + (u_{st} - 1)^2\, g_2 + \left(\bar{y}_{st} - \bar{Y}\right)\, (u_{st} - 1)\, g_3 + \left(\bar{y}_{st} - \bar{Y}\right)^2 g_4, \tag{15}$$

where

$$g_1 = \left.\frac{\partial g}{\partial u_{st}}\right|_{\bar{y}_{st}=\bar{Y}, u_{st}=1}, \quad g_2 = \frac{1}{2}\left.\frac{\partial^2 g}{\partial u_{st}^2}\right|_{\bar{y}_{st}=\bar{Y}, u_{st}=1}, \quad g_3 = \frac{1}{2}\left.\frac{\partial^2 g}{\partial \bar{y}_{st}\partial u_{st}}\right|_{\bar{y}_{st}=\bar{Y}, u_{st}=1},$$
$$g_4 = \frac{1}{2}\left.\frac{\partial^2 g}{\partial \bar{y}_{st}^2}\right|_{\bar{y}_{st}=\bar{Y}, u_{st}=1}.$$

To obtain the bias and the MSE, let us define $\xi_0 = \left(\bar{y}_{st} - \bar{Y}\right)/\bar{Y}$ and $\xi_1 = \left(\bar{x}_{st} - \bar{X}\right)/\bar{X}$. Using these notations,

$$E(\xi_0) = E(\xi_1) = 0,$$
$$V_{r,s} = \sum_{h=1}^{L} W_h^{r+s} \frac{E\lfloor(\bar{x}_h - \bar{X}_h)^r (\bar{y}_h - \bar{Y}_h)^s\rfloor}{\bar{X}^r \bar{Y}^s}. \tag{16}$$

From Equation (16), we can write

$$E(\xi_0^2) = \frac{\sum_{h=1}^{L} W_h^2 \gamma_h S_{yh}^2}{\bar{Y}^2} = V_{0,2}, \quad E\left(\xi_1^2\right) = \frac{\sum_{h=1}^{L} W_h^2 \gamma_h S_{xh}^2}{\bar{X}^2} = V_{2,0},$$
$$E(\xi_0 \xi_1) = \frac{\sum_{h=1}^{L} W_h^2 \gamma_h S_{xyh}}{\bar{X}\bar{Y}} = V_{1,1}, \quad \text{and} \quad \gamma_h = \frac{N_h - n_h}{N_h n_h}.$$

The class of estimators t_c, in terms of ξ_0 and ξ_1, can be written as,

$$t_c - \bar{Y} = \bar{Y}\varepsilon_0 + \varepsilon_1 g_1 + \varepsilon_1^2 g_2 + \bar{Y}\varepsilon_0\varepsilon_1 g_3 + \bar{Y}^2\varepsilon_0^2 g_4, \tag{17}$$

and the bias and the MSE of t_c are, respectively, given by

$$B(t_c) = V_{2,0}g_2 + \bar{Y}V_{1,1}g_3 + \bar{Y}^2 V_{0,2}g_4, \tag{18}$$

$$\text{MSE}(t_c) = \bar{Y}^2 V_{0,2} + V_{2,0}g_1^2 + 2\bar{Y}V_{1,1}g_1. \tag{19}$$

By using the optimal value of $g_1^* = -\bar{Y}V_{1,1}/V_{2,0}$, the minimum MSE of the estimators in the class t_c is found as

$$\text{MSE}(t_c)_{\min} = \bar{Y}^2 V_{0,2}\left[1 - \frac{V_{1,1}^2}{V_{2,0}V_{0,2}}\right], \tag{20}$$

which is also the MSE of combined regression type estimators.

Similarly, for separate estimators we can define a general separate class in the stratified random sampling as

$$t_s = \sum_{h=1}^{L} W_h t_{sh}, \tag{21}$$

where $t_{sh} = g_h(\bar{y}_h, u_h)$. Here, $u_h = \bar{x}_h/\bar{X}_h$ and $g_h(\bar{y}_h, u_h)$ are functions of $\bar{y}_h$ and u_h.

To obtain the bias and MSE, let us define $\xi_{0h} = (\bar{y}_h - \bar{Y}_h)/\bar{Y}_h$ and $\xi_{1h} = (\bar{x}_h - \bar{X}_h)/\bar{X}_h$. Using these notations,

$$E(\xi_{0h}) = E(\xi_{1h}) = 0, \quad E(\xi_{1h}^2) = \gamma_h C_{xh}^2, \quad E(\xi_{0h}^2) = \gamma_h C_{yh}^2, \quad E(\xi_{0h}\xi_{1h}) = \gamma_h C_{xyh},$$

where

$$C_{yh}^2 = \frac{S_{yh}^2}{\bar{Y}_h^2}, \quad C_x^2 = \frac{S_{xh}^2}{\bar{X}_h^2}, \quad C_{yx} = \frac{S_{yxh}}{\bar{Y}_h\bar{X}_h}, \quad S_{yh}^2 = \frac{\sum_{i=1}^{N}(y_{hi} - \bar{Y}_h)^2}{N_h - 1},$$

$$S_{xh}^2 = \frac{\sum_{i=1}^{N}(x_{hi} - \bar{X}_h)^2}{N_h - 1}, \quad S_{yxh} = \frac{\sum_{i=1}^{N}(y_{hi} - \bar{Y}_h)(x_{hi} - \bar{Y}_h)}{N_h - 1}$$

and applying the same procedure of the combined class of estimators by adapting the regularity conditions for the hth stratum, the bias and the MSE of estimators in the class t_s are, respectively, found as follows:

$$B\,(t_s) = \sum_{h=1}^{L} W_h\gamma_h\left[C_{xh}^2 g_{2h} + \bar{Y}_h C_{yxh}g_{3h} + \bar{Y}_h^2 C_{yh}^2 g_{4h}\right], \tag{22}$$

$$\text{MSE}\,(t_s) = \sum_{h=1}^{L} W_h^2\gamma_h\left[\bar{Y}^2 C_{yh}^2 + C_{xh}^2 g_{1h}^2 + 2\bar{Y}C_{xyh}g_{1h}\right], \tag{23}$$

where

$$g_{1h} = \left.\frac{\partial g_h}{\partial u_h}\right|_{\bar{y}_h=\bar{Y}_h, u_h=1}, \quad g_{2h} = \frac{1}{2}\left.\frac{\partial^2 g_h}{\partial u_h^2}\right|_{\bar{y}_h=\bar{Y}_h, u_h=1}, \quad g_{3h} = \frac{1}{2}\left.\frac{\partial^2 g_h}{\partial \bar{y}_h\partial u_h}\right|_{\bar{y}_h=\bar{Y}_h, u_h=1},$$

$$g_{4h} = \frac{1}{2}\left.\frac{\partial^2 g_h}{\partial \bar{y}_h^2}\right|_{\bar{y}_h=\bar{Y}_h, u_h=1}.$$

By using the optimal value of $g_{1h}^{*} = -\bar{Y}C_{xyh}/C_{xh}^{2}$, the minimum MSE of the estimators in the class t_s is found as

$$\mathrm{MSE}(t_s)_{\min} = \sum_{h=1}^{L} W_h^2 \lambda_h S_{hy}^2 (1-\rho_h^2), \tag{24}$$

which is also the MSE of the separate regression type estimator. Note that t_c and t_s contain many estimators in the literature such as Singh and Vishwakarma [9], Kaur [4] etc. We can claim that the minimum MSE of any subclass of t_c or t_s cannot be reduced further than the regression estimator.

4.3 *More general class of estimators*

Let $\underline{u} = (u_1, u_2, \ldots, u_p)$, where $u_i = \hat{\theta}_{i(st)}/\theta_{i(st)}$ $i = 1, \ldots, p$ assume values in a bounded closed, convex subset R_p of p-dimensional space containing the point $\underline{\varepsilon} = (1, 1, \ldots 1)$. Here $\hat{\theta}_{i(st)} = \sum_{h=1}^{L} W_h \hat{\theta}_{i(h)}$, $\theta_{i(st)} = \sum_{h=1}^{L} W_h \theta_{i(h)}$. $\theta_{i(h)}$ is any real numbers or the function of the known parameters of the auxiliary variable X for hth stratum and $\hat{\theta}_{i(h)}$ is a consistent estimator of $\theta_{i(h)}$. To obtain the MSE equation, let us define $\underline{\xi} = \underline{u} - \underline{\varepsilon} = \lfloor \xi_1, \ldots, \xi_p \rfloor$.
Using these notations,

$$E(\xi_0) = E(\xi_i) = 0 \quad \text{where } i = 1, 2, \ldots, p,$$

$$E\left(\xi_0 \underline{\xi}\right) = \underline{b} = [V_{100\ldots01}\ V_{010\ldots01}\ \ldots\ V_{000\ldots11}],$$

$$E\left(\underline{\xi}'\underline{\xi}\right) = \underline{A} = \begin{bmatrix} V_{200\ldots00} & V_{110\ldots00} & \cdots & V_{100\ldots10} \\ V_{110\ldots00} & V_{020\ldots00} & \ldots & V_{010\ldots10} \\ \vdots & \vdots & \ddots & \vdots \\ V_{100\ldots10} & V_{010\ldots10} & \ldots & V_{000\ldots20} \end{bmatrix},$$

where

$$V_{r,s,\ldots,t,v} = \sum_{h=1}^{L} W_h^{r+s+\cdots+t+v} \frac{E[(\hat{\theta}_{1(h)} - \theta_{1(h)})^r (\hat{\theta}_{2(h)} - \theta_{2(h)})^s \cdots (\hat{\theta}_{p(h)} - \theta_{p(h)})^t (\bar{y}_h - \bar{Y}_h)^v]}{\theta_{1(st)}^r \theta_{2(st)}^s \ldots \theta_{p(st)}^t \bar{Y}^v}.$$

Now we suggest a more general class of estimators using more auxiliary information as

$$t_{nc} = H(\bar{y}_{st}, \underline{u}), \tag{25}$$

where $H(\bar{y}_{st}, \underline{u})$ is a function of $\bar{y}_{st}$ and u_i, $i = 1, \ldots, p$, such that $H(\bar{Y}, \underline{\varepsilon}) = \bar{Y}$ and H is bounded and continuous with bounded and continuous first- and second-order partial derivatives in R^{p+1}. This general class also contains the estimator suggested by Koyuncu and Kadilar [5].

Expanding $H(\bar{y}_{st}, \underline{u})$ about the point $(\bar{Y}, \underline{\varepsilon})$ in a second-order Taylor's series and further assuming $\partial H/\partial \bar{y}_{st}|_{\bar{y}_{st}=\bar{Y}, \underline{u}=\underline{\varepsilon}} = 1$, we obtain

$$t_{nc} - \bar{Y} = \bar{Y}\xi_0 + (\underline{u} - \underline{\varepsilon})\underline{H_1} + (\underline{u} - \underline{\varepsilon})\underline{H_2}(\underline{u} - \underline{\varepsilon})' + (\bar{y}_{st} - \bar{Y})^2 \underline{H_3} + (\bar{y}_{st} - \bar{Y})(\underline{u} - \underline{\varepsilon})\underline{H_4}, \tag{26}$$

where

$$\underline{H_1} = \left.\frac{\partial H}{\partial \underline{u}}\right|_{\bar{y}_{st}=\bar{Y},\underline{u}=\underline{\varepsilon}}, \quad \underline{H_2} = \frac{1}{2}\left.\frac{\partial^2 H}{\partial \underline{u}\underline{u}'}\right|_{\bar{y}_{st}=\bar{Y},\underline{u}=\underline{\varepsilon}}, \quad \underline{H_3} = \frac{1}{2}\left.\frac{\partial^2 H}{\partial \bar{y}_{st}^2}\right|_{\bar{y}_{st}=\bar{Y},\underline{u}=\underline{\varepsilon}},$$
$$\underline{H_4} = \frac{1}{2}\left.\frac{\partial^2 H}{\partial \bar{y}_{st}\partial \underline{u}}\right|_{\bar{y}_{st}=\bar{Y},\underline{u}=\underline{\varepsilon}}$$

and the MSE of t_{nc} is given by

$$\mathrm{MSE}(t_{nc}) = [\bar{Y}^2 V_{0,0,\ldots,0,2} + \underline{H_1}'\underline{A}\underline{H_1} + 2\bar{Y}\underline{b}\underline{H_1}]. \tag{27}$$

By using the optimal value of $\underline{H_1}^* = -\bar{Y}\underline{A}^{-1}\underline{b}'$, the minimum MSE of the estimators in the class t_{nc} is found as

$$\mathrm{MSE}(t_{nc})_{\min} = \bar{Y}^2 V_{0,0,\ldots,0,2}\left[1 - \frac{\underline{b}\underline{A}^{-1}\underline{b}'}{V_{0,0,\ldots,0,2}}\right]. \tag{28}$$

Remark If we define one of the u_i as

$$\frac{a_{st}^*}{A_{st}^*} = \frac{\sum_{h=1}^{L} W_h \bar{x}_h A_h}{\sum_{h=1}^{L} W_h \bar{X}_h A_h},$$

we can write $V^*_{r,s,\ldots t,k,v}$ instead of $V_{r,s,\ldots t,v}$ in $\underline{A}$ matrix and $\underline{b}$ vector. Here

$$V^*_{r,s,\ldots,t,k,v} = \sum_{h=1}^{L} W_h^{r+s+\cdots+t+v} A_h^k \frac{E[(\hat{\theta}_{1(h)} - \theta_{1(h)})^r(\hat{\theta}_{2(h)} - \theta_{2(h)})^s \cdots (\hat{\theta}_{(p-1)(h)} - \theta_{(p-1)(h)})^t(\bar{x}_h - \bar{X}_h)^k(\bar{y}_h - \bar{Y}_h)^v]}{\theta^r_{1(st)}\theta^s_{2(st)}\cdots\theta^t_{(p-1)(st)}A_{st}^{*k}\bar{Y}^v}.$$

For separate estimators, let us define $\underline{u_{(h)}} = (u_{1(h)}, u_{2(h)}, \ldots, u_{p(h)})$, where $u_{i(h)} = \hat{\theta}_{i(h)}/\theta_{i(h)}$ $i = 1, \ldots, p$ assume values in a bounded closed, convex subset R_p of p-dimensional space containing the point $\underline{\varepsilon} = (1, 1, \ldots 1)$. Let $\underline{\xi_{(h)}} = \underline{u_{(h)}} - \underline{\varepsilon} = \lfloor \xi_{1h}, \xi_{2h}, \ldots, \xi_{ph} \rfloor$.

Using these notations,

$$E(\xi_{0h}) = E(\xi_{ih}) = 0, \quad \text{where} \quad i = 1, 2, \ldots, p,$$
$$E(\xi_{0h}\underline{\xi_{(h)}}) = \underline{b_{(h)}} = [V_{100\ldots01(h)}\ V_{010\ldots01(h)}\ \ldots\ V_{000\ldots11(h)}],$$
$$E(\underline{\xi_{(h)}}'\underline{\xi_{(h)}}) = \underline{A_{(h)}} = \begin{bmatrix} V_{200\ldots00(h)} & V_{110\ldots00(h)} & \cdots & V_{100\ldots10(h)} \\ V_{110\ldots00(h)} & V_{020\ldots00(h)} & \cdots & V_{010\ldots10(h)} \\ \vdots & \vdots & \ddots & \vdots \\ V_{100\ldots10(h)} & V_{010\ldots10(h)} & \cdots & V_{000\ldots20(h)} \end{bmatrix},$$

where

$$V_{r,s,\ldots,t,v(h)} = \frac{E[(\hat{\theta}_{1(h)} - \theta_{1(h)})^r(\hat{\theta}_{2(h)} - \theta_{2(h)})^s \cdots (\hat{\theta}_{p(h)} - \theta_{p(h)})^t(\bar{y}_h - \bar{Y}_h)^v]}{\theta^r_{1(h)}\theta^s_{2(h)}\ldots\theta^t_{p(h)}\bar{Y}^v_{(h)}}.$$

A more general separate estimator is given by

$$t_{ns} = \sum_{h=1}^{L} W_h t_{n(h)}, \tag{29}$$

where $t_{n(h)} = H_{(h)}(\bar{y}_{(h)}, \underline{u_{(h)}})$, $H_{(h)}(\bar{y}_{(h)}, \underline{u_{(h)}})$ is a function of $\bar{y}_{(h)}$ and $u_{i(h)}$ $\quad i = 1, \ldots, p$, such that $H_{(h)}(\bar{Y}_h, \underline{\varepsilon}) = \bar{Y}_h$ and $H_{(h)}$ is bounded and continuous with bounded and continuous first- and second-order partial derivatives in R^{p+1}. Expanding $H_{(h)}(\bar{y}_{(h)}, \underline{u_{(h)}})$ about the point $(\bar{Y}_h, \underline{\varepsilon})$ in a second-order Taylor's series and further assuming $\partial H/\partial \bar{y}_h|_{\bar{y}_h=\bar{Y}_h, \underline{u_{(h)}}=\underline{\varepsilon}} = 1$, we obtain

$$\begin{aligned} t_{n(h)} - \bar{Y}_h = \bar{Y}_h \xi_{0h} + (\underline{u_{(h)}} - \underline{\varepsilon})\underline{H_{1(h)}} + (\underline{u_{(h)}} - \underline{\varepsilon})\underline{H_{2(h)}}(\underline{u_{(h)}} - \underline{\varepsilon})' + (\bar{y}_h - \bar{Y}_h)^2 \underline{H_{3(h)}} \\ + (\bar{y}_h - \bar{Y}_h)(\underline{u_{(h)}} - \underline{\varepsilon})\underline{H_{4(h)}}, \end{aligned} \tag{30}$$

where

$$\underline{H_{1(h)}} = \frac{\partial H_{(h)}}{\partial \underline{u}}|_{\bar{y}_h=\bar{Y}_h, \underline{u_{(h)}}=\underline{\varepsilon}}, \quad \underline{H_{2(h)}} = \frac{\partial^2 H_{(h)}}{\partial \underline{u_{(h)} u_{(h)}}'}|_{\bar{y}_h=\bar{Y}_h, \underline{u_{(h)}}=\underline{\varepsilon}}, \quad \underline{H_{3(h)}} = \frac{1}{2}\frac{\partial^2 H_{(h)}}{\partial \bar{y}_h^2}|_{\bar{y}_h=\bar{Y}_h, \underline{u_{(h)}}=\underline{\varepsilon}},$$

$$\underline{H_{4(h)}} = \frac{1}{2}\frac{\partial^2 H}{\partial \bar{y}_h \partial \underline{u_{(h)}}}|_{\bar{y}_h=\bar{Y}_h, \underline{u_{(h)}}=\underline{\varepsilon}}$$

and the MSE of $t_{n(h)}$ is given by

$$\mathrm{MSE}(t_{n(h)}) = [\bar{Y}_h^2 V_{0,0,\ldots,0,2(h)} + \underline{H_{1(h)}}'\underline{A_{(h)} H_{1(h)}} + 2\bar{Y}_h \underline{b_{(h)} H_{1(h)}}]. \tag{31}$$

By using the optimal value of $\underline{H_{1(h)}}^* = -\bar{Y}_h \underline{A_{(h)}^{-1}}\underline{b_{(h)}}'$, the minimum MSE of the estimators in the class t_{ns} is found as

$$\mathrm{MSE}(t_{ns})_{\min} = \sum_{h=1}^{L} W_h^2 \bar{Y}_h^2 V_{0,0,\ldots,0,2(h)} \left[1 - \frac{\underline{b_{(h)}}\underline{A_{(h)}^{-1}}\underline{b_{(h)}}'}{V_{0,0,\ldots,0,2(h)}}\right]. \tag{32}$$

Note that the sections of general class of estimators and more general class of estimators contain estimators satisfying regularity conditions. In the next section, we will propose some estimators, which do not satisfy some of the regularity conditions.

5. Suggested combined estimators in the stratified random sampling

In this section, using the prior value of certain population parameter(s) of the auxiliary variable and following Gupta and Shabbir [2], we suggest some classes in the stratified random sampling and study their properties considering the affect of the population parameter(s) to the MSE. In the stratified random sampling, the combined version of the estimator suggested by Gupta and Shabbir [2] can be given by:

$$\bar{y}_{pc} = [w_1 \bar{y}_{st} + w_2(\bar{X} - \bar{x}_{st})]\left(\frac{\eta_{st}\bar{X} + \lambda_{st}}{\eta_{st}\bar{x}_{st} + \lambda_{st}}\right), \tag{33}$$

where η_{st} and λ_{st} are either real numbers or the functions of the known parameters of the auxiliary variable, X, such as $\lambda_1 = \sum_{h=1}^{L} W_h S_{xh}$, $\lambda_2 = \sum_{h=1}^{L} W_h C_{xh}$, $\lambda_3 = \sum_{h=1}^{L} W_h \beta_{1h}(x)$, $\lambda_4 = \sum_{h=1}^{L} W_h \beta_{2h}(x)$ and $\lambda_5 = \sum_{h=1}^{L} W_h \rho_h$. Some new estimators, which are generated from Equation (33) are given in Table 1. The Bias and MSE of $\bar{y}_{pc}$ are, respectively, given by

$$B(\bar{y}_{pc}) = (w_1 - 1)\bar{Y} - w_1 \bar{Y}\tau_{st} V_{1,1} + w_2 \bar{X}\tau_{st} V_{2,0} + w_1 \bar{Y}\tau_{st}^2 V_{2,0}, \quad (34)$$

$$\begin{aligned} \mathrm{MSE}(\bar{y}_{pc}) = {} & \bar{Y}^2 + w_2^2 \bar{X}^2 V_{2,0} + w_1^2 \bar{Y}^2 (V_{0,2} + 1 + 3\tau_{st}^2 V_{2,0} - 4\tau_{st} V_{1,1}) - 2w_2 \overline{YX}\tau_{st} V_{2,0} \\ & - 2w_1 \bar{Y}^2 (\tau_{st}^2 V_{2,0} + 1 - \tau_{st} V_{1,1}) + 2w_2 w_1 \overline{YX}(2\tau_{st} V_{2,0} - V_{1,1}), \end{aligned} \quad (35)$$

Table 1. Some members of suggested estimators.

Simple random sampling					
Separate estimators					
	Some members of $\bar{y}_p$			Some members of $\bar{y}_{ps}$	
	η	λ		η_h	λ_h
$\bar{y}_{p(0)}$	0	1	$\bar{y}_{ps(0)}$	0	1
$\bar{y}_{p(1)}$	1	ρ	$\bar{y}_{ps(1)}$	1	ρ_h
$\bar{y}_{p(2)}$	1	C_x	$\bar{y}_{ps(2)}$	1	C_{xh}
$\bar{y}_{p(3)}$	1	$\beta_{2(x)}$	$\bar{y}_{ps(3)}$	1	$\beta_{2h}(x)$
$\bar{y}_{p(4)}$	$\beta_{2(x)}$	C_x	$\bar{y}_{ps(4)}$	$\beta_{2h}(x)$	C_{xh}
$\bar{y}_{p(5)}$	C_x	$\beta_{2(x)}$	$\bar{y}_{ps(5)}$	C_{xh}	$\beta_{2h}(x)$
Combined estimators					
	Some members of $\bar{y}_{pc}$			Some members of $\bar{y}_{pc}^*$	
	η_{st}	λ_{st}		A_h	λ_{st}
$\bar{y}_{pc(0)}$	0	1	$\bar{y}_{pc(1)}^*$	C_{xh}	$\sum_{h=1}^{L} W_h \beta_{2h}(x)$
$\bar{y}_{pc(1)}$	1	$\sum_{h=1}^{L} W_h \rho_h$	$\bar{y}_{pc(2)}^*$	C_{xh}	$\sum_{h=1}^{L} W_h \rho_h$
$\bar{y}_{pc(2)}$	1	$\sum_{h=1}^{L} W_h C_{xh}$	$\bar{y}_{pc(3)}^*$	ρ_h	$\sum_{h=1}^{L} W_h \beta_{2h}(x)$
$\bar{y}_{pc(3)}$	1	$\sum_{h=1}^{L} W_h \beta_{2h}(x)$	$\bar{y}_{pc(4)}^*$	ρ_h	$\sum_{h=1}^{L} W_h C_{xh}$
$\bar{y}_{pc(4)}$	$\sum_{h=1}^{L} W_h \beta_{2h}(x)$	$\sum_{h=1}^{L} W_h C_{xh}$	$\bar{y}_{pc(5)}^*$	$\beta_{1h}(x)$	$\sum_{h=1}^{L} W_h \beta_{2h}(x)$
$\bar{y}_{pc(5)}$	$\sum_{h=1}^{L} W_h C_{xh}$	$\sum_{h=1}^{L} W_h \beta_{2h}(x)$	$\bar{y}_{pc(6)}^*$	$\beta_{1h}(x)$	$\sum_{h=1}^{L} W_h \rho_h$
			$\bar{y}_{pc(7)}^*$	$\beta_{2h}(x)$	$\sum_{h=1}^{L} W_h C_{xh}$
			$\bar{y}_{pc(8)}^*$	$\beta_{2h}(x)$	$\sum_{h=1}^{L} W_h \rho_h$

where $\tau_{st} = \eta_{st}\bar{X}/(\eta_{st}\bar{X} + \lambda_{st})$. We obtain the optimum values as

$$w_1^* = \frac{V_{2,0}(1 - \tau_{st}^2 V_{2,0})}{(V_{0,2}V_{2,0} + V_{2,0} - \tau_{st}^2 V_{2,0}^2 - V_{1,1}^2)}, \quad w_2^* = \frac{\bar{Y}}{\bar{X}}\left(\tau_{st} + \frac{(1 - \tau_{st}^2 V_{2,0})(V_{1,1} - 2\tau_{st}V_{2,0})}{(V_{0,2}V_{2,0} + V_{2,0} - \tau_{st}^2 V_{2,0}^2 - V_{1,1}^2)}\right). \tag{36}$$

Replacing these optimum values in Equation (35), we have

$$\mathrm{MSE}(\bar{y}_{pc})_{\min} = \frac{\mathrm{MSE}(\bar{y}_{\mathrm{reg}(st)})(1 - \tau_{st}^2 V_{2,0})}{(\mathrm{MSE}(\bar{y}_{\mathrm{reg}(st)})/\bar{Y}^2) + (1 - \tau_{st}^2 V_{2,0})}. \tag{37}$$

Motivated by the combined version of the estimator in Gupta and Shabbir [2], we propose a new family as

$$\bar{y}_{pc}^* = [w_1\bar{y}_{st} + w_2(\bar{X} - \bar{x}_{st})]\ \left(\frac{A_{st}^* + \lambda_{st}}{a_{st}^* + \lambda_{st}}\right), \tag{38}$$

where $A_{st}^* = \sum_{h=1}^{L} W_h\bar{X}_h A_h$, $a_{st}^* = \sum_{h=1}^{L} W_h\bar{x}_h A_h$ and A_h may be some population information of the auxiliary variable for the hth stratum such as S_{xh}, coefficient of variation C_{xh}, skewness $\beta_{1(x)h}$, kurtosis $\beta_{2(x)h}$, correlation coefficient $\rho_{h(xy)}$. Some new estimators that are generated from Equation (38) are given in Table 1. To obtain the MSE, we should define new ξ and V terms as

$$\xi_1^* = \frac{a_{st}^* - A_{st}^*}{A_{st}^*} = \frac{\sum_{h=1}^{L} W_h A_h \left(\bar{x}_h - \bar{X}_h\right)}{A_{st}^*}, \tag{39}$$

$$V_{s,t,r}^* = \frac{\sum_{h=1}^{L} W_h^2 (A_h)^t E\left[\left(\bar{x}_h - \bar{X}_h\right)^{s+t}\left(\bar{y}_h - \bar{Y}_h\right)^r\right]}{\bar{X}^{s+t}\bar{Y}^r}, \tag{40}$$

respectively. Using the notations in Equations (39) and (40), we obtain the bias and the MSE of $\bar{y}_{pc}^*$, respectively, as follows:

$$B(\bar{y}_{pc}^*) = (w_1 - 1)\bar{Y} - w_1\bar{Y}\kappa_{st}V_{011}^* + w_2\bar{X}\kappa_{st}V_{110}^* + w_1\bar{Y}\kappa_{st}^2 V_{020}^*, \tag{41}$$

$$\mathrm{MSE}(\bar{y}_{pc}^*) = w_1^2\bar{Y}^2 A + w_2^2\bar{X}^2 V_{200}^* + \bar{Y}^2 - 2w_1w_2\overline{YX}B + 2w_1\bar{Y}^2 C - 2w_2\overline{XY}\kappa_{st}V_{110}^*, \tag{42}$$

where

$$\kappa_{st} = \frac{\bar{X}}{A_{st}^* + \lambda_{st}}, \quad A = 1 + V_{002}^* - 4\kappa_{st}V_{011}^* + 3\kappa_{st}^2 V_{020}^*, \quad B = V_{101}^* - 2\kappa_{st}V_{110}^*,$$

$$C = \kappa_{st}V_{011}^* - \kappa_{st}^2 V_{020}^* - 1.$$

The optimum values of w_1 and w_2 are found as

$$w_1^* = \frac{\kappa_{st}V_{110}^* B - CV_{200}^*}{AV_{200}^* - B^2}, \quad w_2^* = \frac{\bar{Y}(\kappa_{st}V_{110}^* A - BC)}{\bar{X}(V_{200}^* A - B^2)}. \tag{43}$$

Replacing these optimum values in Equation (42), the minimum MSE of $\bar{y}_{pc}^*$ can be written as follows:

$$\mathrm{MSE}(\bar{y}_{pc}^*)_{\min} = \bar{Y}^2\left[1 - \frac{(\kappa_{st}^2 V_{110}^{*2} A + C^2 V_{200}^* - 2\kappa_{st}V_{110}^* BC)}{(AV_{200}^* - B^2)}\right]. \tag{44}$$

Note that $\bar{y}_{pc}$ and $\bar{y}_{pc}^*$ are not members of t_{nc} because the conditions of $H(\bar{Y}, \underline{\varepsilon}) = \bar{Y}$ and $\partial H/\partial\bar{y}_{st}|_{\bar{y}_{st}=\bar{Y},\underline{u}=\underline{\varepsilon}} = 1$ are not satisfied for these estimators.

Note that various transformations of the auxiliary variable affect the minimum values of MSEs for $\bar{y}_{pc}$ and $\bar{y}^*_{pc}$. Therefore, the minimum MSEs of $\bar{y}_{pc}$ and $\bar{y}^*_{pc}$ are different from the MSE of the regression estimator and change according to the choice of the auxiliary information, whereas, in the classes of t_c and t_p, there is one minimum MSE, which is equal to the MSE of the regression estimator.

6. Efficiency comparisons for combined estimators

In this section, we compare the efficiencies for the combined estimators as follows:

$$\mathrm{MSE}(\bar{y}_{pc(0)})_{\min} - \mathrm{MSE}(\bar{y}_{pc})_{\min} > 0,$$

$$\frac{1}{\bar{Y}^2 + \mathrm{MSE}(\bar{y}_{\mathrm{reg}(st)})} - \frac{1 - \tau_{st}^2 V_{2,0}}{\mathrm{MSE}(\bar{y}_{\mathrm{reg}(st)}) + \bar{Y}^2(1 - \tau_{st}^2 V_{2,0})} > 0. \tag{45}$$

If the condition (45) is satisfied, then $\bar{y}_{pc}$ is more efficient than $\bar{y}_{pc(0)}$ whose properties are shown in Table 1.

$$\mathrm{MSE}(\bar{y}_{\mathrm{reg}(st)}) - \mathrm{MSE}(\bar{y}_{pc})_{\min} > 0,$$

$$1 - \frac{1 - \tau_{st}^2 V_{2,0}}{(\mathrm{MSE}(\bar{y}_{\mathrm{reg}(st)})/\bar{Y}^2) + (1 - \tau_{st}^2 V_{2,0})} > 0. \tag{46}$$

If the condition (46) is satisfied, then $\bar{y}_{pc}$ is more efficient than $\bar{y}_{\mathrm{reg}(st)}$. Note that $\bar{y}_{pc(0)}$ is a member of $\bar{y}_{pc}$ and condition (46) is always satisfied for $\bar{y}_{pc(0)}$. So we can say that this estimator is always more efficient than $\bar{y}_{\mathrm{reg}(st)}$.

$$\mathrm{MSE}(\bar{y}_{pc(0)}) - \mathrm{MSE}(\bar{y}^*_{pc})_{\min} > 0,$$

$$\frac{\kappa_{st}^2 V_{110}^{*2} A + C^2 V_{200}^* - 2\kappa_{st} V_{110}^* BC}{A V_{200}^* - B^2} - \frac{V_{2,0}}{V_{2,0} + V_{0,2} V_{2,0} - V_{1,1}^2} > 0. \tag{47}$$

If the condition (47) is satisfied, then $\bar{y}^*_c$ is more efficient than $\bar{y}_{pc(0)}$.

$$\mathrm{MSE}(\bar{y}_{\mathrm{reg}(st)}) - \mathrm{MSE}(\bar{y}^*_{pc})_{\min} > 0,$$

$$V_{0,2}[1 - \rho_{st}^2] - \left[1 - \frac{(\kappa_{st}^2 V_{110}^{*2} A + C^2 V_{200}^* - 2\kappa_{st} V_{110}^* BC)}{(A V_{200}^* - B^2)}\right] > 0. \tag{48}$$

If the condition (49) is satisfied, then $\bar{y}^*_{pc}$ is more efficient than $\bar{y}_{\mathrm{reg}(st)}$.

$$\mathrm{MSE}(\bar{y}_{pc(0)})_{\min} - \mathrm{MSE}(\bar{y}^*_{pc})_{\min} > 0,$$

$$\frac{\kappa_{st}^2 V_{110}^{*2} A + C^2 V_{200}^* - 2\kappa_{st} V_{110}^* BC}{A V_{200}^* - B^2} - V_{2,0}\tau_{st}^2 - \frac{V_{2,0}(1 - \tau_{st}^2 V_{2,0})^2}{V_{0,2} V_{2,0} + V_{2,0} - \tau_{st}^2 V_{2,0}^2 - V_{1,1}^2} > 0. \tag{49}$$

If the condition (49) is satisfied, then $\bar{y}^*_{pc}$ is more efficient than $\bar{y}_{pc(0)}$.

Suggested separate estimators in the stratified random sampling

Separate version of the estimator, suggested by Gupta and Shabbir [2], can be given by:

$$\bar{y}_{ps} = \sum_{h=1}^{L} W_h[w_{1h}\bar{y}_h + w_{2h}(\bar{X}_h - \bar{x}_h)]\left(\frac{\eta_h \bar{X}_h + \lambda_h}{\eta_h \bar{x}_h + \lambda_h}\right), \tag{50}$$

where η_h and λ_h are either real numbers or the functions of the known parameters of the auxiliary variable for the hth stratum. Applying the standard techniques mentioned in the previous section, we get the bias and the MSE as

$$B(\bar{y}_{ps}) = \sum_{h=1}^{L} W_h[(w_{1h}-1)\bar{Y}_h + w_{1h}\bar{Y}_h\gamma_h(\tau_h^2 C_{xh}^2 - \tau_h C_{xyh}) + w_{2h}\bar{X}_h\tau_h\gamma_h C_{xh}^2], \quad (51)$$

$$\begin{aligned} \mathrm{MSE}(\bar{y}_{ps}) = \sum_{h=1}^{L} W_h^2 \{ & \bar{Y}_h^2(w_{1h}-1)^2 + w_{1h}^2\bar{Y}_h^2\gamma_h(C_{yh}^2 - 4\tau_h C_{xyh} + 3\tau_h^2 C_{xh}^2) \\ & + w_{2h}^2\bar{X}_h^2\gamma_h C_{xh}^2 - 2w_{1h}\bar{Y}_h^2\gamma_h(\tau_h^2 C_{xh}^2 - \tau_h C_{xyh}) - 2w_{2h}\bar{X}_h\bar{Y}_h\tau_h\gamma_h C_{xh}^2 \\ & -2w_{1h}w_{2h}\bar{X}_h\bar{Y}_h\gamma_h(C_{xyh} - 2\tau_h C_{xh}^2)\}, \end{aligned} \quad (52)$$

respectively, where $\tau_h = \eta_h\bar{X}_h/(\eta_h\bar{X}_h + \lambda_h)$ and the optimum values of w_{1h} and w_{2h} are, respectively, found as

$$w_{1h}^* = \frac{1-\gamma_h\tau_h^2 C_{xh}^2}{1+\gamma_h C_{yh}^2 - \gamma_h\rho_h^2 C_{yh}^2 - \gamma_h\tau_h^2 C_{xh}^2}, \quad (53)$$

$$w_{2h}^* = \frac{\bar{Y}_h}{\bar{X}_h}\left[\tau_h + \frac{(1-\gamma_h\tau_h^2 C_{xh}^2)(C_{xyh} - 2\tau_h C_{xh}^2)}{C_{xh}^2 + \gamma_h C_{yh}^2 C_{xh}^2 - \gamma_h C_{xyh}^2 - \tau_h^2\gamma_h C_{xh}^4}\right]. \quad (54)$$

Using Equations (53) and (54) in Equation (52), the minimum MSE of $\bar{y}_{ps}$ can be written as follows:

$$\mathrm{MSE}(\bar{y}_{ps})_{\min} = \sum_{h=1}^{L} W_h^2\bar{Y}_h^2 \frac{\gamma_h C_{yh}^2(1-\rho_{xyh}^2)(1-\gamma_h C_{xh}^2\tau_h^2)}{\gamma_h C_{yh}^2(1-\rho_{xyh}^2) + (1-\gamma_h C_{xh}^2\tau_h^2)}. \quad (55)$$

Note that we do not consider the definition A_{st}^* for separate estimators. If we use this definition in Equation (50), $\bar{y}_{ps}$ changes as

$$\bar{y}_{ps}^{+} = \sum_{h=1}^{L} W_h[w_{1h}\bar{y}_h + w_{2h}(\bar{X}_h - \bar{x}_h)]\left(\frac{A_{st}^* + \lambda_h}{a_{st}^* + \lambda_h}\right), \quad (56)$$

so we cannot say this is a separate estimator exactly and its MSE is worse than the MSE of $\bar{y}_{ps}$.

However, we can think a new estimator as

$$\bar{y}_{ps}^{*} = \sum_{h=1}^{L} W_h[w_{1h}\bar{y}_h + w_{2h}(\bar{X}_h - \bar{x}_h)]\left(\frac{W_h\eta_h\bar{X}_h + \lambda_h}{W_h\eta_h\bar{x}_h + \lambda_h}\right). \quad (57)$$

In this case we use $\tau_h^* = W_h\eta_h\bar{X}_h/(W_h\eta_h\bar{X}_h + \lambda_h)$ instead of $\tau_h = \eta_h\bar{X}_h/(\eta_h\bar{X}_h + \lambda_h)$ in Equations (51)–(55), respectively. Note that we can think $\bar{y}_{ps}^*$ is a subclass of $\bar{y}_{ps}$ because if we define η_h as $\eta_h^* = W_h\eta_h$ as a function of known auxiliary information in Equation (50) we can get Equation (57).

7. Efficiency comparisons for separate estimators

In this section, we compare the efficiencies for the combined estimators as follows:

$$\mathrm{MSE}(\bar{y}_{ps(0)})_{\min} - \mathrm{MSE}(\bar{y}_{ps})_{\min} > 0,$$

$$\frac{1}{1+\gamma_h C_{yh}^2(1-\rho_{xyh}^2)} - \frac{1-\gamma_h C_{xh}^2\tau_h^2}{\gamma_h C_{yh}^2(1-\rho_{xyh}^2) + 1 - \tau_h^2\gamma_h C_{xh}^2} > 0. \quad (58)$$

If the condition (58) is satisfied, then $\bar{y}_{ps}$ is more efficient than $\bar{y}_{ps(0)}$ whose properties are shown in Table 1.

$$\mathrm{MSE}(\bar{y}_{(\mathrm{reg})s}) - \mathrm{MSE}(\bar{y}_{ps})_{\min} > 0,$$

$$1 - \frac{1 - \gamma_h C_{xh}^2 \tau_h^2}{\gamma_h C_{yh}^2 (1 - \rho_{xyh}^2) + 1 - \tau_h^2 \gamma_h C_{xh}^2} > 0. \tag{59}$$

If the condition (59) is satisfied, then $\bar{y}_{ps}$ is more efficient than $\bar{y}_{(\mathrm{reg})s}$. Note that $\bar{y}_{ps(0)}$ is a member of $\bar{y}_{ps}$ and condition (59) is always satisfied for $\bar{y}_{ps(0)}$. So we can say that this estimator is always more efficient than $\bar{y}_{(\mathrm{reg})s}$.

8. Numerical examples

8.1 *Simple random sampling*

For numerical comparisons of estimators in the simple random sampling, we consider the following data set used by Gupta and Shabbir [2]:

$$N = 200, \quad n = 50, \quad \bar{Y} = 500, \quad \bar{X} = 25, \quad C_y = 15, \quad C_x = 2, \quad \rho = 0.90,$$
$$\beta_{2(x)} = 50, \quad \gamma = 0.015.$$

Although Gupta and Shabbir [2] claim that various transformations of the auxiliary variable do not affect the value of the minimum MSE, we show that the specific values of η and λ play a role on the minimum MSE. For this reason, we decide to calculate the minimum MSE values of $\bar{y}_p$ using different values of η and λ as shown in Table 1. The minimum MSE values for the members of $\bar{y}_p$, given in Table 1, have been obtained using Equations (11). Besides MSEs of the classical ratio estimator $\bar{y}_0$ and the regression estimator $\bar{y}_{\mathrm{reg}}$ have been obtained. These values are given in Table 3. From Table 3, we observe that $\bar{y}_{p(1)}$ is the most efficient estimator for the simple random sampling data. As a result, $\bar{y}_{p(i)}$ $(i = 1, 2, 3, 4, 5)$ are more efficient than $\bar{y}_0$, $\bar{y}_{p(0)}$, and $\bar{y}_{\mathrm{reg}}$.

8.2 *Stratified random sampling*

We use the data concerning the number of teachers as the study variable and the number of students as the auxiliary variable in both primary and secondary schools for 923 districts at six regions (as 1, Marmara; 2, Agean; 3, Mediterranean; 4, Central Anatolia; 5, Black Sea; 6, East and Southeast Anatolia) in Turkey in 2007 (Source: Ministry of Education, Republic of Turkey). The summary statistics of the data are given in Table 2. We used the Neyman allocation method for determining the sample sizes of each stratum [1].

The minimum MSE values for the members of $\bar{y}_{pc}$, $\bar{y}_{pc}^*$ and $\bar{y}_{ps}$, given in Table 3, have been obtained using Equations (37), (44) and (55), respectively. In addition to these MSE values, MSE values for the estimators of the classical combined ratio $\bar{y}_{0(c)}$, the classical separate ratio $\bar{y}_{0(s)}$, the combined regression $\bar{y}_{\mathrm{reg}(c)}$, and the separate regression $\bar{y}_{\mathrm{reg}(s)}$ have also been computed. All of these MSE values are given in Table 3. From Table 3, we can infer that separate estimators are more efficient than combined estimators and that $\bar{y}_{ps(2)}$ is the most efficient in the separate estimators for this data set. When we further examine Table 3, we see that the differences among the MSE values for the members of $\bar{y}_{pc}^*$ are big, whereas these differences for the members of $\bar{y}_{ps}$ and $\bar{y}_{pc}$ are small. As a result, $\bar{y}_{pc(4)}^*$ is the most efficient estimator in the combined estimators and it is also more efficient than the combined regression estimator, $\bar{y}_{\mathrm{reg}(c)}$. From this result, we can say that the best efficiency is obtained when A_h and λ_{st} are defined as ρ_h and $\sum_{h=1}^{L} W_h C_{xh}$,

Table 2. Data statistics (data for the stratified random sampling).

$N_1 = 127$	$N_2 = 117$	$N_3 = 103$
$N_4 = 170$	$N_5 = 205$	$N_6 = 201$
$n_1 = 31$	$n_2 = 21$	$n_3 = 29$
$n_4 = 38$	$n_5 = 22$	$n_6 = 39$
$\bar{Y}_1 = 703.74$	$\bar{Y}_2 = 413$	$\bar{Y}_3 = 573.17$
$\bar{Y}_4 = 424.66$	$\bar{Y}_5 = 267.03$	$\bar{Y}_6 = 393.84$
$C_{y1} = 1.256$	$C_{y2} = 1.562$	$C_{y3} = 1.803$
$C_{y4} = 1.909$	$C_{y5} = 1.512$	$C_{y6} = 1.807$
$\bar{X}_1 = 20804.59$	$\bar{X}_2 = 9211.79$	$\bar{X}_3 = 14309.30$
$\bar{X}_4 = 9478.85$	$\bar{X}_5 = 5569.95$	$\bar{X}_6 = 12997.59$
$C_{X1} = 1.465$	$C_{X2} = 1.648$	$C_{X3} = 1.925$
$C_{X4} = 1.922$	$C_{X5} = 1.526$	$C_{X6} = 1.777$
$S_{xy1} = 25237153.52$	$S_{xy2} = 9747942.85$	$S_{xy3} = 28294397.04$
$S_{xy4} = 14523885.53$	$S_{xy5} = 3393591.75$	$S_{xy6} = 15864573.97$
$\rho_{st1} = 0.936$	$\rho_{st2} = 0.996$	$\rho_{st3} = 0.994$
$\rho_{st4} = 0.983$	$\rho_{st5} = 0.989$	$\rho_{st6} = 0.965$
$\beta_2(x_1) = 4.593$	$\beta_2(x_2) = 18.543$	$\beta_2(x_3) = 15.446$
$\beta_2(x_4) = 10.162$	$\beta_2(x_5) = 21.947$	$\beta_2(x_6) = 23.114$
$w_1 = 0.138$	$w_2 = 0.127$	$w_3 = 0.112$
$w_4 = 0.184$	$w_5 = 0.222$	$w_6 = 0.218$

Table 3. MSE values of proposed estimators.

Simple random sampling			
Separate estimators			
$\bar{y}_p$	MSE	$\bar{y}_{ps}$	MSE
$\bar{y}_{p(0)}$	97677.08	$\bar{y}_{ps(0)}$	106.0068178
$\bar{y}_{p(1)}$	95468.42[a]	$\bar{y}_{ps(1)}$	105.9795401
$\bar{y}_{p(2)}$	95842.14	$\bar{y}_{ps(2)}$	105.9785024[a]
$\bar{y}_{p(3)}$	97673.69	$\bar{y}_{ps(3)}$	105.9796034
$\bar{y}_{p(4)}$	95505.23	$\bar{y}_{ps(4)}$	105.9795366
$\bar{y}_{p(5)}$	96642.58	$\bar{y}_{ps(5)}$	105.9795754
$\bar{y}_{\text{reg}}$	160312.5	$\bar{y}_{\text{reg}(s)}$	106.4267643
$\bar{y}_0$	656250	$\bar{y}_{0(s)}$	128.8116162
Combined estimators			
$\bar{y}_{pc}$	MSE	$\bar{y}^*_{pc}$	MSE
$\bar{y}_{pc(0)}$	194.0853	$\bar{y}^*_{pc(1)}$	185.9231308
$\bar{y}_{pc(1)}$	194.0826725	$\bar{y}^*_{pc(2)}$	183.6392752
$\bar{y}_{pc(2)}$	194.0826728	$\bar{y}^*_{pc(3)}$	177.6944494
$\bar{y}_{pc(3)}$	194.0826796	$\bar{y}^*_{pc(4)}$	177.6742894[a]
$\bar{y}_{pc(4)}$	194.0826721[a]	$\bar{y}^*_{pc(5)}$	291.4648988
$\bar{y}_{pc(5)}$	194.0826765	$\bar{y}^*_{pc(6)}$	271.8398998
$\bar{y}_{\text{reg}(c)}$	194.2832	$\bar{y}^*_{pc(7)}$	714.9849344
$\bar{y}_{0(c)}$	216.4183	$\bar{y}^*_{pc(8)}$	692.3880483

Notes: [a]The most efficient estimator.

respectively, for this data set. Note that we get more efficient estimators when we define A_h as the correlation coefficient in the family of estimators, $\bar{y}^*_{pc}$. We conclude that the minimum MSE values of the $\bar{y}_p$, $\bar{y}_{ps}$, $\bar{y}_{pc}$ and $\bar{y}^*_{pc}$ can change according to the definition of known population parameter(s).

9. Conclusion

In this paper, the properties of the regression type and the Gupta and Shabbir [2] estimators are discussed in the simple random sampling and we adapt these estimators for the stratified random sampling. Moreover, we consider a new class for combined estimators and examine the effects of various transformations of the auxiliary information on the families of estimators.

We show that $\bar{y}_{p(0)}$, $\bar{y}_{pc(0)}$ and $\bar{y}_{ps(0)}$ are always more efficient than the classical regression estimator in the simple random sampling, the classical combined regression estimator, and the classical separate regression estimators, respectively. We also show that the suggested estimators $\bar{y}_{pc}$ and $\bar{y}^*_{pc}$ in the stratified random sampling are more efficient than the classical combined regression estimator when the conditions of Equation (46) and (48) are satisfied, respectively.

Although Gupta and Shabbir [2] claim that various transformations of the auxiliary variable do not affect the minimum value of MSE, we prove that the specific values of η and λ have an important role on the minimum MSE. We also prove that this important result is also valid for the stratified random sampling.

References

[1] W.G. Cochran, *Sampling Techniques*, John Wiley and Sons, New York, 1977.

[2] S. Gupta and J. Shabbir, *On improvement in estimating the population mean in simple random sampling*, J. Appl. Stat. 35(5) (2008), pp. 559–566.

[3] C. Kadilar and H. Cingi, *Ratio estimators in stratified random sampling*, Biom. J. 45 (2003), pp. 218–225.

[4] P. Kaur, *On the estimation of population mean in stratified sampling*, Biom. J. 27(1) (1985), pp. 101–105.

[5] N. Koyuncu and C. Kadilar, *Ratio and product estimators in stratified random sampling*, J. Statist. Plann. Inference 139(8) (2009), pp. 2552–2558.

[6] J. Shabbir and S. Gupta, *Improved ratio estimators in stratified sampling*, Amer. J. Math. Management Sci. 25 (2005), pp. 293–311.

[7] H.P. Singh and R. Tailor, *Use of known correlation coefficient in estimating the finite population mean*, Stat. Transit. 6 (2003), pp. 555–560.

[8] H.P. Singh, R. Tailor, S. Singh, and J.M. Kim, *A modified estimator of population mean using power transformation*, Statist. Papers 49 (2008), pp. 37–58.

[9] H.P. Singh and G.K. Vishwakarma, *An efficient variant of the product and ratio estimators in stratified random sampling*, Stat. Transit. 7(6) (2006), pp. 1311–1325.

[10] H.P. Singh and G.K. Vishwakarma, *A family of estimators of population mean using auxiliary information in stratified sampling*, Comm. Statist. Theory Methods 37(7) (2008), pp. 1038–1050.

[11] B.V.S. Sisodia and V.K. Dwivedi, *A modified ratio estimator using coefficient of variation of auxiliary variable*, J. Indian Soc. Agricultural Statist. 33 (1981), pp. 13–18.

[12] S.K. Srivastava, *A generalized estimator for mean of a finite population using multi-auxiliary information*, JASA 66 (1971), pp.404–407.

[13] L.N. Upadhyaya and H.P. Singh, *Use of transformed auxiliary variable in estimating the finite population mean*, Biom. J. 41(5) (1999), pp. 627–636.