A Generalized Class of Estimators for Estimating Population Mean in the Presence of Non-Response

Iram Saleem[1], Aamir Sanaullah[2], and Muhammad Hanif[1]⋆

[1]Department of Statistics, National College of Business Administration and Economics, Lahore, Pakistan
[2]Department of Statistics, COMSATS Institute of Information and Technology, Lahore, Pakistan.

ARTICLE INFO

ABSTRACT

Koyuncu and Kadilar proposed an estimator based on single auxiliary variable with complete response in stratified random sampling. In this paper, we extended Koyuncu and Kadilar's estimator to a more generalized class of estimators using two-auxiliary variables in stratified random sampling for the situation of non-response and further introduced its another improved generalized class of estimators. The mathematical conditions under which proposed class of estimators are efficient as compare to Hansen and Hurwtiz estimator, and ratio estimators modified for stratified sampling have been derived. An empirical study has also been carried out to examine the performance of the suggested estimators.

1. INTRODUCTION

Non-response refers to the situation, when an investigator fails to get necessary information from some of the units of the selected sample. The problem of non-response was first analyzed by Hansen and Hurwitz [1]. They developed a classical non-response concept to obtain information from the sub-sample of non-response group. An estimator for the population mean in the presence of non-response was constructed and also derived its variance with the optimum sampling fraction for the non-respondents. It is suitable for the surveys, in which first attempt is made on mail questionnaires and second attempt is selected from the non-respondent persons by personnel interviews. Following Hansen–Hurwitz methodology Cochran [2] proposed ratio type estimator for dealing with non-response. Chaudhary *et al.* [3], Haq and Shabbir [4] and Sanaullah *et al.* [5] presented some improved estimators for stratified random sampling under non-response.

In order to improve the efficiency of an estimator, auxiliary information is often used to estimate the unknown population mean of study variable. Cochran [6] discussed classical ratio estimators. Further, Cochran [2], Kadilar and Cingi [7], Shabbir and Gupta [8], Koyuncu and Kadilar [9], Sanaullah *et al.* [10] and Sanaullah *et al.* [11] utilized auxiliary information under stratified random sampling scheme.

2. SAMPLING DESIGN AND PROCEDURE DEALING WITH NON-RESPONSE

Consider a finite population of size N is stratified into L homogenous strata. Let N_h be the size of hth stratum (h=1,2,3,..., L) such that $\sum_{h=1}^{L} N_h = N$ and (y_{hi}, x_{hi}, z_{hi}) be the observations of study variable y and auxiliary variables x and z on the ith unit of hth stratum, respectively. Let $\bar{y}_h$, $\bar{x}_h$ and $\bar{z}_h$ be the sample means of hth stratum corresponding to the population means $\overline{Y}_h$, $\overline{X}_h$ and $\overline{Z}_h$ respectively. Usually it is

⋆Corresponding author. Email: dmianhanif@gmail.com

not possible to collect complete information from all the units selected in the sample $n_h \left(\sum_{h=1}^{L}, n_h, =, n \right)$. Let $n_{h(1)}$ units from a sample of

nh provide their responses and $n_{h(2)}$ units do not. Adapting Hansen and Hurwitz [1] sub-sampling methodology, a sub-sample of size r_h ($r_h = \frac{n_{h(2)}}{f_h}$; $fh>1$) from $n_{h(2)}$ non-respondents group is selected at random and $1/fh$ denotes the sampling fraction among the non-respondent group in the hth stratum. In practice, r_h is usually not integer and has to be rounded. In accordance with most of the current literature on the topic, let us assume that the followed-up r_h ($\subset n_{h(2)}$) units respond on the second call. Moreover, let Uh denote a dummy variable which takes value uhi on the ith population unit of stratum h and has mean $\overline{U}_h$. Hereafter, $\overline{U}_h$ may stand for Yh; Xh or for a second auxiliary variable Zh. Let;

$$\overline{u}_{n_{h(1)}} = \frac{\sum_{i=1}^{n_{h(1)}} u_{i(1)}}{n_{h(1)}}, \quad \overline{u}_{r_h} = \frac{\sum_{i=1}^{r_{(1)}} u_{i(2)}}{r_h}$$

and

$$\overline{u}_h^* = \frac{n_{h_{(1)}}}{n_h} \overline{u}_{(1)n_{h_1}} + \frac{n_{h_{(2)}}}{n_h} \overline{u}_{(2)r_h} \tag{1}$$

where $\overline{u}_{(1)n_{h(1)}}$ is mean of $n_{h(1)}$ respondents on first call and $\overline{u}_{(2)r_h}$ is mean of r_h units respond on the second call, and $\overline{u}_h^*$ denotes the unbiased Hansen–Hurwitz [1] of $\overline{U}_h$ for stratum h.

A modified Hansen and Hurwitz [1] unbiased estimator for stratified sampling may be given as,

$$t_1 = \sum_{h=1}^{L} P_h \overline{u}_h^*, \tag{2}$$

The variance of t_1 is,

$$\mathrm{Var}\,(t_1) = \sum_{h=1}^{L} \lambda_h P_h^2 S_{u_h}^2 + \sum_{h=1}^{L} \lambda_h^* P_h^2 S_{u_{h(2)}}^2. \tag{3}$$

where $S_{uh}^2 = \sum_{i=1}^{N_h} \left(u_{hi} - \overline{U}_h \right)^2 / (N_h - 1)$, and $S_{uh(2)}^2 = \sum_{i=1}^{N_{h(2)}} \left(u_i - \overline{U}_{h(2)} \right)^2 / \left(N_{h(2)} - 1 \right)$ are the mean square error of the entire group and the

non-response group of the study variable with $P_h = N_h/N$, $\lambda_h = \left(\frac{1}{n_h} - \frac{1}{N_h} \right)$, $\lambda_h^* = \left(\frac{f_h - 1}{n_h} \right) W_{h(2)}$, $W_{h(2)} = N_{h(2)}/N_h$, and $f_h = n_{h(2)}/r_h$.

The modified form of ratio and product estimators from stratified random sampling under non-response defined by Cochran [2] may be written as:

$$t_2 = \overline{y}_{st}^* \left[\frac{\overline{X}}{\overline{x}_{st}^*} \right] \left[\frac{\overline{Z}}{\overline{z}_{st}^*} \right], \tag{4}$$

and

$$t_3 = \overline{y}_{st}^* \left[\frac{\overline{x}_{st}^*}{\overline{X}} \right] \left[\frac{\overline{z}_{st}^*}{\overline{Z}} \right], \tag{5}$$

where $\overline{y}_{st}^*$ and $\overline{x}_{st}^*$ are Hansen–Hurwitz [1] estimators modified to the stratified sampling for population means $\overline{X}$ and $\overline{Y}$ respectively.

The MSE of the estimators t_2 and t_3 are given respectively as

$$\mathrm{MSE}\,(t_2) \approx \overline{Y}^2 \sum_{h=1}^{L} \left[\lambda_h \left(\frac{S_{hy}^2}{\overline{Y}^2} + \frac{S_{hx}^2}{\overline{X}^2} + \frac{S_{hz}^2}{\overline{Z}^2} - 2\frac{S_{hxy}}{\overline{YX}} - 2\frac{S_{hyz}}{\overline{YZ}} + 2\frac{S_{hyz}}{\overline{XZ}} \right) + \right.$$

$$\left. \lambda_h^* \left(\frac{S_{hy(2)}^2}{\overline{Y}^2} + \frac{S_{hx(2)}^2}{\overline{X}^2} + \frac{S_{hz(2)}^2}{\overline{Z}^2} - 2\frac{S_{hxy(2)}}{\overline{YX}} - 2\frac{S_{hyz(2)}}{\overline{YZ}} + 2\frac{S_{hxz(2)}}{\overline{XZ}} \right) \right], \tag{6}$$

and

$$\mathrm{MSE}\,(t_3) \approx \overline{Y}^2 \sum_{h=1}^{L} \lambda_h \left(\frac{S_{hy}^2}{\overline{Y}^2} + \frac{S_{hx}^2}{\overline{X}^2} + \frac{S_{hz}^2}{\overline{Z}^2} + 2\frac{S_{hxy}}{\overline{YX}} + 2\frac{S_{hyz}}{\overline{YZ}} + 2\frac{S_{hyz}}{\overline{XZ}} \right) +$$

$$\lambda_h^* \left(\frac{S_{hy(2)}^2}{\overline{Y}^2} + \frac{S_{hx(2)}^2}{\overline{X}^2} + \frac{S_{hz(2)}^2}{\overline{Z}^2} + 2\frac{S_{hxy(2)}}{\overline{YX}} + 2\frac{S_{hyz(2)}}{\overline{YZ}} + 2\frac{S_{hxz(2)}}{\overline{XZ}} \right) \right]. \tag{7}$$

Chaudhary *et al.* [3] presented ratio estimator for stratified sampling when non-response is present only on study variable as,

$$t_4 = \overline{y}_{st}^* \left[\frac{a\overline{X} + b}{\alpha \left(a\overline{x}_{st} + b \right) + (1 - \alpha) \left(a\overline{X} + b \right)} \right]^g, \tag{8}$$

The MSE of t_3 is

$$\mathrm{MSE}\,(t_4) \approx \sum_{h=1}^{L} P_h^2 \lambda_h \left[S_{yh}^2 + \alpha^2 v^2 g^2 R^2 S_{xh}^2 - 2\alpha v g R \rho_{xyh} S_{xh} S_{yh} \right] + \sum_{h=1}^{L} P_h^2 \lambda_h^* S_{yh(2)}^2, \tag{9}$$

where

$$v = \frac{a\overline{X}}{a\overline{X} + b}, R = \frac{\overline{Y}}{\overline{X}},$$

The MSE(t_5) is minimum for

$$\alpha = \frac{\displaystyle\sum_{h=1}^{L} P_h^2 \lambda_h \rho_{xyh} S_{yh} S_{xh}}{v g R \displaystyle\sum_{h=1}^{L} P_h^2 \lambda_h S_{xh}}. \tag{10}$$

The aim of this paper is to propose a more generalized class of estimators for estimating population mean considering the non-response in stratified random sampling using two auxiliary variables. Also the purpose is to introduce another improved form of proposed generalised estimator. Another purpose is to determine the optimum size of the sample and the sub-sampling fractions of the non-respondent group for the fixed cost.

3. PROPOSED GENERALIZED CLASS OF ESTIMATORS

In this section, following Koyuncu and Kadilar [12] we proposed a generalized class of estimators for estimating a finite population mean in stratified random sampling considering the presence of non-response using two auxiliary variables as,

$$t_a = \overline{y}_{st}^* \left[\frac{a_X \overline{X} + b_X}{\alpha_X \left(a_X \overline{x}_{st}^* + b_X \right) + \left(1 - \alpha_X \right) \left(a_X \overline{X} + b_X \right)} \right]^{g_X} \left[\frac{a_Z \overline{Z} + b_Z}{\alpha_Z \left(a_Z \overline{z}_{st}^* + b_X \right) + \left(1 - \alpha_Z \right) \left(a_Z \overline{Z} + b_Z \right)} \right]^{g_Z}, \tag{11}$$

where the constants a_x & a_z ($\neq 0$), and b_x & b_z are either real numbers or the functions of the auxiliary variable, in form of coefficient of variations, standard deviations, correlation coefficients, skewness or kurtosis from the population, whereas g_x and g_z are known constants take the value (0, 1, −1) to produce respectively unbiased estimator, different families of ratio-cum-ratio and families of product-cum-product type estimators, and α_x and α_z are the constants to be determined such that MSE of the proposed estimator t_a is minimum. Taking different values of the constants we may obtain many families of ratio-cum-ratio and families of product-cum-product estimators such as some examples are presented in Table 1.

Table 1 | Families of the estimators t_a.

Family of ratio-cum-ratio $g_x = g_z = 1$	Family of product-cum-product $g_x = g_z = -1$	a_X	b_X	a_Z	b_Z	α_X	α_Z
t_a^1	t_a^2	1	0	1	0	1	1
t_a^3	t_a^4	1	ρ_{xy}	1	ρ_{yz}	1	1
t_a^5	t_a^6	σ_x	1	σ_z	1	1	1
t_a^7	t_a^8	ρ_{xy}	1	ρ_{yz}	1	1	1

3.1. Bias and MSE of Proposed Generalized Class of Estimator

To obtain the bias and MSE, we consider

$$\bar{y}_{st}^* = \sum_{h=1}^{L} P_h \bar{y}_h^* = \overline{Y}\left(1 + e_o^*\right), \bar{x}_{st}^* = \overline{X}\left(1 + e_1^*\right), \bar{z}_{st}^* = \overline{Z}\left(1 + e_2^*\right)$$

$$E(e_i^*) = 0 \qquad \text{for} \qquad i = 0,1,2$$

and

$$\left.\begin{aligned}
V_{r,s,t}^* &= \sum_{h=1}^{L} P_h^{r+s+t} \frac{E\left((\bar{x}_h'^* - \overline{X}_h)^r (\bar{y}_h^* - \overline{Y}_h)^s (\bar{z}_h^* - \overline{Z}_h)^t\right)}{\overline{X}^r \overline{Y}^s \overline{Z}^t} \\
E(e_o^*)^2 &= \frac{1}{\overline{Y}^2} \sum_{h=1}^{L} P_h^2 \left(\lambda_h S_{yh}^2 + \lambda_h^* S_{yh2}^2\right) = V_{020}^* \quad E(e_1^*)^2 = \frac{1}{\overline{X}^2} \sum_{h=1}^{L} P_h^2 \left(\lambda_h S_{xh}^2 + \lambda_h^* S_{xh2}^2\right) = V_{200}^* \\
E(e_2^*)^2 &= \frac{1}{\overline{Z}^2} \sum_{h=1}^{L} P_h^2 \left(\lambda_h S_{zh}^2 + \lambda_h^* S_{zh2}^2\right) = V_{002}^* \quad E\left(e_o^* e_2^*\right) = \frac{1}{\overline{YZ}} \sum_{h=1}^{L} P_h^2 \left(\lambda_h S_{yzh} + \lambda_h^* S_{yzh2}\right) = V_{011}^* \\
E\left(e_o^* e_1^*\right) &= \frac{1}{\overline{YX}} \sum_{h=1}^{L} P_h^2 \left(\lambda_h S_{xyh} + \lambda_h^* S_{xyh2}\right) = V_{110}^* \quad E\left(e_1^* e_2^*\right) = \frac{1}{\overline{ZX}} \sum_{h=1}^{L} P_h^2 \left(\lambda_h S_{xzh} + \lambda_h^* S_{xzh2}\right) = V_{101}^*
\end{aligned}\right\} \tag{12}$$

On rewriting we may get Eq. (11) as

$$t_a = \overline{Y}\left(1 + e_0^*\right)\left[1 + \alpha_x v_x e_1^*\right]^{-g_x} \left[1 + \alpha_z v_z e_2^*\right]^{-g_z}, \tag{13}$$

where

$$v_x = \frac{a_x \overline{X}}{a_x \overline{X} + b_x} \quad \text{and} \quad v_z = \frac{a_z \overline{Z}}{a_z \overline{Z} + b_z}.$$

We assume that $|\alpha_x v_x e_1^*| < 1$ and $|\alpha_z v_z e_2^*| < 1$, so that we may expand the series, $\left[1 + \alpha_x v_x e_1^*\right]^{-g_x}$ and $\left[1 + \alpha_z v_z e_2^*\right]^{-g_z}$, we get

$$t_a = \overline{Y}\left(1 + e_0^*\right)\left[1 - \left(g_x \alpha_x v_x\right) e_1^* + \frac{g_x\left(g_x + 1\right)}{2}\left(\alpha_x v_x\right)^2 e_1^{*2} + \dots\right]$$

$$\left[1 - g_z \alpha_z v_z e_2^* + \frac{g_z\left(g_z + 1\right)}{2}\left(\alpha_z v_z\right)^2 e_2^{*2} + \dots\right]. \tag{14}$$

It is assumed that the contribution of terms involving powers in e_0^*, e_1^* and e_2^* higher than two is negligible. We have,

$$t_a - \overline{Y} = \overline{Y}\left[e_0^* - \left(g_x \alpha_x v_x\right) e_1^* - \left(g_x \alpha_x v_x\right) e_0^* e_1^* + \frac{g_x\left(g_x + 1\right)}{2}\left(\alpha_x v_x\right) e_1^{*2} - \left(g_z \alpha_z v_z\right) e_2^*\right.$$

$$\left. - \left(g_z \alpha_z v_z\right) e_0^* e_2^* + \left(g_x \alpha_x v_x\right)\left(g_z \alpha_z v_z\right) e_1^* e_2^* + \frac{g_z\left(g_z + 1\right)}{2}\left(\alpha_z v_z\right)^2 e_2^{*2}\right]. \tag{15}$$

Taking expectation of Eq. (15), we will get the bias as,

$$\text{Bias}\left(t_a\right) = \left[g_x \alpha_x v_x V_{110}^* + \frac{g_x\left(g_x + 1\right)}{2} \alpha_x^2 v_x^2 V_{200}^*\right.$$

$$\left. - \left(g_z \alpha_z v_z\right) V_{011}^* + \left(g_x \alpha_x v_x\right)\left(g_z \alpha_z v_z\right) V_{101}^* + \frac{g_z\left(g_z + 1\right)}{2}\left(\alpha_z^2 v_z^2\right) V_{002}^*\right]. \tag{16}$$

we take square of Eq. (15) and retain terms up to the order n^{-1} then we take expectation to get the MSE of the estimator *ta* as,

$$\text{MSE}\left(t_a\right) = \overline{Y}^2\left[V_{020}^* + \left(g_x \alpha_x v_x\right)^2 V_{200}^* + \left(g_z \alpha_z v_z\right)^2 V_{002}^* - 2\left(g_x \alpha_x v_x\right) V_{110}^*\right.$$

$$\left. - 2\left(g_z \alpha_z v_z\right) V_{011}^* + 2\left(g_x \alpha_x v_x\right)\left(g_z \alpha_z v_z\right) V_{101}^*\right]. \tag{17}$$

Minimization of Eq. (17) with respect to α_x and α_z yields the optimum values as

$$\alpha_x = \frac{\left(V_{011}^* V_{101}^* - V_{110}^* V_{002}\right)}{g_x v_x \left(V_{101}^{*2} - V_{002}^* V_{200}^*\right)} \quad \text{and} \quad \alpha_z = \frac{\left(V_{101}^* V_{110}^* - V_{011}^* V_{200}\right)}{g_z v_z \left(V_{101}^{*2} - V_{002}^* V_{200}^*\right)} \tag{18}$$

Using the Eq. (18), the expression of minimum MSE may be obtained as

$$\text{MSE}\,(t_a) = \overline{Y}^2 \left[V_{020}^* - \left(\frac{V_{002}^* V_{110}^{*2} + V_{200}^* V_{011}^{*2} - 2 V_{101}^* V_{011}^* V_{110}^*}{V_{200}^* V_{002}^* - V_{101}^{*2}} \right) \right]. \tag{19}$$

For generalized family of ratio-cum-ratio estimators presented in Table 1, we can give the MSE expression in Eq. (17) as,

$$\text{MSE}\left(t_a^i\right) = \begin{cases} \overline{Y}^2 \left[V_{020}^* + V_{200}^* + V_{002}^* - 2 V_{110}^* - 2 V_{011}^* + 2 V_{101}^* \right] = \text{MSE}\,(t_2) \quad i = 1 \\ \\ \overline{Y}^2 \left[\begin{matrix} V_{020}^* + v_{x(\frac{i-1}{2})}^2 V_{200}^* + v_{z(\frac{i-1}{2})}^2 V_{002}^* - 2 v_{x(\frac{i-1}{2})} V_{110}^* \\ -2 v_{z(\frac{i-1}{2})} V_{011}^* + 2 v_{x(\frac{i-1}{2})} v_{z(\frac{i-1}{2})} V_{101}^* \end{matrix} \right] \quad i = 3, 5, 7 \end{cases}$$

and for generalized family of product-cum-product estimators, the MSE expression can be given as,

$$\text{MSE}\left(t_a^j\right) = \begin{cases} \overline{Y}^2 \left[V_{020}^* + V_{200}^* + V_{002}^* + 2 V_{110}^* + 2 V_{011}^* + 2 V_{101}^* \right] = \text{MSE}\,(t_3) \quad j = 2 \\ \\ \overline{Y}^2 \left[\begin{matrix} V_{020}^* + v_{x(\frac{j}{2}-1)}^2 V_{200}^* + v_{z(\frac{j}{2}-1)}^2 V_{002}^* + 2 v_{x(\frac{j}{2}-1)} V_{110}^* \\ +2 v_{z(\frac{j}{2}-1)} V_{011}^* + 2 v_{x(\frac{j}{2}-1)} v_{z(\frac{j}{2}-1)} V_{101}^* \end{matrix} \right] \quad j = 4, 6, 8 \end{cases}$$

Where

$$v_{x1} = \frac{\overline{X}}{\overline{X} + \rho_{xy}} \quad \text{and} \quad v_{z1} = \frac{\overline{Z}}{\overline{Z} + \rho_{yz}}$$

$$v_{x2} = \frac{\sigma_x \overline{X}}{\sigma_x \overline{X} + 1} \quad \text{and} \quad v_{z2} = \frac{\sigma_z \overline{Z}}{\sigma_z \overline{Z} + 1}$$

$$v_{x3} = \frac{\rho_{xy} \overline{X}}{\rho_{xy} \overline{X} + 1} \quad \text{and} \quad v_{z3} = \frac{\rho_{yz} \overline{Z}}{\rho_{yz} \overline{Z} + 1}$$

One can think many more estimators from Eq. (11), and the bias and MSE expression for these estimators can be expressed by Eqs. (16) and (17) respectively.

4. ANOTHER PROPOSED GENERALIZED ESTIMATORS

In this section, we have shown another improved and generalized form for the estimator *ta* proposed in Section 3. The proposed estimator *ts* for estimating population mean is given as,

$$t_s = \eta \overline{y}_{st}^* \left[\frac{a_X \overline{X} + b_X}{\alpha_X \left(a_X \overline{x}_{st}^* + b_X\right) + \left(1 - \alpha_X\right)\left(a_X \overline{X} + b_X\right)} \right]^{g_x} \left[\frac{a_Z \overline{Z} + b_Z}{\alpha_Z \left(a_Z \overline{z}_{st}^* + b_X\right) + \left(1 - \alpha_Z\right)\left(a_Z \overline{Z} + b_Z\right)} \right]^{g_z} \tag{20}$$

or may also be consider as,

$$t_s = \eta \overline{Y} \left(1 + e_0^*\right) \left[1 + \alpha_x v_x e_1^*\right]^{-g_x} \left[1 + \alpha_z v_z e_2^*\right]^{-g_z}, \tag{21}$$

We assume that $|\alpha_x v_x e_1^*| < 1$ and $|\alpha_z v_z e_2^*| < 1$, so that we may expand the series, $\left[1 + \alpha_x v_x e_1^*\right]^{-g_x}$ and $\left[1 + \alpha_z v_z e_2^*\right]^{-g_z}$, we get

$$t_s = \eta \overline{Y} \left(1 + e_0^*\right) \left[1 - \left(g_x \alpha_x v_x\right) e_1^* + \frac{g_x \left(g_x + 1\right)}{2} \left(\alpha_x v_x\right)^2 e_1^{*2} + \ldots\right]$$

$$\left[1 - g_z \alpha_z v_z e_2^* + \frac{g_z \left(g_z + 1\right)}{2} \left(\alpha_z v_z\right)^2 e_2^{*2} + \ldots\right]. \tag{22}$$

It is assumed that the contribution of terms involving powers in e_0^*, e_1^* and e_2^* higher than two is negligible. We have,

$$t_s - \overline{Y} = \overline{Y}(\eta - 1) + \overline{Y}\left[e_0^* - (g_x \alpha_x v_x) e_1^* - (g_x \alpha_x v_x) e_0^* e_1^* + \frac{g_x(g_x + 1)}{2}(\alpha_x v_x) e_1^{*2} - (g_z \alpha_z v_z) e_2^* \right.$$

$$\left. - (g_z \alpha_z v_z) e_0^* e_2^* + (g_x \alpha_x v_x)(g_z \alpha_z v_z) e_1^* e_2^* + \frac{g_z(g_z + 1)}{2}(\alpha_z v_z)^2 e_2^{*2} \right]. \tag{23}$$

The bias and MSE expressions of *ts* may be given respectively as,

$$\text{Bias}(t_s) = \overline{Y}(\eta - 1) + \text{Bias}(t_a) \tag{24}$$

and

$$\text{MSE}(t_s) = \overline{Y}^2 (\eta - 1)^2 + \eta^2 \text{MSE}(t_a) + 2\overline{Y}\eta(\eta - 1)\text{Bias}(t_a) \tag{25}$$

We differentiate Eq. (25) w.r.t $\bar{\eta}$, and equate it to zero, we get

$$\bar{\eta} = \frac{\overline{Y}\left(\overline{Y} + \text{Bias}(t_a)\right)}{\left[\overline{Y}^2 + \text{MSE}(t_a) + 2\overline{Y}\text{Bias}(t_a)\right]}, \tag{26}$$

On substitution the optimum value of η as given in Eq. (26), in the result Eq. (25), the min MSE of the proposed estimator *ts* is obtained as,

$$\text{min MSE}(t_s) = \overline{Y}^2 - \frac{\overline{Y}^2\left(\overline{Y} + \text{Bias}(t_a)\right)^2}{\left[\overline{Y}^2 + \text{MSE}(t_a) + 2\overline{Y}\text{Bias}(t_a)\right]}$$

Or

$$\text{min MSE}(t_s) = \overline{Y}^2 \left[1 - \frac{\left(1 + \frac{\text{Bias}(t_a)}{\overline{Y}}\right)^2}{\left[1 + \frac{\text{MSE}(t_a)}{\overline{Y}^2} + 2\frac{\text{Bias}(t_a)}{\overline{Y}}\right]} \right] \tag{27}$$

We can think many more improved estimators taking different values of the constants in Eq. (20), such as some example are given in Table 2. The MSE of the ratio-cum-ratio and product-cum-product estimators given in Table 2 can be given using Eq. (27) as,

$$\text{MSE}(t_s^i) = \overline{Y}^2 \left[1 - \frac{\left(1 + \frac{\text{Bias}(t_a^i)}{\overline{Y}}\right)^2}{\left[1 + \frac{\text{MSE}(t_a^i)}{\overline{Y}^2} + 2\frac{\text{Bias}(t_a^i)}{\overline{Y}}\right]} \right] \quad \text{for } i = 1, 3, 5, 7 \tag{28}$$

$$\text{MSE}(t_s^j) = \overline{Y}^2 \left[1 - \frac{\left(1 + \frac{\text{Bias}(t_a^j)}{\overline{Y}}\right)^2}{\left[1 + \frac{MSE(t_a^j)}{\overline{Y}^2} + 2\frac{\text{Bias}(t_a^j)}{\overline{Y}}\right]} \right] \quad \text{for } j = 2, 4, 6, 8 \tag{29}$$

Table 2 | Families of the proposed estimator *ts*.

Family of ratio-cum-ratio $g_x = g_z = 1$	Family of product-cum-product $g_x = g_z = -1$	(a_x, b_x, a_z, b_z)	α_x	α_z	η
t_s^1	t_s^2	$(1, 0, 1, 0)$	1	1	η
t_s^3	t_s^4	$(1, \rho_{xy}, 1, \rho_{yz})$	1	1	η
t_s^5	t_s^6	$(\sigma_x, 1, 1, \sigma_y)$	1	1	η
t_s^7	t_s^8	$(\rho_{xy}, 1, 1, \rho_{yz})$	1	1	η

5. MATHEMATICAL COMPARISON

In this section, the MSE of the suggested family of estimator has been compared with the mean estimator, stratified ratio estimator and the class of estimators. Let us consider following notations as:

$$A_i = \begin{bmatrix} v^2_{x(\frac{i-1}{2})} V^*_{200} + v^2_{z(\frac{i-1}{2})} V^*_{002} - 2v_{x(\frac{i-1}{2})} V^*_{110} \\ -2v_{z(\frac{i-1}{2})} V^*_{011} + 2v_{x(\frac{i-1}{2})} v_{z(\frac{i-1}{2})} V^*_{101} \end{bmatrix} \qquad i = 3, 5, 7$$

$$A_j = \begin{bmatrix} v^2_{x(\frac{j}{2}-1)} V^*_{200} + v^2_{z(\frac{j}{2}-1)} V^*_{002} + 2v_{x(\frac{j}{2}-1)} V^*_{110} \\ +2v_{z(\frac{j}{2}-1)} V^*_{011} + 2v_{x(\frac{j}{2}-1)} v_{z(\frac{j}{2}-1)} V^*_{101} \end{bmatrix} \qquad j = 4, 6, 8$$

$$B_1 = v^2_{z(\frac{i-1}{2})} V^*_{002} - 2v_{z(\frac{i-1}{2})} V^*_{011} - V^*_{200} - V^*_{002} + 2V^*_{110} + 2V^*_{011} - 2V^*_{101}$$

$$B_2 = \left(\frac{V^*_{002} V^{*2}_{110} + V^*_{200} V^{*2}_{011} - 2V^*_{101} V^*_{011} V^*_{110}}{V^*_{200} V^*_{002} - V^{*2}_{101}} \right)$$

and

$$C^i = \frac{\left(1 + \frac{\text{Bias}(t^i_a)}{\overline{Y}}\right)^2}{\left[1 + \frac{\text{MSE}(t^i_a)}{\overline{Y}^2} + 2\frac{\text{Bias}(t^i_a)}{\overline{Y}}\right]}$$

The efficiency conditions may be written as:

i.

$$\begin{aligned} &\text{Var}(t_1) > \text{MSE}(t^i_a) \qquad\qquad i = 3, 5, 7\\ &\text{If}\\ &V^*_{020} - A_i > 0 \end{aligned} \tag{30}$$

ii.

$$\begin{aligned} &\text{MSE}(t_2) > \text{MSE}(t^i_a)\\ &\text{If}\\ &\min\left(\frac{-A_i \pm \sqrt{A^2_i - B_1 V^*_{200}}}{V^*_{200}}\right) \le v_{x(\frac{i-1}{2})} \le \max\left(\frac{-A_i \pm \sqrt{A^2_i - B_1 V^*_{200}}}{V^*_{200}}\right)\\ &[14pt] \end{aligned} \tag{31}$$

iii.

$$\begin{aligned} &\text{Var}(t_1) > \min \text{MSE}(t^i_a)\\ &\text{If}\\ &B_2 > 0 \end{aligned} \tag{32}$$

iv.

$$\begin{aligned} &\text{MSE}(t_2) > \min \text{MSE}(t^i_a)\\ &\text{If}\\ &B_2 > V^*_{200} + V^*_{002} - 2V^*_{110} - 2V^*_{011} + 2V^*_{101} \end{aligned} \tag{33}$$

v.

$$\begin{aligned} &\text{Var}(t_1) > \text{MSE}(t^i_s)\\ &\text{If}\\ &C^i > 1 - V^*_{020} \end{aligned} \tag{34}$$

vi.

$$\text{MSE}(t_2) > \text{MSE}(t_s^i)$$
$$\text{If}$$
$$C^i > 1 - \left(V_{020}^* + V_{200}^* + V_{002}^* - 2V_{110}^* - 2V_{011}^* + 2V_{101}^* \right)$$

(35)

vii.

$$\min \text{MSE}(t_a^i) > \min \text{MSE}(t_s^i)$$
$$\text{If}$$
$$C^i > 1 + B_2 - V_{020}^*$$

(36)

6. COST FUNCTION AND SAMPLE SIZE ESTIMATION

Let us assume a linear cost function, the total cost of the sample survey could be written as

$$C' = \sum_{h=1}^{L} c_{h0} n_h + \sum_{h=1}^{L} c_{h1} n_h + \sum_{h=1}^{L} c_{h2} r_h$$

(37)

where ch_0 denotes the per unit cost of making first attempt, ch_1 denotes per unit cost for processing the result of all characteristics in first attempt and chh denotes the per unit cost for processing the result of all characteristics in second attempt in the hth stratum.

The total expected cost of the survey could be given as

$$C = E(C') = \sum_{h=1}^{L} (c_{h0} + c_{h1} W_{h1}) n_h + \sum_{h=1}^{L} c_{h2} r_h$$

(38)

6.1. Cost Function for Estimator t_a

Let us consider the function for estimator ta

$$\varphi(n_h, r_h, \delta) = \text{Var}(t_a) + \delta \{ C - C_0 \}$$

$$\varphi(n_h, r_h, \delta) = \overline{Y}^2 \left[V_{020}^* + u_X^2 V_{200}^* + u_Z^2 V_{002}^* - 2u_X V_{110}^* - 2u_Z V_{011}^* + 2u_X u_Z V_{101}^* \right]$$
$$+ \delta \left[\sum_{h=1}^{L} (c_{h0} + c_{h1} W_{h1}) n_h + \sum_{h=1}^{L} c_{h2} r_h - C_0 \right]$$

(39)

Let

$$S_h = \left(\frac{1}{\overline{Y}^2} S_{y_h}^2 + \frac{1}{\overline{X}^2} S_{x_h}^2 + \frac{1}{\overline{Z}^2} S_{z_h}^2 - 2u_x \frac{1}{\overline{XY}} S_{xy_h} - 2u_z \frac{1}{\overline{YZ}} S_{yz_h} + 2u_x u_z \frac{1}{\overline{XZ}} S_{xz_h} \right)$$

and

$$S_h^* = \left(\frac{1}{\overline{Y}^2} S_{y_h}^{*2} + \frac{1}{\overline{X}^2} S_{x_h}^{*2} + \frac{1}{\overline{Z}^2} S_{z_h}^{*2} - 2u_x \frac{1}{\overline{XY}} S_{xy_h}^* - 2u_z \frac{1}{\overline{YZ}} S_{yz_h}^* + 2u_x u_z \frac{1}{\overline{XZ}} S_{xz_h}^* \right)$$

(40)

So we can write Eq. (39) as

$$\varphi(n_h, r_h, \delta) = \overline{Y}^2 \sum_{h=1}^{L} P_h^2 \left[\left(\frac{1}{n_h} - \frac{1}{N_h} \right) S_h + \left(\frac{W_{2h}(f_h - 1)}{r_h} \right) S_h^* \right] + \delta \left[\sum_{h=1}^{L} (c_{h0} + c_{h1} W_{h1}) n_h + \sum_{h=1}^{L} c_{h2} r_h - C_0 \right]$$

(41)

Where δ is lagrangian multiplier. On differentiating Eq. (41) we can get the values of nh, rh, and δ as,

$$n_h = \sqrt{\frac{\overline{Y}^2 P_h^2 S_h}{\delta (c_{h0} + c_{h1} W_{h1})}}$$

(42)

Also

$$r_h = \sqrt{\frac{\overline{Y}^2 P_h^2 \left[W_{2h}(f_h - 1) S_h^* \right]}{\delta c_{h2}}}$$

(43)

$$\frac{C_0}{\overline{Y} \sum_{h=1}^{L} P_h \left[\sqrt{(c_{h0} + c_{h1} W_{h1}) S_h} + \sqrt{\left[W_{2h} \left(f_h - 1 \right) S_h^* \right] c_{h2}} \right]} = \frac{1}{\sqrt{\delta}}$$

Let

$$D = \sum_{h=1}^{L} P_h \left[\sqrt{(c_{h0} + c_{h1} W_{h1}) S_h} + \sqrt{\left[W_{2h} \left(f_h - 1 \right) S_h^* \right] c_{h2}} \right]$$

$$\frac{1}{\sqrt{\delta}} = \frac{C_0}{\overline{Y}D} \tag{44}$$

Subsituting δ from Eqs. (44) to (42), we have

$$n'_h = \frac{C_0 \sqrt{P_h^2 S_h}}{D \sqrt{(c_{h0} + c_{h1} W_{h1})}} \tag{45}$$

Similarly, substitute the value of δ from Eqs. (44) to (43) we have

$$r'_h = \frac{C_0 \sqrt{P_h^2 \left[W_{2h} \left(f_h - 1 \right) S_h^* \right]}}{D \sqrt{c_{h2}}} \tag{46}$$

7. EMPIRICAL STUDY

In order to see the performance of the suggested family of estimators as compare to class of estimators under stratified random sampling. The statistics of the two stratified populations have been given in Table 3.

Population-I: [Source: Koyuncu and Kadilar [9]]

We consider No. of teachers as study variable *(Y)*, No. of students as auxiliary variable *(X)*, and No. of classes in primary and secondary schools as another auxiliary variable *(Z)* for 923 districts at six 6 regions (1: Marmara, 2: Agean, 3: Mediterranean, 4: Central Anatolia, 5: Black Sea, and 6: East and Southeast Anatolia) in Turkey in 2007.

Population-II: [source: detailed livelihood assessment of flood affected districts of Pakistan September 2011, Food Security Cluster, Pakistan]

We consider food expenditure as study variable *(Y)*, household earn as auxiliary variable *(X)*, and total expenditure in May (2011) as another auxiliary variable *(Z)* for (6 940) male and (1 678) female households in flood affected districts of Pakistan.

Neyman allocation has been used in order to allocate sample sizes to different strata in the two populations separately as:

$$n_h = n \frac{N_h S_h}{\sum_{h=1}^{L} N_h S_h}$$

From Table 3, we observe a positive correlation among study variable and the auxiliary variable in order to use ratio estimators for estimating population mean. The comparison of the proposed estimators have been made with respect to Hansen-Hurwtiz [1] and ratio estimators modified for stratified sampling. The information for the two stratified populations is given in Table 3.

The MSE values of the proposed class of estimators are computed in Table 4 two different data sets. The percent relative efficiencies of the estimators are given in Table 5. Efficiency of each estimator has been tested by increasing the non-response rate from 10% to 30% each with three different values of *fh* (2, 2.5&3). From Table 4 it is observed that MSE's of the estimators are increased if non-response increases from 10% to 30% however from Table 5 it is noticed that *PRE* of the proposed estimator is also increased if non-response increases from 10% to 30% which shows that the proposed estimators *ta*, and *ts* as compare to the existing estimators (*t₁* and *t₂*) can perform more efficiently even at higher non-response rate. If we compare proposed estimators *ta* and *ts* with each other, we observe that MSE values of the proposed estimators *ts* are smaller than the MSE values of proposed estimator *ta* at each non-response rate. If we compare the two proposed estimators *ta* and *ts* on basis their *PRE* values, the conclusion can be drawn that *ts* is an improved form of *ta*.

Table 3 | Data statistics for the two populations.

Stratum (h)		Population-I						Population-II		
		1	2	3	4	5	6	1	2	3
Stratified mean, S.Ds and correlation coefficients	N_h	127	117	103	170	205	201	21	34	26
	n_h	31	21	29	38	22	39	06	04	02
	n'_h	70	50	75	95	70	90	15	17	08
	S_{yh}	883.84	644.92	1 033.40	810.58	403.65	711.72	12.14	8.34	5.47
	S_{xh}	30 486.7	15 180.77	27 549.69	18 218.93	8 497.77	23 094.14	76.71	31.94	49.55
	S_{zh}	555.58	365.46	612.95	458.03	260.85	397.05	19.48	07.10	13.21
	$\overline{Y}_h$	703.74	413	573.17	424.66	267.03	393.84	37.55	37.25	26.39
	$\overline{X}_h$	20 804.59	9 211.79	14 309.30	9 478.85	5 569.95	12 997.59	116.57	093.00	26.39
	$\overline{Z}_h$	498.28	318.33	431.36	311.32	227.20	313.71	114.14	106.50	118.88
	ρ_{xyh}	0.9360	0.9960	0.9940	0.9830	0.9890	0.9650	0.7914	0.8339	0.7696
	ρ_{xzh}	0.9396	0.9696	0.9770	0.9640	0.9670	0.9960	0.9894	0.8820	0.9669
	ρ_{yzh}	0.9790	0.9760	0.9840	0.9830	0.9640	0.9830	0.7781	0.6651	0.5935
Wh=10% Non-response	S_{yh_2}	510.57	386.77	1 872.88	1 603.30	264.19	497.84	08.66	10.05	03.95
	S_{xh_2}	9 446.93	9 198.29	52 429.99	34 794.9	4 972.56	12 485.10	42.14	13.28	74.22
	S_{zh_2}	303.92	278.51	960.71	821.29	190.85	287.99	6.25	5.20	20.53
	ρ_{xy2}	0.9961	0.9975	0.9998	0.9741	0.9950	0.9284	0.9997	0.9995	0.9840
	ρ_{xz2}	0.9901	0.9895	0.9964	0.9609	0.9865	0.9752	0.9707	1.0000	0.9999
	ρ_{yz2}	0.9931	0.9871	0.9972	0.9942	0.9850	0.9647	0.9649	0.9996	0.9819
Wh=20% Non-response	S_{yh_2}	396.77	406.15	1 654.40	1 333.35	335.83	903.91	7.96	8.47	4.06
	S_{xh_2}	7 439.16	8 880.46	4 5784.78	2 9219.30	6 540.43	28 411.44	36.50	25.82	59.32
	S_{zh_2}	244.56	274.42	965.42	680.28	214.49	469.86	5.20	8.18	16.54
	ρ_{xy2}	0.9954	0.9931	0.9960	0.9761	0.9966	0.9869	0.9905	0.8026	0.8601
	ρ_{xz2}	0.9897	0.9884	0.9789	0.9629	0.9820	0.9825	0.9623	0.9858	0.9956
	ρ_{yz2}	0.9898	0.9798	0.9846	0.9940	0.9818	0.9874	0.9297	0.8062	0.8112
Wh=30% Non-response	S_{yh_2}	500.26	356.95	1 383.70	1 193.47	289.41	825.24	12.70	09.86	4.50
	S_{xh_2}	14 017.99	7 812.00	38 379.77	26 090.60	5 611.32	24 571.95	37.69	24.02	52.26
	S_{zh_2}	284.44	247.63	811.21	631.28	188.30	437.90	9.42	6.83	14.54
	ρ_{xy2}	0.9639	0.9919	0.9955	0.9801	0.9961	0.9746	0.9288	0.8335	0.8275
	ρ_{xz2}	0.9107	0.9848	0.9771	0.9650	0.9794	0.9642	0.9062	0.8859	0.9907
	ρ_{yz2}	0.9739	0.9793	0.9839	0.9904	0.9799	0.9829	0.9696	0.5877	0.7542

Table 4 | MSEs values of estimators.

		MSEs of the different existing estimators and Proposed class of estimators			
Wh	**fh**	t_1	t_2	$t_a^{\min}$	$t_s^{\min}$
	2	2143.99	1411.42	76.1190	**76.0883**
10%	2.5	2370.94	1501.89	81.7661	**81.7304**
	3	2597.83	1592.34	87.1097	**87.0691**
	2	2540.33	1667.89	83.2651	**83.2284**
20%	2.5	2965.45	1886.62	92.3106	**92.2641**
	3	3390.54	2105.42	101.171	**101.116**
	2	2703.09	1753.48	88.1595	**88.1179**
30%	2.5	3209.59	2015.08	99.6831	**99.6291**
	3	3716.08	2276.639	110.882	**110.815**

Table 5 | PRE values of estimators.

		PREs of the different existing estimators and Proposed class of estimators			
Wh	**fh**	t_1	t_2	$t_a^{\min}$	$t_s^{\min}$
	2	100	151.90305	2816.6292	**2817.7657**
10%	2.5	100	157.86376	2899.6614	**2900.9279**
	3	100	163.14543	2982.2511	**2983.6417**
	2	100	152.30801	3050.8941	**3052.2394**
20%	2.5	100	157.18322	3212.4696	**3214.0887**
	3	100	161.03865	3351.2963	**3353.1192**
	2	100	154.15574	3066.1358	**3067.5833**
30%	2.5	100	159.27854	3219.7935	**3221.5387**
	3	100	163.22658	3351.3826	**3353.4088**

8. CONCLUSION

From the results of the numerical study, we infer that both the proposed estimators using auxiliary information perform more efficiently (with substantial gain in precision) than t_1 and t_2. Hence proposed estimators are recommended for their practical use.

REFERENCES

1. M.H. Hansen, W.N. Hurwitz, J. Am. Stat. Assoc. 41(236) (1946), 517–529.
2. W.G. Cochran, Sampling Techniques, Third ed., John Wiley, New York, 1977.
3. M.K. Chaudhary, R. Singh, R.K. Shukla, M. Kumar, F. Smarandache, Pak. J. Stat. Oper. Res. 5(1) (2009), 47–54.
4. A. Haq, J. Shabbir, Commun. Stat. Theory Methodol. 42(1) (2013), 782–799.
5. A. Sanaullah, M. Noor-ul-Amin, M. Hanif, Pak. J. Stat. 31(1) (2015), 71–94.
6. W.G. Cochran, J. Am. Stat. Assoc. 37 (1942), 199–212.
7. C. Kadilar, H. Cingi, Biom. J. 45(2) (2003), 218–225.
8. J. Shabbir, S. Gupta, Am. J. Math. Manag. Sci. 25(3-4) (2005), 293–311.
9. N. Koyuncu, C. Kadilar, J. Stat. Plan. Inference. 139(8) (2009b), 2552–2558.
10. A. Sanaullah, H. Khan, H.A. Ali, R. Singh, J. Reliab. Stat. Stud. 5(2) (2012), 119–132.
11. A. Sanaullah, H.A. Ali, M. Noor ul Amin, M. Hanif, Appl. Math. Comput. 226 (2014), 541–547.
12. N. Koyuncu, C. Kadilar, Commun Stat Theory Methodol. 38(1) (2009a), 2398–2417.