

# An Efficient Family of Synthetic Estimators for Small Areas and its Applications

V. K. Singh\* and S. K. Seth

Department of Statistics, Banaras Hindu University, Varanasi-221005, India.

Received: 13 Oct. 2014, Revised: 13 Nov. 2014, Accepted: 15 Nov. 2014

Published online: 1 Jan. 2015

---

**Abstract:** In making estimates for small areas with adequate level of precision, it is often necessary to use “indirect estimators” that “borrow strength” by using the values of the variable of interest from related areas, comparatively bigger in size and/or time periods and thus, by increasing the effective sample size. If a reliable direct estimator for a large area, covering several small areas, is used to derive an indirect estimator for a small area, then such an estimator is called a “synthetic estimator”. The synthetic estimators as compared to other methods of estimation, are largely used in small area estimation problems due to its simplicity and applicability to general sampling design and potential to increase accuracy in estimation.

This paper is devoted to the development of a class of synthetic estimators for population mean of small area which exhibits some nice properties and includes some well-known synthetic estimators. The class utilizes information on an auxiliary characteristic. It has been demonstrated with the help of empirical data that the suggested class performs better than other synthetic estimators of similar kind.

**Keywords:** Family of estimators, synthetic estimator, small domain.

---

## 1 Introduction

While surveys are normally planned with specific need of population parameters, sometimes interest also lies in parts of the population known as “subpopulations” or “domains” of interest. In case such domains get sufficient representation of sampled units in the main sample, the domain parameters may be estimated satisfactorily through direct estimators, but generally subpopulations are too small to provide reliable direct estimates. Thus, a domain is regarded as small if the domain-specific sample is not large enough to support direct estimates of adequate precision (Rao, [8]).

The topic of Small Area Estimation (SAE) refers to reliable estimation of parameters of small domains with greater precision using indirect estimators which might be based upon data from other domains and/or data already collected through large scale surveys. In the theory of small area estimation, such indirect estimators are termed as “Synthetic Estimators” which borrow strength by using values of the variables of interest from related areas, thus, increasing the effective sample size (Holt *et al*, [3]). Gonzalez[1] described synthetic estimator as :

*“An unbiased estimate is obtained from a sample survey for a large area; when this estimator is used to derive estimates for sub-areas under the assumption that the small areas have the same characteristics as the large area, we identify these estimates as Synthetic Estimates.”*

The term “synthetic estimates” was first used by the U.S. National Centre for Health Statistics[4] of the United State when it computed estimates of long and short term physical disabilities based on the national health interview survey. The synthetic method of estimation is by far one of the most widely used small area estimation methods due to its simple common sense approach. However, it is evident that at some point, as the sample size in a small area increases, a direct estimator becomes more desirable than a synthetic one, but if small domain sample sizes are relatively small, the synthetic estimator perform better than simple direct estimators (Schaible *et al*, [10]). Gonzalez and waksberg[2] also studied errors of synthetic and direct estimates for standard metropolitan statistical areas and counties of United States of America.

In the context of SAE, it was shown by Tikkiwal and Ghiya[15], Tikkiwal and Pandey[16], Pandey and Tikkiwal[5], Singh and Seth[12] and others that when an auxiliary variable, closely related with study variable, was used to develop synthetic ratio or regression estimators or general class of synthetic and composite ratio-type estimators, the small area estimators

---

\* Corresponding author e-mail: [vijay\\_usha\\_2000@yahoo.com](mailto:vijay_usha_2000@yahoo.com)

outperform the estimators which do not use auxiliary information.

The purpose of the present paper is to suggest an efficient class of synthetic estimators for estimating domain mean, utilizing the information on an auxiliary variable and to discuss its properties. The suggested class includes some well-known synthetic estimators and some ratio and product type estimators which might be useful for SAE. Using some empirical data it has been demonstrated that the proposed class performs better than other similar synthetic estimators.

## 2 The problem and notations

Let us consider a finite population  $U = \{Y_1, Y_2, \dots, Y_N\}$  of size  $N$ , where  $Y$  represents the characteristic under study. Let the population be divided into 'A' non-overlapping small domains  $U_a$  of size  $N_a (a = 1, 2, 3, \dots, A)$  for which estimates are required. Let  $X$  be an auxiliary variable for which the information are available in the population. Let a simple random sample of size  $n$  be selected from the population  $U$  such that  $n_a$  units in the sample comes from small domain  $U_a$ . Obviously then we have

$$\sum_{a=1}^A N_a = N \text{ and } \sum_{a=1}^A n_a = n.$$

We define the following population and sample values for the characteristics  $X$  and  $Y$  :

$\bar{Y}(\bar{X})$ : mean of the variable  $Y(X)$  in the population.

$\bar{Y}_a(\bar{X}_a)$ : mean of the variable  $Y(X)$  for the domain  $U_a$ .

$\bar{y}(\bar{x})$ : mean of the variable  $Y(X)$  in the sample of size  $n$ .

$\bar{y}_a(\bar{x}_a)$ : mean of the variable  $Y(X)$  in the sample of size  $n_a$ .

In the similar fashion, we can define mean squares and coefficients of variation in the population, in the domain 'a' and in the samples as :

$$S_Y^2(S_X^2), C_Y(C_X), S_{XY}, C_{XY}, S_{Y_a}^2(S_{X_a}^2), C_{Y_a}(C_{X_a}), C_{Y_a X_a}, S_{Y_a X_a}$$

where

$$S_Z^2 = \frac{1}{N-1} \sum_{i=1}^N (z_i - \bar{Z})^2$$

$$S_{XY} = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{Y})(x_i - \bar{X})$$

$$S_{Z_a}^2 = \frac{1}{N_a-1} \sum_{i=1}^{N_a} (z_{a_i} - \bar{Z}_a)^2$$

$$S_{Y_a X_a} = \frac{1}{N_a-1} \sum_{i=1}^{N_a} (y_{a_i} - \bar{Y}_a)(x_{a_i} - \bar{X}_a)$$

$$C_Z = S_Z/\bar{Z}, \quad C_{XY} = S_{XY}/\bar{X}\bar{Y}, \quad C_{Z_a} = S_{Z_a}/\bar{Z}_a,$$

$$C_{Y_a X_a} = S_{Y_a X_a}/\bar{Y}_a\bar{X}_a; \quad Z = X, Y;$$

$z_{a_i}$  being the  $i^{th}$  observation of the domain 'a'.

Let us denote the coefficient of correlation in the population and the domain 'a' as  $\rho_{YX}$  and  $\rho_{Y_a X_a}$  respectively.

Further let

$$V_{ij} = E \left[ \left( \frac{\bar{y} - \bar{Y}}{\bar{Y}} \right)^i \left( \frac{\bar{x} - \bar{X}}{\bar{X}} \right)^j \right]$$

for  $i, j = 0, 1, 2, \dots$

## 3 Generalized class of synthetic estimators

Our aim is here to propose a generalized class of synthetic estimators for estimating  $\bar{Y}_a$  based on auxiliary information  $X$ . Let us define the class of estimators as

$$T_{\alpha, r}^a = \bar{y} \Psi^*(\alpha, \bar{X}_a, \bar{x}_r) \quad (1)$$

where  $\Psi^*(\alpha, \bar{X}_a, \bar{x}_r)$  is a function of the parameter  $\alpha$ ,  $\bar{X}_a$  and  $\bar{x}_r$  such that

$$\Psi^*(\alpha, \bar{X}_a, \bar{x}_r) = \frac{\eta^* \{ \Phi_1(\alpha) \}}{\eta^* \{ \Phi_2(\alpha) \}} \quad (2)$$

with  $\Phi_1(\alpha) = \frac{fB}{A+fB+C}$ ,  $\Phi_2(\alpha) = \frac{C}{A+fB+C}$ ,  $f = \frac{n}{N}$ ,

$$\eta^* \{ \Phi_i(\alpha) \} = \Phi_i(\alpha) + \{1 - \Phi_i(\alpha)\} \frac{\bar{X}_a}{\bar{x}_r}; \quad i = 1, 2;$$

$$\bar{x}_r = r\bar{x} + (1-r)\bar{X}_a, \quad r = n/(N+n),$$

$A = (\alpha - 1)(\alpha - 2)$ ,  $B = (\alpha - 1)(\alpha - 4)$ ,  $C = (\alpha - 2)(\alpha - 3)(\alpha - 4)$ ;  $\alpha > 0$ .

**Remark 1.** The class  $T_{\alpha,r}^a$  define a one-parameter family of synthetic estimators for the domain 'a', borrowing information of  $\bar{y}$  and  $\bar{x}$ , the sample means of the variables Y and X respectively, based on a sample of size n where  $n > n_a$ .

**Remark 2.** In fact, the estimator  $T_{\alpha,r}^a$  is a synthetic version of an efficient family of factor-type estimator(FTE), proposed by Singh and Shukla[13]and Shukla[11].

**Remark 3.** The estimator  $T_{\alpha,r}^a$  is motivated by the estimator  $T_{\alpha,a}$  of Singh and Seth[12]. However since  $0 \leq r \leq 0.5$ , the two classes are not similar, except when the auxiliary information were not utilized at all, which happens when  $\alpha = 4$  in both the estimators, for which  $T_{4,r}^a \equiv T_{4,a} = \bar{y}$ .

#### 4 Some particular cases of $T_{\alpha,r}^a$

It is worth while here to consider some members of the family generated by  $T_{\alpha,r}^a$ , which are, in fact, synthetic versions of some well-known estimators existing in the literature of survey sampling. Further, in order to show that the two families,  $T_{\alpha,r}^a$  and  $T_{\alpha,a}$  are not equivalent, the corresponding members of the family generated by  $T_{\alpha,a}$  are also presented.

(i) Consider  $\alpha = 1$ , then

$$T_{1,r}^a = \bar{y} \left( \frac{\bar{X}_a}{\bar{x}_r} \right) = \bar{y} \left[ \frac{\bar{X}_a}{r\bar{x} + (1-r)\bar{X}_a} \right] \tag{3}$$

whereas

$$T_{1,a} = \bar{y} \left( \frac{\bar{X}_a}{\bar{x}} \right). \tag{4}$$

Thus,  $T_{1,r}^a \equiv T_{1,a}$  only if  $r=0$ , which is not admissible value of  $r$  due to obvious reason.

It is obvious from (3) and (4) that both are ratio-type synthetic estimators. The estimator  $T_{1,a}$  has been considered by Rao[8], Tikkiwal and Ghiya[15]and others.

**Remark 4.** It is remarkable that the estimator  $T_{1,r}^a$  is of the form

$$T_W = \bar{y} \left[ \frac{\bar{X}}{\lambda\bar{x} + (1-\lambda)\bar{X}} \right] \tag{5}$$

defined by Walsh[17] and Reddy[7]. Therefore, a class of synthetic estimators based upon  $T_W$  may be considered as follows :

$$T_{W,a} = \bar{y} \left[ \frac{\bar{X}_a}{\lambda\bar{x} + (1-\lambda)\bar{X}_a} \right] \tag{6}$$

and if  $\lambda = r$  (a known value),  $T_{W,a}$  is similar to  $T_{1,r}^a$ .

(ii) let  $\alpha = 2$ , then we have

$$T_{2,r}^a = \bar{y} \left( \frac{\bar{x}_r}{\bar{X}_a} \right) = \bar{y} \left[ \frac{r\bar{x} + (1-r)\bar{X}_a}{\bar{X}_a} \right] \tag{7}$$

and

$$T_{2,a} = \bar{y} \left( \frac{\bar{x}}{\bar{X}_a} \right) \tag{8}$$

which are product-type synthetic estimators.

$T_{2,r}^a$  may be considered a synthetic product estimator based on Walsh[17]estimator.

(iii) for  $\alpha = 3$ , we have

$$T_{3,r}^a = \bar{y} \left[ \frac{N\bar{X}_a - n\bar{x}_r}{(N-n)\bar{X}_a} \right] \tag{9}$$

whereas

$$T_{3,a} = \bar{y} \left[ \frac{N\bar{X}_a - n\bar{x}}{(N-n)\bar{X}_a} \right]. \tag{10}$$

Both the members may be considered as synthetic version of dual to ratio estimator proposed by Srivenkataramana[14].

(iv) For  $\alpha = 4$ , we see that

$$T_{4,r}^a \equiv T_{4,a} = \bar{y} \quad (11)$$

which is simple synthetic estimator as discussed under Remark 3, section 3.

The above discussions show that the proposed class of synthetic estimators for SAE includes a number of ratio and/or product synthetic estimators. Thus, a study of properties of  $T_{\alpha,r}^a$  would be helpful in the study of these estimators.

## 5 Asymptotic property of $T_{\alpha,r}^a$

Since  $\alpha$  is a parameter, such that  $\alpha > 0$ , it may take any value in the range  $(0, \infty)$ . In the previous section, we located some important members of the family of estimators generated by  $T_{\alpha,r}^a$  for some specific values of  $\alpha$ . However, depending upon the situations, one can choose a large value of  $\alpha$ . Thus it is advisable to observe the existence of the estimator for infinitely large value of the parameter. Re-writing the expression (1) in terms of  $\alpha$ , dividing the numerator and denominator by  $\alpha^3$  and taking limit of the expression as  $\alpha \rightarrow \infty$ , it can be seen that

$$\lim_{\alpha \rightarrow \infty} T_{\alpha,r}^a = T_{1,r}^a = \bar{y} \left( \frac{\bar{X}_a}{\bar{x}_r} \right). \quad (12)$$

Thus, in fact, even for quite larger value of the parameter the estimator  $T_{\alpha,r}^a$  converges to a well-defined estimator. This property makes the class more applicable as compared to other class of synthetic estimators, like

$$\bar{y}_{syn,a} = \bar{y} \left( \frac{\bar{x}}{\bar{X}_a} \right)^\beta \quad (13)$$

proposed by Tikkiwal and Ghiya[15] and

$$\bar{y}_{syn,a} = W_1 \bar{y} \left( \frac{\bar{x}_1}{\bar{X}_{1a}} \right)^{\beta_1} + W_2 \bar{y} \left( \frac{\bar{x}_2}{\bar{X}_{2a}} \right)^{\beta_2} \quad (14)$$

proposed by Rai and Pandey[6], both of which do not exhibit convergence property.

## 6 Design-bias and mean square error(MSE)of $T_{\alpha,r}^a$

**6.1 Theorem 1.** Under the large sample approximations, the bias of the estimator is given by

$$B [T_{\alpha,r}^a] = \left\{ \frac{Q_1}{Q_2} \bar{Y} - \bar{Y}_a \right\} + \frac{Q_1}{Q_2} \bar{Y} r \bar{X} \left\{ \frac{fB}{Q_1} - \frac{C}{Q_2} \right\} \left\{ V_{11} - \frac{Cr\bar{X}}{Q_2} V_{02} \right\} \quad (15)$$

where  $Q_1 = \{A + f(1-r)B + C\} \bar{X}_a + frB\bar{X}$

and  $Q_2 = \{A + fB + (1-r)C\} \bar{X}_a + rC\bar{X}$ .

It is to be observed that the estimator would be heavily biased unless

$$\frac{Q_1}{Q_2} = \frac{\bar{Y}_a}{\bar{Y}}$$

The proof of the expression (15) is presented in the Appendix.

**6.2 Theorem 2.** The expression of MSE of  $T_{\alpha,r}^a$  upto the order  $O(n^{-1})$  is

$$M [T_{\alpha,r}^a] = \left(\frac{Q_1}{Q_2}\bar{Y} - \bar{Y}_a\right)^2 + \left(\frac{Q_1}{Q_2}\bar{Y}\right)^2 V_{20} + \left(\frac{Q_1}{Q_2}\bar{Y}\right) r^2 \bar{X}^2 \left\{ \frac{C}{Q_2} - \frac{fB}{Q_1} \right\} \\ \left[ \left(\frac{Q_1}{Q_2}\bar{Y}\right) \left(\frac{C}{Q_2} - \frac{fB}{Q_1}\right) + 2 \left(\frac{C}{Q_2}\right) \left\{ \frac{Q_1}{Q_2}\bar{Y} - \bar{Y}_a \right\} \right] V_{02} \\ - 2 \left(\frac{Q_1}{Q_2}\bar{Y}\right) r\bar{X} \left\{ \frac{C}{Q_2} - \frac{fB}{Q_1} \right\} \left\{ 2 \frac{Q_1}{Q_2}\bar{Y} - \bar{Y}_a \right\} V_{11}. \quad (16)$$

The derivation of the expression(16) has been presented in the Appendix.

### 7 Bias and MSE of some particular cases of $T_{\alpha,r}^a$

In the section 4 above, we have observed that the proposed family involves a number of synthetic ratio and/or product type estimators, which might be discussed for their importance in SAE problems. Some particular cases of  $T_{\alpha,r}^a$  are, therefore, discussed here with their design-bias and MSE expressions.

**Case I:**  $\alpha = 1$ . The estimator  $T_{\alpha,r}^a$  reduces to  $T_{1,r}^a$  given in expression (3). Its bias and MSE are

$$B [T_{1,r}^a] = \left(\frac{\bar{X}_a}{\bar{X}_W}\bar{Y} - \bar{Y}_a\right) + \left(\frac{\bar{X}_a}{\bar{X}_W}\bar{Y}\right) r\bar{X} \left(\frac{1}{\bar{X}_W}\right) \left\{ V_{11} - r\bar{X} \left(\frac{1}{\bar{X}_W}\right) V_{02} \right\}. \quad (17)$$

$$M [T_{1,r}^a] = \left(\frac{\bar{X}_a}{\bar{X}_W}\bar{Y} - \bar{Y}_a\right)^2 + \left(\frac{\bar{X}_a}{\bar{X}_W}\bar{Y}\right)^2 V_{20} \\ + \left(\frac{\bar{X}_a}{\bar{X}_W}\bar{Y}\right) r^2 \bar{X}^2 \left(\frac{1}{\bar{X}_W}\right)^2 \left[ \left(\frac{\bar{X}_a}{\bar{X}_W}\bar{Y}\right) + 2 \left\{ \frac{\bar{X}_a}{\bar{X}_W}\bar{Y} - \bar{Y}_a \right\} \right] V_{02} \\ - 2 \left(\frac{\bar{X}_a}{\bar{X}_W}\bar{Y}\right) r\bar{X} \left(\frac{1}{\bar{X}_W}\right) \left\{ 2 \frac{\bar{X}_a}{\bar{X}_W}\bar{Y} - \bar{Y}_a \right\} V_{11}, \quad (18)$$

where  $\bar{X}_W = r\bar{X} + (1-r)\bar{X}_a$ .

**Remark 5.** Replacing  $r$  by  $\lambda$  (a constant), the bias and MSE of the estimator  $T_{W,a}$ , given in the expression (6), can be obtained from the expressions (17) and (18) respectively.

**Case II:**  $\alpha = 4$ . Then  $T_{4,r}^a = \bar{y}$ , the simple synthetic estimator, which does not utilize the auxiliary information at the estimation stage. Its bias and MSE are respectively

$$B [T_{4,r}^a] = (\bar{Y} - \bar{Y}_a) \quad (19)$$

$$M [T_{4,r}^a] = (\bar{Y} - \bar{Y}_a)^2 + \bar{Y}^2 V_{20}. \quad (20)$$

**Case III:**  $\alpha = 2$ . Then the bias and MSE of the product-type synthetic estimator,  $T_{2,r}^a$  are given by

$$B [T_{2,r}^a] = \left\{ \frac{\bar{X}_W}{\bar{X}_a}\bar{Y} - \bar{Y}_a \right\} + \left(\frac{\bar{X}_W}{\bar{X}_a}\bar{Y}\right) r\bar{X} \left(\frac{1}{\bar{X}_W}\right) V_{11} \quad (21)$$

$$M [T_{2,r}^a] = \left(\frac{\bar{X}_W}{\bar{X}_a}\bar{Y} - \bar{Y}_a\right)^2 + \left(\frac{\bar{X}_W}{\bar{X}_a}\bar{Y}\right)^2 V_{20} \\ + \left(\frac{\bar{X}_W}{\bar{X}_a}\bar{Y}\right)^2 r^2 \bar{X}^2 \left(\frac{1}{\bar{X}_W}\right)^2 V_{02} \\ + 2 \left(\frac{\bar{X}_W}{\bar{X}_a}\bar{Y}\right) r\bar{X} \left(\frac{1}{\bar{X}_W}\right) \left\{ 2 \left(\frac{\bar{X}_W}{\bar{X}_a}\bar{Y}\right) - \bar{Y}_a \right\} V_{11}. \quad (22)$$

**Case IV:** Similarly letting  $\alpha = 3$  and observing that  $Q_1 = 2\{(1-f)\bar{X}_a - rf(\bar{X} - \bar{X}_a)\}$ ,  $Q_2 = 2(1-f)\bar{X}_a$ ,  $C = 0$  and  $fB = -2f$ , the bias and MSE expressions of the estimator  $T_{3,r}^a$  can be deduced respectively from (15) and (16).

## 8 Optimization of $M [T_{\alpha,r}^a]$

The expression for  $M [T_{\alpha,r}^a]$ , presented in the expression (16) reveals the fact that it would be a function of the unknown parameter  $\alpha$ . Some specific choices of  $\alpha$  have already been considered and corresponding estimators, their biases and MSEs are obtained under sections 4 and 7. Besides this, it is quite reasonable to locate the estimator in the class of synthetic estimators generated by  $T_{\alpha,r}^a$  which possesses minimum MSE, if it exist. Fortunately, since the defined estimator has been observed to be asymptotically convergent over the parametric space, the optimum estimator could be obtained by minimizing the MSE of  $T_{\alpha,r}^a$ , with respect to  $\alpha$ . Although, from the expression of  $M [T_{\alpha,r}^a]$ , it is evident that  $\frac{\partial M [T_{\alpha,r}^a]}{\partial \alpha} = 0$  would not produce explicit solution for optimum  $\alpha (\alpha_0)$ . Hence, the value of  $\alpha_0$  could be obtained by iteration method for a given set of real data, which would not be difficult with the help of fast-operating computers and appropriately developed programme. It is further noted from the expression (16) that there would be more than one optimum values of  $\alpha$ , some of which might be real, imaginary and/or negative. Since  $\alpha > 0$ , only the real positive values of  $\alpha_0$  would be considered.

**Remark 6.** If for any empirical data, one gets more than one  $\alpha_0$ , for which minimum  $M [T_{\alpha,r}^a]$  are same, the appropriate choice of  $\alpha_0$  among these values of  $\alpha$  would be that which yields the smallest value of  $B [T_{\alpha,r}^a]$ . This is an additional advantage of FTE, which other one-parameter families do not exhibit.

## 9 An empirical study and efficiency comparison

**9.1** In order to demonstrate what have been discussed so far, we have made use of MU284 population given in Appendix B of Sarndal *et al*[9]. The data presented a number of characteristics of 284 municipalities in Sweden spread over four major regions: North, South, East and West. For the application purpose, we have considered only the east, central and south regions (regions indicators: 1, 2, 3, 6, 7 and 8). The mixed populations of these regions was considered as the target population with sizes 190, comprising of 6 small domains 1, 2, 3, 6, 7, and 8 with sizes 25, 48, 32, 41, 15 and 29 respectively. We consider the following characteristics as study and auxiliary variables:

$Y$ : the total number of seats in municipal council.

$X$ : the number of conservative seats in municipal council.

For the entire population of size 190 and the six small area, the following values were obtained:

**Table 1: Population and Domain Values**

Population Values						
$N$	$\bar{Y}$	$\bar{X}$	$S_Y^2$	$S_X^2$	$S_{XY}$	$\rho_{XY}$
190	47.69	8.3	137.71	26.82	41.94	0.69
Domain value	Domain					
	1	2	3	6	7	8
$N_a$	25	48	32	41	15	29
$\bar{Y}_a$	51.16	47.66	50.25	46.56	54.2	40.17
$\bar{X}_a$	16	8.1	9.5	6.73	6.06	4.04
$S_{Y_a}^2$	197.97	166.35	106.77	67.7	130.17	99.29
$S_{X_a}^2$	36	23.2	9.35	8.8	8.06	4.85
$S_{X_a Y_a}$	61.25	55.75	26.38	14.87	25.91	18.48
$\rho_{Y_a X_a}$	0.726	0.898	0.835	0.609	0.799	0.842

**9.2** On the basis of the empirical data, we shall now present a comparison of the suggested class of synthetic estimator with other synthetic estimators in terms of its efficiency. As it is evident, the suggested estimator  $T_{\alpha,r}^a$  depends upon two factors: one, on the values of the parameter  $\alpha$  and second, on the choice of the constant  $r$ ,  $0 \leq r \leq 0.5$ . Obviously, a particular choice of  $r$  fixes the size of the sample. It is, therefore, required to compare the efficiency of different estimators for a given value of  $r$ . The comparison of  $T_{\alpha,r}^a$  has been done with the following estimators:

(i) Direct estimator (Direct ratio estimator):

$$T_{D,a} = \bar{y}_a \left( \frac{\bar{X}_a}{\bar{x}_a} \right) \quad (23)$$

with

$$M [T_{D,a}] = \frac{N_a - n_a}{N_a n_a} [S_{Y_a}^2 + R_{N_a}^2 S_{X_a}^2 - 2R_{N_a} S_{Y_a X_a}]; \tag{24}$$

$$R_{N_a} = \bar{Y}_a / \bar{X}_a.$$

(ii) Indirect estimators:

(a) Ratio synthetic estimator:

$$T_{RS,a} = \bar{y} \left( \frac{\bar{X}_a}{\bar{x}} \right)$$

with

$$M [T_{RS,a}] = \left\{ \frac{\bar{X}_a}{\bar{X}} \bar{Y} - \bar{Y}_a \right\}^2 + \frac{N-n}{Nn} \left( \frac{\bar{X}_a}{\bar{X}} \bar{Y} \right) * \left[ \left( \frac{\bar{X}_a}{\bar{X}} \bar{Y} \right) \{3C_X^2 + C_Y^2 - 4C_{XY}\} - 2\bar{Y}_a \{C_X^2 - C_{XY}\} \right]. \tag{25}$$

(b) Synthetic estimator suggested by Singh and Seth[12]:

$$T_{\alpha,a} = \bar{y} \Psi \{ \alpha, \bar{X}_a, \bar{x} \}.$$

The expression of MSE of  $T_{\alpha,a}$  is presented in Singh and Seth[12].

(iii) Simple synthetic estimator:  $\bar{y}_{SS,a} = \bar{y}$

with

$$M [\bar{y}_{SS,a}] = \{ \bar{Y} - \bar{Y}_a \}^2 + \frac{N-n}{Nn} S_{\bar{Y}}^2. \tag{26}$$

Table 2 present the absolute values of MSE of the estimators  $\bar{y}_{SS,a}$ ,  $T_{D,a}$  and  $T_{RS,a}$  for the domains 1, 2, 3, 6, 7 and 8 taking  $r = 0.1, 0.2, 0.3$  and  $0.4$ , that is  $n = 21, 48, 81$  and  $127$ .

**Table 2: MSE of the estimators  $\bar{y}_{SS,a}, T_{D,a}$  and  $T_{RS,a}$  for different domains**

Estimator	r	Domain					
		1	2	3	6	7	8
$\bar{y}_{SS,a}$	0.1	17.83	5.79	12.34	7.07	48.17	62.34
	0.2	14.21	2.17	8.72	3.44	44.55	58.72
	0.3	13	0.96	7.51	2.29	43.34	57.51
	0.4	12.4	0.36	6.91	1.63	42.74	56.91
$T_{D,a}$	0.1	55.71	52.18	22.32	55.27	166.4	70.46
	0.2	20.89	19.57	8.37	20.72	62.39	26.42
	0.3	9.29	8.69	3.72	9.21	27.74	11.74
	0.4	3.53	3.31	1.41	3.51	10.57	4.47
$T_{RS,a}$	0.1	1825.8	21.59	53.32	70.58	373.25	284.5
	0.2	1718.04	8.96	31.38	65.84	375.74	286.99
	0.3	1682.12	4.74	24.07	64.26	376.56	287.82
	0.4	1664.16	2.64	20.42	63.47	376.98	288.24

**Remark 7.** We observe that  $M [T_{D,a}]$  is based on domain size ( $N_a$ ), size of the sample selected from the domain ( $n_a$ ) and domain parameters, while MSEs of  $T_{RS,a}$  and  $\bar{y}_{SS,a}$  involves population size ( $N$ ), size of the sample taken from the population ( $n$ ) and population parameters and  $\bar{Y}_a, \bar{X}_a$ . Since the value of  $r$  is computed using  $n$  and  $N$  the equivalent sample size ( $n_a$ ) for computing  $M [T_{D,a}]$  has been taken such that  $(n_a/N_a) = (n/N)$ . Further, since estimators  $T_{\alpha,r}^a$  and  $T_{\alpha,r}$  depend upon the parameter  $\alpha$ , we have compared them with respect to  $T_{D,a}$  in terms of percent relative efficiency  $\{M (T_{D,a}) / M (\cdot)\} \times 100$  for  $\alpha = 1, 3, 4$  and optimum  $\alpha (\alpha_0)$  for each of the domains. The values are depicted in Table 3.



**Table 3: Percent Relative Efficiency of  $T_{\alpha,r}^a$  and  $T_{\alpha,a}$  with respect to  $T_{D,a}$**

Estimator	$\alpha$	r	Domain					
			1	2	3	6	7	8
$T_{\alpha,r}^a$	1	0.1	865.1	1272.7	269.9	1489.7	233.5	617.5
		0.2	479.1	1591.1	269.1	1000.9	64.2	1822.1
		0.3	43.1	1424.5	361.2	205.6	21.7	73.2
		0.4	5.7	769.8	617.4	39.1	6.5	10.1
	3	0.1	351	943.6	189.5	859.5	327.8	134
		0.2	357.7	1144.2	128.8	1204.8	102.7	143
		0.3	253.8	1526.1	179.7	382.3	25	76.2
		0.4	1.6	181.1	25.86	5.6	2.5	0.4
	4	0.1	312.5	901.3	180.8	781.7	345.5	113
		0.2	147	901.7	95.9	602.4	140.1	45
		0.3	71.4	906.1	49.5	411.3	64	20.4
		0.4	28.5	920.4	20.5	215.3	24.7	7.9
	$\alpha_0$	0.1	1085.9	1683	630.5	1635.5	719.3	3139.2
		0.2	1095.7	1615.7	616.4	1572.4	630.5	3231.9
		0.3	1102.1	1518.9	613.7	1550.7	605.3	3257.9
		0.4	1126.7	1364.7	625.2	1563.3	606.5	3323.9
$T_{\alpha,a}$	1	0.1	3	241.7	41.9	78.3	44.6	24.8
		0.2	1.2	218.4	26.7	31.5	16.6	9.2
		0.3	1.1	183.5	15.5	14.3	7.4	4.1
		0.4	0.6	125.5	6.9	5.5	2.8	1.6
	3	0.1	986.1	1373.2	296	1625.5	208.6	193.3
		0.2	109.4	1449.3	529.7	250.9	39.9	20
		0.3	4.9	301	75.3	16.3	7	1.3
		0.4	0.4	24.2	1.5	1.2	0.6	0.1
	4	0.1	312.5	901.3	180.8	781.7	345.5	113
		0.2	147	901.7	95.9	602.4	140.1	45
		0.3	71.4	906.1	49.5	411.3	64	20.4
		0.4	28.5	920.4	20.5	215.3	24.7	7.9
	$\alpha_0$	0.1	1152.2	1711.8	631.7	1640.2	768.6	3248.3
		0.2	1154.1	1595.5	619.9	1586.3	690.5	3217
		0.3	1156.7	1505.9	619.7	1564.2	664.1	3202.9
		0.4	1177.3	1366.4	630.5	1570.9	661.3	3244.4

**Remark 8.** From the Table 3, some important conclusions can be drawn.

(i) The suggested synthetic estimator  $T_{\alpha,r}^a$  is always efficient than the estimator  $T_{\alpha,a}$  for  $\alpha = 1$ , that is,  $T_{1,r}^a = \bar{y} \left( \frac{\bar{X}_a}{\bar{x}_r} \right)$  is preferable over the estimator  $T_{1,a} = \bar{y} \left( \frac{\bar{X}_a}{\bar{x}} \right)$ .

(ii) It is observed that the efficiency of the estimator  $T_{\alpha,r}^a$  generally decreases as the value of r increases. This is obvious since by choosing smaller values of r we give more weight to the population mean  $\bar{X}_a$  as compared to the sample mean  $\bar{x}$ .

(iii) Both the estimators  $T_{\alpha,r}^a$  and  $T_{\alpha,a}$  are highly precise than other estimators for optimum choices of  $\alpha$ .

(iv) The estimators seem to be more efficient for domains 2, 3 and 6 than other domains. This might be due to the reason that for these domains the “synthetic assumptions”,  $(\bar{Y}/\bar{Y}_a) = (\bar{X}/\bar{X}_a)$  are closely met in comparison to other domains.

### 10 Simulation study

In order to see the performance of the suggested estimator over different samples, a simulation study has been done in this section. The same set of data has been used for the purpose.



We have selected 500 independent simple random samples of size 19 from the population of size 190. Further, to assess the relative performance of the estimators under consideration, their Absolute Relative Bias (ARB) and Simulated Relative Standard Error (SRSE) were obtained for each domain on the basis of the selected samples as follows:

$$ARB [T_{k,a}] = \frac{|\frac{1}{500} \sum_{s=1}^{500} T_{k,a}^s - \bar{Y}_a|}{\bar{Y}_a} \times 100, \tag{27}$$

$$SRSE [T_{k,a}] = \frac{\sqrt{SMSE (T_{k,a})}}{\bar{Y}_a} \times 100, \tag{28}$$

where

$$SMSE [T_{k,a}] = \frac{1}{500} \sum_{s=1}^{500} (T_{k,a}^s - \bar{Y}_a)^2. \tag{29}$$

$T_{k,a}$  denote a particular synthetic estimator for domain 'a' and  $T_{k,a}^s$  stands for the value of  $T_{k,a}$  for domain for the  $s^{th}$  sample, where a= 1, 2, 3, 6, 7, and 8.

The values of ARB and SRSE of the estimators  $T_{RS,a}$ ,  $T_{\alpha,a}$  and  $T_{\alpha,r}^a$  along with the value of optimum  $\alpha$  for each of the domain are presented in the following table.

**Table 4: ARB and SRSE (in percent) of different estimators**

Estimator		Domain					
		1	2	3	6	7	8
$T_{RS,a}$	SRSE	68.27	80.61	27.8	78.21	150.2	111.53
	ARB	3.05	3.6	1.24	3.5	6.72	4.99
$T_{1,a}$	SRSE	1814.56	34.85	212.55	363.94	787.94	927.01
	ARB	81.15	1.56	9.51	16.28	35.24	41.46
$T_{1,0.1}^a$	SRSE	46.85	5.5	87.91	0.39	340.25	165.04
	ARB	2.09	0.25	3.93	0.02	15.22	7.38
$T_{\alpha_0,a}$	$\alpha_0$	1.89	1.85	1.83	2.95	1.9	1.86
	SRSE	5.68	12.55	12.81	22.53	102.56	11.82
	ARB	0.25	0.56	0.57	1.01	4.59	0.53
$T_{\alpha_0,0.1}^a$	$\alpha_0$	2.21	2.14	2.11	2.22	2.03	2.19
	SRSE	1.88	5.16	7.92	16.77	105.85	1.81
	ARB	0.08	0.23	0.35	0.75	4.73	0.08

The value of SRSE of  $T_{\alpha,r}^a$  for each domain are observed smaller than those of  $T_{\alpha,a}$  showing that the suggested class of synthetic estimators could be preferred over the estimator  $T_{\alpha,a}$ .

## 11 Conclusions

From what we discussed in the above sections, it can be concluded that the suggested factor-type synthetic estimator defines a class of estimators which includes some well-known synthetic estimators as members of class and hence provides a unified way of discussing the properties of such estimators. Further, on the basis of the empirical data and simulation study, it can be concluded that the suggested estimator outperforms other synthetic estimators.

## Acknowledgement

The authors are grateful to the referees for referring the paper and for providing useful suggestions.

## Appendix

### Bias and MSE of $T_{\alpha,r}^a$

1. Since  $\bar{y}$  and  $\bar{x}$  are unbiased estimators of population means  $\bar{Y}$  and  $\bar{X}$  respectively, we can write

$$\bar{y} = \bar{Y}(1 + e_1), \bar{x} = \bar{X}(1 + e_2)$$

such that  $E(e_1) = E(e_2) = 0$  and

$$E(e_1^2) = \frac{V(\bar{y})}{\bar{Y}^2} = \frac{N-n}{Nn} C_Y^2 = V_{20}, E(e_2^2) = \frac{V(\bar{x})}{\bar{X}^2} = \frac{N-n}{Nn} C_X^2 = V_{02}$$

$$E(e_1 e_2) = \frac{Cov(\bar{y}, \bar{x})}{\bar{Y}\bar{X}} = \frac{N-n}{Nn} C_{YX} = V_{11}.$$

Expressing  $T_{\alpha,r}^a$  in terms of  $e_1$  and  $e_2$  we have

$$T_{\alpha,r}^a = \bar{Y}(1 + e_1) \frac{Q_1}{Q_2} \left[ \frac{(1 + frB\bar{X}e_2/Q_1)}{(1 + rC\bar{X}e_2/Q_2)} \right]. \quad (A1)$$

Now realizing that  $|\frac{rC\bar{X}e_2}{Q_2}| < 1$  and expanding the denominator of (A1) upto the second power of  $e_2$ , we get

$$T_{\alpha,r}^a = \bar{Y} \frac{Q_1}{Q_2} \left[ (1 + e_1) \left( 1 + \frac{frB\bar{X}}{Q_1} e_2 \right) \left\{ 1 - \frac{rC\bar{X}e_2}{Q_2} + \left( \frac{rC\bar{X}}{Q_2} \right)^2 e_2^2 \right\} \right]. \quad (A2)$$

Simplifying the above expression, taking expectation of both the sides retaining terms upto the second powers of  $e_1$  and  $e_2$ , using the expression of  $E(e_1^2)$ ,  $E(e_2^2)$ ,  $E(e_1 e_2)$  and observing that  $B[T_{\alpha,r}^a] = E[T_{\alpha,r}^a] - \bar{Y}_a$ , we get the expression of  $B[T_{\alpha,r}^a]$  as given in (15).

2. Further, we have

$$M[T_{\alpha,r}^a] = E[T_{\alpha,r}^a - \bar{Y}_a]^2 = E[T_{\alpha,r}^a]^2 + \bar{Y}_a^2 - 2\bar{Y}_a E[T_{\alpha,r}^a]. \quad (A3)$$

Now

$$E[T_{\alpha,r}^a]^2 = E \left[ \bar{Y}(1 + e_1) \frac{Q_1}{Q_2} \left\{ \frac{(1 + frB\bar{X}e_2/Q_1)}{(1 + rC\bar{X}e_2/Q_2)} \right\} \right]^2.$$

Expanding the right hand side and retaining terms up to the order  $O(n^{-1})$ , taking expectation of the expression (A2) and substituting the values in (A3), the expression (16) can easily be deduced.

## References

- [1] Gonzalez, M.E.(1973): Use and evaluation of synthetic estimates, Proceedings of the Social Statistics, American Statistical Association, 33-36.
- [2] Gonzalez, M.E. and Wakesberg, J.(1973): Estimation of the error of synthetic estimates, Paper presented at the first meeting of the International Association of Survey Statisticians, Vienna, Austria.
- [3] Holt, D., Smith, T.M.F. and Tomberlin, T.J.(1979): A model-based approach to estimation for small groups of population. *Journal of American Statistical Association.*, 74, 405-410.
- [4] National Centre for Health Statistics(1968): Synthetic State Estimates of Disability, P.H.S. Publication 1759, Washington, DC: U.S. Government Printing Office.
- [5] Pandey, K.K. and Tikkiwal, G.C.(2010): Generalized class of synthetic estimators for small areas under systematic sampling scheme, *Statistics in Transition*, 11(1), 75-89.
- [6] Rai, P.K. and Pandey, K.K.(2013): Synthetic estimators using auxiliary information in small domains, *Statistics in Transition*, 14(1), 31-44.
- [7] Reddy, V.N.(1973): On ratio and product methods of estimation, *Sankhya*, B, 35, 307-317.
- [8] Rao, J.N.K.(2003): Small Area Estimation, Wiley Inter-science, John Wiley and Sons, New Jersey.
- [9] Sarndal, C.E., Swensson, B. and Wretman, J.H.(1992): Model Assisted Survey Sampling, Springer-Verlog, New York.
- [10] Schaible, W.L., Brock, D.B., Casady, R.J. and Schnack, G.A.(1977): An empirical comparison of the simple synthetic and composite estimators for small area statistics, *Proceedings of American Statistical Association, Social Statistics Section*, 1017-1021.
- [11] Shukla, D.(1988): On Some Factor-Type Estimators for Population Parameter in Sample Surveys, *Unpublished Ph.D. Thesis* submitted to Banaras Hindu University, Varanasi, India.

- [12] Singh, V.K. and Seth, S.K. (2014): On the use of one parameter family of synthetic estimators for small areas, *Journal of Statistics Applications and Probability*, 3(3), 1-9.
- [13] Singh, V.K. and Shukla, D.(1993): An efficient one-parameter family of factor-type estimators in sample survey, *Metron*, 51(1-2), 139-159.
- [14] Srivenkataramana, T.(1980): A Dual to ratio estimator in sample surveys, *Biometrika*, 67(1), 199-204.
- [15] Tikkiwal, G.C. and Ghiya, A.(2000): A generalised class of synthetic estimators with application to crop acreage estimation for small domains, *Biometrical Journal*, 42(7), 865-876.
- [16] Tikkiwal, G.C. and Pandey, K.K.(2007): On synthetic and composite estimators for small area estimation under Lahiri-Midzuno sampling scheme, *Statistics in Transition*, 8(1), 111-123.
- [17] Walsh, J.E. (1971): Generalization of ratio estimator for population total, *Sankhya A*,42,99-106.



**V. K. Singh** presently working as Professor of Statistics, Department of Statistics, Faculty of Science, Banaras Hindu University, Varanasi, India since 2000. Joined the Department as Assistant Professor in 1972. Did my M. Sc. ( Statistics ) and Ph. D. ( Statistics ) from Banaras Hindu University in 1972 and 1979 respectively. Having 42 years teaching experience and 40 years research experience. Field of specializations are Sampling Theory, Stochastic Modelling, Mathematical Demography and Operations Research. Published 78 research papers in reputed international/national journals. Guided 12 Ph. D. scholars for their Ph. D. Degree. Visited United Kingdom, Australia and Sri Lanka for attending International Conferences and organizing Symposiums. Convened 2 national conferences. Life member of Indian Statistical Association,

Member of International Association of Survey Statisticians ( IASS ), Associate Editor of Assam Statistical Review, India.



**Sunil Kumar Seth** pursuing PhD degree in statistics from Department of Statistics, Banaras Hindu University Varanasi-221005, India. On the topic Small Area Estimation. He has done my M.Sc. in Statistics and P.G.D.C.A. from V.B. S. Purvanchal University Jaunpur, UP, India.