Rajesh Singh
Editor

Florentin Smarandache
Editor

THE EFFICIENT USE OF SUPPLEMENTARY INFORMATION IN FINITE POPULATION SAMPLING

Estimators	Values of α_1	Values of α_2	$PRE(\bar{y}_i)$
$\bar{y}_{st}$	0	0	100
$\bar{y}_1$	1	0	1029.469
$\bar{y}_5$	1	1	149.686
$\bar{y}_8$	1	1	115.189
$MSE(\bar{y}_9)_{min}$	6.2918	-0.8870	2854.549

2014

THE EFFICIENT USE OF SUPPLEMENTARY INFORMATION IN FINITE POPULATION SAMPLING

Rajesh Singh

Department of Statistics, BHU, Varanasi (U.P.), India

Editor

Florentin Smarandache

Department of Mathematics, University of New Mexico, Gallup, USA

Editor

2014

Education Publishing
1313 Chesapeake Avenue
Columbus, Ohio 43212
USA
Tel. (614) 485-0721

Peer-Reviewers:
Dr. A. A. Salama, Faculty of Science, Port Said University, Egypt.
Said Broumi, Univ. of Hassan II Mohammedia, Casablanca, Morocco.
Pabitra Kumar Maji, Math Department, K. N. University, WB, India.
Mumtaz Ali, Department of Mathematics, Quaid-i-Azam University, Islamabad, 44000, Pakistan.

EAN: 9781599732756
ISBN: 978-1-59973-275-6

Contents

Preface

The purpose of writing this book is to suggest some improved estimators using auxiliary information in sampling schemes like simple random sampling, systematic sampling and stratified random sampling.

This volume is a collection of five papers, written by nine co-authors (listed in the order of the papers): Rajesh Singh, Mukesh Kumar, Manoj Kr. Chaudhary, Cem Kadilar, Prayas Sharma, Florentin Smarandache, Anil Prajapati, Hemant Verma, and Viplav Kr. Singh.

In first paper dual to ratio-cum-product estimator is suggested and its properties are studied. In second paper an exponential ratio-product type estimator in stratified random sampling is proposed and its properties are studied under second order approximation. In third paper some estimators are proposed in two-phase sampling and their properties are studied in the presence of non-response.

In fourth chapter a family of median based estimator is proposed in simple random sampling. In fifth paper some difference type estimators are suggested in simple random sampling and stratified random sampling and their properties are studied in presence of measurement error.

The authors hope that book will be helpful for the researchers and students who are working in the field of sampling techniques.

Dual To Ratio Cum Product Estimator In Stratified Random Sampling

†Rajesh Singh, Mukesh Kumar and Manoj K. Chaudhary

Department of Statistics, B.H.U., Varanasi (U.P.), India

Cem Kadilar
Department of Statistics, Hacettepe University,
Beytepe 06800, Ankara, Turkey
† Corresponding author, rsinghstat@yahoo.com

Abstract

Tracy et al.[8] have introduced a family of estimators using Srivenkataramana and Tracy ([6],[7]) transformation in simple random sampling. In this article, we have proposed a dual to ratio-cum-product estimator in stratified random sampling. The expressions of the mean square error of the proposed estimators are derived. Also, the theoretical findings are supported by a numerical example.

Key words: Auxiliary information, dual, ratio-cum-product estimator, stratified random sampling, mean square error and efficiency.

1. Introduction

In planning surveys, stratified sampling has often proved as useful in improving the precision of un-stratified sampling strategies to estimate the finite population mean of the study variable, $\overline{Y} = \frac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{N_h} y_{hi}$. Let y, x and z respectively, be the study and auxiliary variates on each unit U_h (h=1,2,3, ---, N) of the population U. Here the size of the stratum U_h is N_h, and the size of simple random sample in stratum U_h is n_h, where h=1, 2,---,L. In this study, under stratified random sampling without replacement scheme, we suggest estimators to estimate $\overline{Y}$ by considering the estimators in Plikusas [3] and in Tracy et al. [8].

To obtain the bias and MSE of the proposed estimators, we use the following notations in the rest of the article:

$$\bar{y}_{st} = \sum_{h=1}^{L} w_h \bar{y}_h = \bar{Y}(1 + e_0),$$

$$\bar{x}_{st} = \sum_{h=1}^{L} w_h \bar{x}_h = \bar{X}(1 + e_1),$$

$$\bar{z}_{st} = \sum_{h=1}^{L} w_h \bar{z}_h = \bar{Z}(1 + e_2),$$

where, $w_h = \dfrac{N_h}{N}$.

Such that,

$$E(e_0) = E(e_1) = E(e_2) = 0,$$

$$V_{rst} = \sum_{h=1}^{L} w_h^{r+s+t} \frac{E[(\bar{y}_h - \bar{Y}_h)^r (\bar{x}_h - \bar{X}_h)^s (\bar{z}_h - \bar{Z}_h)^t]}{\bar{Y}^r \bar{X}^s \bar{Z}^t}, \qquad (1)$$

where $\bar{y}_h$ and $\bar{Y}_h$ are the sample and population means of the study variable in the stratum h, respectively. Similar expressions for X and Z can also be defined.

Using (1), we can write

$$E(e_0^2) = \frac{\sum_{h=1}^{L} w_h^2 \gamma_h S_{yh}^2}{\bar{Y}^2} = V_{200},$$

$$E(e_1^2) = \frac{\sum_{h=1}^{L} w_h^2 \gamma_h S_{xh}^2}{\bar{X}^2} = V_{020},$$

$$E(e_2^2) = \frac{\sum_{h=1}^{L} W_h^2 \gamma_h S_{zh}^2}{\bar{Z}^2} = V_{002},$$

$$E(e_0 e_1) = \frac{\sum_{h=1}^{L} W_h^2 \gamma_h S_{xyh}}{\bar{X}\bar{Y}} = V_{110},$$

$$E(e_1 e_2) = \frac{\sum_{h=1}^{L} W_h^2 \gamma_h S_{xzh}}{\bar{X}\bar{Z}} = V_{011},$$

where

$$S_{yh}^2 = \frac{\sum_{i=1}^{N_h} (y_h - \bar{Y}_h)^2}{N_h - 1}, \qquad S_{xh}^2 = \frac{\sum_{i=1}^{N_h} (x_h - \bar{X}_h)^2}{N_h - 1}, \quad S_{zh}^2 = \frac{\sum_{i=1}^{N_h} (z_h - \bar{Z}_h)^2}{N_h - 1},$$

$$S_{xyh} = \frac{\sum_{i=1}^{N_h} (x_h - \bar{X}_h)(y_h - \bar{Y}_h)}{N_h - 1}, \qquad S_{yzh} = \frac{\sum_{i=1}^{N_h} (y_h - \bar{Y}_h)(z_h - \bar{Z}_h)}{N_h - 1},$$

$$S_{xzh} = \frac{\sum_{i=1}^{N_h} (x_h - \bar{X}_h)(z_h - \bar{Z}_h)}{N_h - 1}, \qquad \gamma_h = \frac{1 - f_h}{n_h}, \quad f_h = \frac{n_h}{N_h}, \qquad W_h = \frac{N_h}{N}.$$

The combined ratio and the combined product estimators are, respectively, defined as

$$\bar{y}_1 = \bar{y}_{st}\left(\frac{\bar{X}}{\bar{x}_{st}}\right), \tag{2}$$

$$\bar{y}_2 = \bar{y}_{st}\left(\frac{\bar{z}_{st}}{\bar{Z}}\right) \tag{3}$$

And the MSE of $\bar{y}_1$ and $\bar{y}_2$ to the first degree of approximation are, respectively, given by

$$MSE(\bar{y}_1) \cong \bar{Y}^2 (V_{200} + V_{020} - 2V_{110}) \tag{4}$$

$$MSE(\bar{y}_2) \cong \bar{Y}^2 (V_{200} + V_{020} + 2V_{110}) \tag{5}$$

Note that $\bar{Y} = \bar{Y}_{st} = \sum_{h=1}^{L} W_h \bar{Y}_h$. Similar expressions for X and Z can also be defined.

2. Classical Estimators

Srivenkataramana and Tracy ([6],[7]) considered a simple transformation as

$$u_i = A - x_i, \qquad (\text{ i=1, 2,…,N})$$

$$\Rightarrow \bar{u} = A - \bar{x},$$

where A is a scalar to be chosen. This transformation renders the situation suitable for a product method instead of ratio method. Clearly $\bar{u}_{st}(= A - \bar{x}_{st})$ is unbiased for $\bar{U}(= A - \bar{X})$. Using this transformation, an estimator in the stratified random sampling is defined as

$$\bar{y}_3 = \bar{y}_{st}\left(\frac{\bar{u}_{st}}{\bar{U}}\right) \tag{6}$$

This is a product type estimator (alternative to combined ratio type estimator) in stratified random sampling.

The exact expression for MSE of $\bar{y}_3$ is given by

$$MSE(\bar{y}_3) = \bar{Y}^2(V_{200} + \theta^2 V_{020} - 2\theta V_{110}) \tag{7}$$

where $\theta = \dfrac{\bar{X}}{(A - \bar{X})}$.

In some survey situations, information on a second auxiliary variable, Z, correlated negatively with the study variable, Y, is readily available. Let $\bar{Z}$ be the known population mean of Z. To estimate $\bar{Y}$, Singh[4] considered ratio-cum-product estimator as $\bar{y}_4 = \bar{y}\left(\frac{\bar{X}}{\bar{x}}\right)\left(\frac{\bar{z}}{\bar{Z}}\right)$,

where Perri[2] used $\bar{t}_x = \bar{x} + \alpha(\bar{X} - \bar{x})$ and $\bar{t}_z = \bar{z} + \beta(\bar{Z} - \bar{z})$ instead of $\bar{x}$ and $\bar{z}$, respectively. Here, α and β are constants that make the MSE minimum.

Adapting $\bar{y}_4$ to the stratified random sampling, the ratio cum product estimator using two auxiliary variables can be defined as

$$\bar{y}_5 = \bar{y}_{st}\left(\frac{\bar{X}}{\bar{x}_{st}}\right)\left(\frac{\bar{z}_{st}}{\bar{Z}}\right) \qquad (8)$$

The approximate MSE of this estimator is

$$MSE(\bar{y}_5) \cong \bar{Y}^2[V_{200} + V_{020} + V_{002} + 2(V_{101} - V_{110} - V_{011})] \qquad (9)$$

3. Suggested Estimators

Tracy et al. [8] introduced a product estimator using two auxiliary variables in the simple random sampling given by

$$\bar{y}_6 = \bar{y}\left(\frac{\bar{u}}{\bar{U}}\right)\left(\frac{\bar{z}}{\bar{Z}}\right) \qquad (10)$$

Motivated by Tracy et al. [8], we propose the following product estimator for the stratified random sampling scheme as

$$\bar{y}_7 = \bar{y}_{st}\left(\frac{\bar{u}_{st}}{\bar{U}}\right)\left(\frac{\bar{z}_{st}}{\bar{Z}}\right) \qquad (11)$$

Expressing $\bar{y}_7$ in terms of e's, we can write (11) as

$$\bar{y}_7 = \bar{Y}(1 + e_0)(1 - \theta e_1)(1 + e_2)$$

The $MSE(\bar{y}_7)$ to the first order of approximation, is given as

$$MSE(\bar{y}_7) = \bar{Y}^2[V_{200} + \theta^2 V_{020} + V_{002} - 2(\theta V_{110} - V_{101} + \theta V_{011})] \qquad (12)$$

and this MSE equation is minimised for

$$\theta = \frac{V_{110} + V_{011}}{V_{020}} = \theta_{opt}(say)$$

Note that the corresponding A is

$$A_{opt} = \frac{(1-\theta_{opt})\bar{X}}{\theta_{opt}},$$

By putting the optimum value of θ in (12), we can obtain the minimum MSE equation for the first proposed estimator, $\bar{y}_7$.

Remark 3.1 : The value of $\bar{X}$ is known, but the exact values of V_{110}, V_{011} and V_{020} are rarely available in practice. However in repeated surveys or studies based on multiphase sampling, where information is gathered on several occasions it may be possible to guess the values of V_{110}, V_{011} and V_{020} quite accurately. Even though this approach may reduce the precision, it may be satisfactory provided the relative decrease in precision is marginal, see Tracy et al. [8].

Plikusas[3] defined dual to ratio cum product estimator in stratified random sampling as

$$\bar{y}_{Pl} = \bar{y}_{st}\frac{\bar{x}^{*}_{st}}{\bar{X}} \cdot \frac{\bar{Z}}{\bar{z}^{*}_{st}} \qquad (13)$$

where

$$\bar{x}^{*}_{st} = (1+g_h\bar{X}_h) - g_h\bar{x}_h, \quad \bar{z}^{*}_{st} = (1+g_h\bar{Z}_h) - g_h\bar{z}_h.$$

and $g_h = \frac{n_h}{(N_h - n_h)}$.

Considering the estimator in (13) and motivated by Singh et al. [5], we propose a family of dual to ratio cum product estimator as –

$$\bar{y}_p = \bar{y}_{st}\left(\frac{\bar{x}^*_{st}}{\bar{X}}\right)^{\alpha_1}\left(\frac{\bar{Z}}{\bar{z}^*_{st}}\right)^{\alpha_2} \tag{14}$$

To obtain the MSE of the second proposed estimator, $\bar{y}_9$, we write

$$\bar{y}_{st} = \bar{Y}(1+e_0),$$

$$\bar{x}^*_{st} = (1+e_1^*),$$

$$\bar{z}^*_{st} = (1+e_2^*).$$

Expressing (14) in terms of e's, we have

$$\bar{y}_p = \bar{Y}(1+e_0)(1+e_1')^{\alpha_1}(1+e_2')^{-\alpha_2} \tag{15}$$

Expanding the right hand side of (15), to the first order of approximation, we get

$$(\bar{y}_p - \bar{Y}) = \bar{Y}[e_0 + \alpha_1 e_0 e_1 - \alpha_2 e_0 e_2' + \alpha_1 e_1' - \alpha_2 e_2' - \alpha_1\alpha_2 e_1' e_2'$$

$$+ \frac{\alpha_1(\alpha_1 - 1)}{2} e_1'^2 + \frac{\alpha_2(\alpha_2 + 1)}{2} e_2'^2] \tag{16}$$

Squaring both sides of (16) and then taking expectation, we obtain the MSE of the second proposed estimator, $\bar{y}_9$, to the first order approximation, as

$$MSE(\bar{y}_p) \cong \bar{Y}^2\{V'_{200} + \alpha_1^2 V'_{020} + \alpha_2^2 V'_{002} + 2(\alpha_1 V'_{110} - \alpha_1\alpha_2 V'_{011} - \alpha_2 V'_{101})\} \tag{17}$$

where

$$V'_{rst} = \sum_{h=1}^{L} \frac{W^{r+s+t}(-g)^{s+t} E(\bar{y}_h - \bar{Y}_h)(\bar{x}_h - \bar{X}_h)(\bar{z}_h - \bar{Z}_h)}{\bar{Y}^r \bar{X}^s \bar{Z}^t} \tag{18}$$

This MSE equation is minimized for the optimum values of α_1 and α_2 given by

$$\alpha_1' = \frac{V'_{101}V'_{011} - V'_{110}V'_{002}}{V'_{020}V'_{002} - V'^2_{011}} \tag{19}$$

$$\alpha_2' = \frac{V_{020}' V_{101}' - V_{110}' V_{011}'}{V_{020}' V_{002} - V_{011}^2} \tag{20}$$

Putting these values of α_1' and α_2' in MSE $(\bar{y}_9)$, given in (17), we obtain the minimum MSE of the second proposed estimator, $\bar{y}_9$.

4. Theoretical Efficiency Comparisons

In this section, we first compare the efficiency between the first proposed estimator, $\bar{y}_7$, with the classical combined estimator, $\bar{y}_{st}$, as follows:

$$MSE(\bar{y}_7) < V(\bar{y}_{st})$$

$$\bar{Y}^2[V_{200} + \theta^2 V_{020} + V_{002} - 2(\theta V_{110} - V_{101} + \theta V_{011})] < \bar{Y}^2 V_{200}.$$

The estimator $\bar{y}_7$ is better than the usual estimator $\bar{y}_{st}$, if and only if,

$$\frac{B_1}{2B_2} < 1, \tag{21}$$

where, $B_1 = \theta^2 V_{020} + V_{002}$ and $B_2 = \theta V_{110} - V_{101} + \theta V_{011}$.

If the condition (21) is satisfied, the first proposed estimator, $\bar{y}_7$, performs better than the classical combined estimator.

We also find the condition under which the second proposed estimator, $\bar{y}_9$, performs better than the classical combined estimator in theory as follows:

$$MSE(\bar{y}_9) < V(\bar{y}_{st}),$$

$$\bar{Y}^2\{V_{200}' + \alpha_1^2 V_{020}' + \alpha_2^2 V_{002}' + 2(\alpha_1 V_{110}' - \alpha_1\alpha_2 V_{011}' - \alpha_2 V_{101}')\} < \bar{Y}^2 V_{200},$$

$$\frac{\alpha_1^2 V'_{020} + \alpha_2^2 V'_{002} - 2\alpha_1\alpha_2 V'_{011}}{\alpha_2 V'_{101} - \alpha_1 V'_{110}} < 1,$$

The estimator $\bar{y}_9$ is better than the usual estimator $\bar{y}_{st}$, if and only if,

$$\frac{C}{D} < 1, \tag{22}$$

where, $C = \alpha_1^2 V'_{020} + \alpha_2^2 V'_{002} - 2\alpha_1\alpha_2 V'_{011}$ and $D = \alpha_2 V'_{101} - \alpha_1 V'_{110}$.

5. Numerical Example

In this section, we use the data set earlier used in Koyuncu and Kadilar[1] . The population statistics are given in Table 1. In this data set, the study variable (Y) is the number of teachers, the first auxiliary variable (X) is the number of students, and the second auxiliary variable (Z) is the number of classes in both primary and secondary schools for 923 districts at 6 regions (as 1: Marmara, 2: Agean, 3: Mediterranean, 4: Central Anatolia, 5: Black Sea, 6: East and South east Anatolia) in Turkey in 2007, see Koyuncu and Kadilar[1]. Koyuncu and Kadilar[1] have used Neyman allocation for allocating the samples to different strata. Note that all correlations between the study and auxiliary variables are positive. Therefore, we decide not to use product estimators for this data set for efficiency comparison. For this reason, we apply the classical combined estimator, $\bar{y}_{st}$, combined ratio estimator , $\bar{y}_1$, the ratio-cum-product estimator, $\bar{y}_5$, Plikusas [3] estimator, $\bar{y}_8$, and the second proposed estimator, $\bar{y}_9$, to the data set. For the efficiency comparison, we compute percent relative efficiencies as

$$PRE(\bar{y}_i) = \frac{MSE(\bar{y}_{st})}{MSE(\bar{y}_i)} \times 100, \qquad i = st, 1, 5, 8, 9.$$

Table 1. Data Statistics of Population

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
S_{y1} = 883.835	S_{y2}= 644	S_{y3}= 1033.467
S_{y4} = 810.585	S_{y5}= 403.654	S_{y6}=711.723
$\bar{Y}_1$ = 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
S_{x1} =30486.751	S_{x2} =15180.760	S_{x3} =27549.697
S_{x4} =18218.931	S_{x5} =8997.776	S_{x6} =23094.141
$\bar{X}_1$ =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
S_{xy1} =25237153.52	S_{xy2} =9747942.85	S_{xy3} =28294397.04
S_{xy1} = 14523885.53	S_{xy1} =3393591.75	S_{xy6} =15864573.97
ρ_{xy1} = 0.936	ρ_{xy2} = 0.996	ρ_{xy3} =0.994
ρ_{xy4} = 0.983	ρ_{xy5} = 0.989	ρ_{xy6} = 0.965
S_{z1} = 555.5816	S_{z2} = 365.4576	S_{z3} =612.9509281
S_{z4} = 458.0282	S_{z5} = 260.8511	S_{z6} = 397.0481
Z_1 = 498.28	Z_2= 318.33	Z_3 = 431.36
$\bar{Z}_4$ = 498.28	$\bar{Z}_5$ = 227.20	$\bar{Z}_6$= 313.71
S_{yz1} = 480688.2	S_{yz2} = 230092.8	S_{yz1} = 623019.3
S_{yz1} = 364943.4	S_{yz1} = 101539	S_{yz1} = 277696.1
S_{xz1} = 15914648	S_{xz2} = 5379190	S_{xz3} = 164900674.56
S_{xz4} = 8041254	S_{xz5} = 2144057	S_{xz1} = 8857729

ρ_{yz1} = 0.978914 ρ_{yz2} = 0.9762 ρ_{y3} = 0.983511

ρ_{yz4} = 0.982958 ρ_{yz5} = 0.964342 ρ_{yzi} = 0.982689

Table 2. Percent Relative Efficiencies (PRE) of estimators

Estimators	Values of α_1	Values of α_2	PRE($\bar{y}_i$)
$\bar{y}_{st}$	0	0	100
$\bar{y}_1$	1	0	1029.469
$\bar{y}_5$	1	1	149.686
$\bar{y}_3$	1	1	115.189
MSE($\bar{y}_9$)$_{min}$	6.2918	-0.8870	2854.549

Table3. The MSE values according to A

Value of θ	**Corresponding value of A**	**MSE($\bar{y}_7$)**
<0.8	**-**	>V(yst)
0.8	25779.79	2186.879
0.9	24188.44	1814.999
1.00	22915.37	1492.895
1.10	21873.76	1220.564
1.20	21005.75	998.009
1.30	20271.29	825.227
1.40	19641.74	702.221
1.50	19096.14	628.989
1.5971(opt)	***18631.62(opt)***	***605.511****
1.60	18618.74	605.532
1.70	18197.50	631.849
1.80	17823.06	707.941
1.90	17488.04	833.807
2.00	17186.53	1009.448
2.10	16913.72	1234.864
2.20	16665.72	1510.054
2.30	16439.29	1835.019
2.40	16231.72	2209.758
>2.40	-	>V(yst)

* MSE (min) at the value A(optimal).

6. **Conclusion**

When we examine Table 2, we observe that the second proposed estimator, $\bar{y}_9$, under optimum condition certainly performs quite better than all other estimators discussed here. Although the correlations are negative, we also examine the performance of the first proposed estimator, $\bar{y}_7$, according to the classical combined estimator. Therefore, for various values of A and θ in Table 3, the MSE values of $\bar{y}_{st}$ and $\bar{y}_7$ are computed. From Table 3, we observe that the first proposed estimator, $\bar{y}_7$, performs better than the estimator, $\bar{y}_{st}$, for a wide range of θ as $\theta \in [0.8, 2.40]$, even in the negative correlations.

References

[1] Koyuncu, N. and Kadilar, C. Family of Estimators of Population Mean Using Two Auxiliary Variables in Stratified Random Sampling Commun. in Statist.—Theor. and Meth, 38, 2009, 2398–2417.

[2] Perri, P.F. Improved ratio-cum-product type estimators. Statist. In Trans, 2007, 851-69.

[3] Plikusas, A. Some overview of the ratio type estimators In: Workshop on survey sampling theory and methodology, Statistics Estonia, 2008.

[4] Singh, M. P. Ratio-cum-product method of estimation. Metrika 12, 1967, 34 -42.

[5] Singh, R., Kumar, M., Chauhan, P., Sawan, N. and Smarandache, F. A general family of dual to ratio-cum-product estimator in sample surveys. Statist. In Trans- New series. IJSA, 2012, 1(1), 101-109.

[6] Srivenkataramana, T. and Tracy, D.S. An alternative to ratio method in sample surveys. Ann. Inst. Statist. Math.32 A, 1980, 111-120.

[7] Srivenkataramana, T. and Tracy, D.S. Extending product method of estimation to positive correlation case in surveys. Austral. J. Statist. 23, 1981, 95-100.

[8] Tracy, D.S., Singh, H.P. and Singh, R. An alternative to the ratio-cum-product estimator in sample surveys. Jour. of Statist. Plann. and Infere. 53, 1996, 375 -387.

Exponential Ratio-Product Type Estimators Under Second Order Approximation In Stratified Random Sampling

†1Rajesh Singh, [1]Prayas Sharma and [2]Florentin Smarandache

[1]Department of Statistics, Banaras Hindu University

Varanasi-221005, India

[2]Department of Mathematics, University of New Mexico, Gallup, USA

† Corresponding author, rsinghstat@yahoo.com

Abstract

Singh et al. (20009) introduced a family of exponential ratio and product type estimators in stratified random sampling. Under stratified random sampling without replacement scheme, the expressions of bias and mean square error (MSE) of Singh et al. (2009) and some other estimators, up to the first- and second-order approximations are derived. Also, the theoretical findings are supported by a numerical example.

Keywords: Stratified Random Sampling, population mean, study variable, auxiliary variable, exponential ratio type estimator, exponential product estimator, Bias and MSE.

1. INTRODUCTION

In survey sampling, it is well established that the use of auxiliary information results in substantial gain in efficiency over the estimators which do not use such information. However, in planning surveys, the stratified sampling has often proved needful in improving the precision of estimates over simple random sampling. Assume that the population U consist of L strata as $U=U_1, U_2,\ldots,U_L$. Here the size of the stratum U_h is N_h, and the size of simple random sample in stratum U_h is n_h, where h=1, 2,---,L.

When the population mean of the auxiliary variable, $\overline{X}$, is known, Singh et al. (2009) suggested a combined exponential ratio-type estimator for estimating the population mean of the study variable $(\overline{Y})$:

$$t_{1S} = \bar{y}\exp\left[\frac{\bar{X} - \bar{x}_{st}}{\bar{X} + \bar{x}_{st}}\right] \tag{1.1}$$

where,

$$\bar{y}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} y_{hi}\,, \qquad \bar{x}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} x_{hi}\,,$$

$$\bar{y}_{st} = \sum_{h=1}^{L} w_h \bar{y}_h\,, \quad \bar{x}_{st} = \sum_{h=1}^{L} w_h \bar{x}_h\,, \text{ and } \quad \bar{X} = \sum_{h=1}^{L} w_h \bar{X}_h\,.$$

The exponential product-type estimator under stratified random sampling is given by

$$t_{2S} = \bar{y}\exp\left[\frac{\bar{X} - \bar{x}_{st}}{\bar{x}_{st} + \bar{X}}\right] \tag{1.2}$$

Following Srivastava (1967) an estimator t_{3s} in stratified random sampling is defined as :

$$t_{3S} = \bar{y}\exp\left[\frac{\bar{X} - \bar{x}_{st}}{\bar{x}_{st} + \bar{X}}\right]^{\alpha} \tag{1.3}$$

where α is a constant suitably chosen by minimizing MSE of t_{3S}. For α=1 , t_{3S} is same as conventional exponential ratio-type estimator whereas for α = -1, it becomes conventional exponential product type estimator.

Singh et al. (2008) introduced an estimator which is linear combination of exponential ratio-type and exponential product-type estimator for estimating the population mean of the study variable $(\bar{Y})$ in simple random sampling. Adapting Singh et al. (2008) estimator in stratified random sampling we propose an estimator t_{4s} as :

$$t_{4S} = \bar{y}\left[\theta\exp\left[\frac{\bar{X} - \bar{x}_{st}}{\bar{x}_{st} + \bar{X}}\right] + (1-\theta)\exp\left[\frac{\bar{X} - \bar{x}_{st}}{\bar{x}_{st} + \bar{X}}\right]\right] \tag{1.4}$$

where θ is the constant and suitably chosen by minimizing mean square error of the estimator t_{4S}. It is observed that the estimators considered here are equally efficient when terms up to first order of approximation are taken. Hossain et al. (2006) and Singh and Smarandache (2013) studied some estimators in SRSWOR under second order approximation. Koyuncu and Kadilar (2009, 2010)), have studied some estimators in stratified random sampling under second order approximation. To have more clear picture about the best estimator, in this study we have derived the expressions of MSE's of the estimators considered in this paper up to second order of approximation in stratified random sampling.

3. Notations used

Let us define, $e_0 = \frac{\bar{y}_{st} - \bar{y}}{\bar{y}}$ and $e_1 = \frac{\bar{x}_{st} - \bar{x}}{\bar{x}}$,

such that

$$E(e_0) = E(e_1) = E(e_2) = 0,$$

$$V_{rs} = \sum_{h=1}^{L} W_h^{r+s} E\left[\left(\bar{x}_h - \bar{X}_h\right)^r \left(\bar{y}_h - \bar{Y}_h\right)^s\right]$$

To obtain the bias and MSE of the proposed estimators, we use the following notations in the rest of the article:

$$\bar{y}_{st} = \sum_{h=1}^{L} w_h \bar{y}_h = \bar{Y}(1 + e_0),$$

$$\bar{x}_{st} = \sum_{h=1}^{L} w_h \bar{x}_h = \bar{X}(1 + e_1),$$

where $\bar{y}_h$ and $\bar{Y}_h$ are the sample and population means of the study variable in the stratum h, respectively. Similar expressions for X and Z can also be defined.

Also, we have

$$E(e_0^2) = \frac{\sum_{h=1}^{L} w_h^2 \gamma_h S_{yh}^2}{\bar{Y}^2} = V_{200},$$

$$E(e_1^2) = \frac{\sum_{h=1}^{L} w_h^2 \gamma_h S_{xh}^2}{\bar{X}^2} = V_{020},$$

$$E(e_0 e_1) = \frac{\sum_{h=1}^{L} w_h^2 \gamma_h S_{xyh}}{\bar{X}\bar{Y}} = V_{110},$$

where

$$S_{yh}^2 = \frac{\sum_{i=1}^{N_h} (y_h - \bar{Y}_h)^2}{N_h - 1}, \qquad S_{xh}^2 = \frac{\sum_{i=1}^{N_h} (x_h - \bar{X}_h)^2}{N_h - 1}, \quad S_{zh}^2 = \frac{\sum_{i=1}^{N_h} (z_h - \bar{Z}_h)^2}{N_h - 1},$$

$$S_{xyh} = \frac{\sum_{i=1}^{N_h} (x_h - \bar{X}_h)(y_h - \bar{Y}_h)}{N_h - 1}, \qquad S_{yzh} = \frac{\sum_{i=1}^{N_h} (y_h - \bar{Y}_h)(z_h - \bar{Z}_h)}{N_h - 1},$$

$$\gamma_h = \frac{1 - f_h}{n_h}, \qquad f_h = \frac{n_h}{N_h}, \qquad w_h = \frac{N_h}{n_h}.$$

Some additional notations for second order approximation:

$$V_{rs} = \sum_{h=1}^{L} W_h^{r+s} \frac{1}{\bar{Y}^r \bar{X}^s} E\left[\left(\bar{y}_h - \bar{Y}_h\right)^s \left(\bar{x}_h - \bar{X}_h\right)^r\right]$$

where , $C_{rs(h)} = \frac{1}{N_h} \sum_{i=1}^{N_h} \left[\left(\bar{y}_h - \bar{Y}_h\right)^s \left(\bar{x}_h - \bar{X}_h\right)^r\right],$

$$V_{12} = \sum_{h=1}^{L} W_h^3 \frac{k_{1(h)} C_{12(h)}}{\bar{Y}\bar{X}^2}, \qquad V_{21} = \sum_{h=1}^{L} W_h^3 \frac{k_{1(h)} C_{21(h)}}{\bar{Y}^2 \bar{X}}, \qquad V_{30} = \sum_{h=1}^{L} W_h^3 \frac{k_{1(h)} C_{30(h)}}{\bar{Y}^3},$$

$$V_{03} = \sum_{h=1}^{L} W_h^3 \frac{k_{1(h)} C_{03(h)}}{\bar{X}^3}, \qquad V_{13} = \sum_{h=1}^{L} W_h^4 \frac{k_{2(h)} C_{13(h)} + 3k_{3(h)} C_{01(h)} C_{02(h)}}{\bar{Y}\bar{X}^3},$$

$$V_{04} = \sum_{h=1}^{L} W_h^4 \frac{k_{2(h)} C_{04(h)} + 3k_{3(h)} C_{02(h)}^2}{\overline{X}^4},$$

$$V_{22} = \sum_{h=1}^{L} W_h^4 \frac{k_{2(h)} C_{22(h)} + k_{3(h)} \left(C_{01(h)} C_{02(h)} + 2C_{11(h)}^2 \right)}{\overline{Y}^2 \overline{X}^2},$$

where $k_{1(h)} = \dfrac{(N_h - n_h)(N_h - 2n_h)}{n^2 (N_h - 1)(N_h - 2)}$,

$$k_{2(h)} = \frac{(N_h - n_h)(N_h + 1)N_h - 6n_h(N_h - n_h)}{n^3 (N_h - 1)(N_h - 2)(N_h - 3)},$$

$$k_{3(h)} = \frac{(N_h - n_h)N_h(N_h - n_h - 1)(n_h - 1)}{n^3 (N_h - 1)(N_h - 2)(N_h - 3)}.$$

4. **First Order Biases and Mean Squared Errors under stratified random sampling**

The expressions for biases and MSE,s of the estimators t_{1S}, t_{2S} and t_{3S} respectively, are :

$$Bias(t_{1S}) = \overline{Y}\left[\frac{3}{8}V_{02} - \frac{1}{2}V_{11}\right] \quad (4.1)$$

$$MSE(t_{1S}) = \overline{Y}^2\left[V_{20} + \frac{1}{2}V_{02} - V_{11}\right] \quad (4.2)$$

$$Bias(t_{2S}) = \overline{Y}\left[\frac{1}{2}V_{11} - \frac{1}{8}V_{02}\right] \quad (4.3)$$

$$MSE(t_{2S}) = \overline{Y}^2\left[V_{20} + \frac{1}{4}V_{02} + V_{11}\right] \quad (4.4)$$

$$Bias(t_{3S}) = \overline{Y}\left[\alpha\frac{1}{4}V_{02} + \alpha^2\frac{1}{8}V_{02} - \frac{1}{2}\alpha V_{11}\right] \quad (4.5)$$

$$MSE(t_{3S}) = \overline{Y}^2\left[V_{20} + \frac{1}{4}\alpha^2 V_{02} - \alpha V_{11}\right] \tag{4.6}$$

By minimizing $MSE(t_{3S})$, the optimum value of α is obtained as $\alpha_o = \frac{2V_{11}}{V_{02}}$. By putting this optimum value of α in equation (4.5) and (4.6), we get the minimum value for bias and MSE of the estimator t_{3S}.

The expression for the bias and MSE of t_{4s} to the first order of approximation are given respectively, as

$$Bias(t_{4s)=}\overline{Y}\left[\theta\left\{\left\{\frac{3}{8}V_{02} - \frac{1}{2}V_{11}\right\} + (1-\theta)\left\{\frac{1}{2}V_{11} - \frac{1}{8}V_{02}\right\}\right\}\right] \tag{4.7}$$

$$MSE(t_{4S}) = \overline{Y}^2\left[V_{20} + \left(\frac{1}{2} - \theta\right)^2 V_{02} + 2\left(\frac{1}{2} - \theta\right)V_{11}\right] \tag{4.8}$$

By minimizing $MSE(t_{4S})$, the optimum value of θ is obtained as $\theta_o = \frac{V_{11}}{V_{02}} + \frac{1}{2}$. By putting this optimum value of α in equation (4.7) and (4.8) we get the minimum value for bias and MSE of the estimator t_{3S}. We observe that for the optimum cases the biases of the estimators t_{3S} and t_{4S} are different but the MSE of t_{3S} and t_{4S} are same. It is also observed that the MSE's of the estimators t_{3S} and t_{4S} are always less than the MSE's of the estimators t_{1S} and t_{2S}. This prompted us to study the estimators t_{3S} and t_{4S} under second order approximation.

5. Second Order Biases and Mean Squared Errors in stratified random sampling

Expressing estimator t_i's(i=1,2,3,4) in terms of e_i's (i=0,1), we get

$$t_{1s} = \overline{Y}(1+e_0)\exp\left[\frac{-e_1}{2+e_1}\right]$$

Or

$$t_{1s}-\overline{Y}=\overline{Y}\left\{e_0-\frac{e_1}{2}-\frac{1}{2}e_0e_1+\frac{3}{8}e_1{}^2+\frac{3}{8}e_0e_1{}^2-\frac{7}{48}e^3-\frac{7}{48}e_0e_1{}^3+\frac{25}{384}e^4\right\} \quad (5.1)$$

Taking expectations, we get the bias of the estimator t_{1S} up to the second order of approximation as

$$\text{Bias}_2(t_{1s})==\frac{\overline{Y}}{2}\left[\left[-V_{11}+\frac{3}{4}V_{02}+\frac{3}{4}V_{12}-\frac{7}{24}V_{03}-\frac{7}{24}V_{13}+\frac{25}{192}V_{04}\right] \quad (5.2)$$

Squaring equation (5.1) and taking expectations and using lemmas we get MSE of t_{1S} up to second order of approximation as

$$MSE(t_{1S})=E\left[\overline{Y}\left(e_0-\frac{e_1}{2}+\frac{3}{8}e_1{}^2-\frac{1}{2}e_0e_1+\frac{3}{8}e_0e_1{}^2-\frac{7}{48}e^3\right)\right]^2$$

Or,

$$MSE(t_{1s})=\overline{Y}^2E\left[\left\{e_0{}^2+\frac{1}{4}e_1{}^2-e_0e_1+e_0{}^2e_1{}^2-e_0{}^2e_1-\frac{3}{8}e_1{}^3-\frac{25}{24}e_0e_1{}^3+\frac{5}{4}e_0e_1{}^2+\frac{55}{192}e_1{}^4\right\}\right]$$

(5.3)

Or,

$$MSE_2(t_{1s})=\overline{Y}^2\left[V_{20}+\frac{1}{4}V_{02}-V_{11}+V_{22}-V_{21}+\frac{5}{4}V_{12}-\frac{25}{24}V_{13}+\frac{55}{192}V_{04}\right] \quad (5.4)$$

Similarly we get the biases and MSE's of the estimators t_{2S}, t_{3S} and t_{4S} up to second order of approximation respectively, as

$$\text{Bias}_2(t_{2s})=\frac{\overline{Y}}{2}\left[V_{11}-\frac{1}{4}V_{02}-\frac{1}{4}V_{12}-\frac{5}{24}V_{13}+\frac{1}{192}V_{04}-\frac{5}{24}V_{03}\right] \quad (5.5)$$

$$MSE_2(t_{2S})=\overline{Y}^2\left[V_{20}+\frac{1}{4}V_{02}+V_{11}+\frac{23}{192}V_{04}-\frac{1}{8}V_{03}+\frac{1}{4}V_{12}-\frac{1}{24}V_{13}+V_{21}\right) \quad (5.6)$$

$$\text{Bias}_2(t_{3S}) = \overline{Y}\left[\left(\frac{\alpha^2}{8}+\frac{\alpha}{4}\right)V_{02}+\left(\frac{\alpha^2}{8}+\frac{\alpha}{4}\right)V_{12}-\frac{\alpha}{2}V_{11}\left(\frac{\alpha^2}{8}+\frac{\alpha^3}{48}\right)V_{03}-\left(\frac{\alpha^2}{8}+\frac{\alpha^3}{48}\right)V_{13}\right.$$

$$\left.+\left(\frac{\alpha^2}{32}+\frac{\alpha^3}{32}+\frac{\alpha^4}{384}\right)V_{04}\right] \tag{5.7}$$

$$\text{MSE}_2(t_{3S}) = \overline{Y}^2\left[V_{20}+\frac{\alpha^2}{4}V_{02}-\alpha V_{11}+\left(\frac{\alpha}{2}+\frac{\alpha^2}{2}\right)V_{22}-\alpha V_{21}+\left(\frac{\alpha}{2}+\frac{\alpha^2}{2}\right)V_{22}+\left(\frac{\alpha}{2}+\frac{3\alpha^2}{4}\right)V_{12}\right.$$

$$\left.-\left(\frac{\alpha^2}{4}+\frac{\alpha^2}{8}\right)V_{03}-\left(\frac{3\alpha^2}{4}+\frac{7\alpha^3}{24}\right)V_{13}+\left(\frac{\alpha^2}{16}+\frac{\alpha^3}{16}+\frac{7\alpha^4}{192}\right)V_{04}\right] \tag{5.8}$$

$$\text{Bias}_2(t_{4S}) = E(t_{4S}-\overline{Y}) = \overline{Y}\left[\left(\frac{1}{2}-\alpha\right)V_{11}-\frac{1}{2}\left(\frac{1}{4}-\alpha\right)\{V_{02}+V_{12}\}+\left\{\frac{1}{16}\left(\frac{1}{24}+\alpha\right)\right\}V_{04}\right.$$

$$\left.-\frac{1}{48}(2\alpha+5)\{V_{03}+V_{13}\}\right] \tag{5.9}$$

$$\text{MSE}_2(t_{4S}) = \overline{Y}^2\left[V_{20}+\left(\frac{1}{2}-\theta\right)^2V_{02}+\left\{\left(\frac{1}{2}-\theta\right)^2+\frac{(4\theta-1)}{4}\right\}V_{22}+\left\{\left(\frac{1}{2}-\theta\right)^2+\frac{(4\theta-1)}{4}\right\}V_{12}\right.$$

$$+\left\{\frac{1}{64}(4\theta-1)^2-\frac{1}{24}\left(\frac{1}{2}-\theta\right)(2\theta+5)\right\}V_{04}+2\left(\frac{1}{2}-\theta\right)V_{21}+\frac{1}{4}\left(\frac{1}{2}-\theta\right)(4\theta-1)V_{03}$$

$$\left.-\left\{-\frac{1}{24}(2\theta+5)+\frac{1}{2}\left(\frac{1}{2}-\theta\right)(4\theta-1)\right\}V_{13}\right] \tag{5.10}$$

The optimum value of α we get by minimizing $\text{MSE}_2(t_{3S})$. But theoretically the determination of the optimum value of α is very difficult, we have calculated the optimum value by using numerical techniques. Similarly the optimum value of θ which minimizes the MSE of the estimator t_{4s} is obtained by using numerical techniques.

6. Numerical Illustration

For the one natural population data, we shall calculate the bias and the mean square error of the estimator and compare Bias and MSE for the first and second order of approximation.

Data Set-1

To illustrate the performance of above estimators , we have considered the natural data given in Singh and Chaudhary (1986, p.162).

The data were collected in a pilot survey for estimating the extent of cultivation and production of fresh fruits in three districts of Uttar- Pradesh in the year 1976-1977.

Table 6.1: **Bias and MSE of estimators**

Estimator	Bias		MSE	
	First order	Second order	First order	Second order
t_{1S}	-1.532898612	-1.475625158	2305.736643	2308.748272
t_{2S}	8.496498176	8.407682289	23556.67462	23676.94086
t_{3S}	-1.532898612	-1.763431841	704.04528	705.377712
t_{4S}	-5.14408	-5.0089	704.04528	707.798567

7. CONCLUSION

In the Table 6.1 the bias and MSE of the estimators t_{1S}, t_{2S}, t_{3S} and t_{4S} are written under first order and second order of approximation. The estimator t_{2S} is exponential product-type estimator and it is considered in case of negative correlation. So the bias and mean squared error for this estimator is more than the other estimators considered here. For the classical exponential ratio-type estimator, it is observed that the biases and the mean squared errors increased for second order. The estimator t_{3S} and t_{4S} have the same mean squared error for the first order but the mean squared error of t_{3S} is less than t_{4S} for the second order. So, on the basis of the given data set we conclude that the estimator t_{3S} is best followed by the estimator t_{4S} among the estimators considered here.

REFERENCES

Koyuncu, N. and Kadilar, C. (2009) : Family of estimators of population mean using two auxiliary variables in stratified random sampling. Commun. in Statist.—Theor. and Meth, 38, 2398–2417.

Koyuncu, N. and Kadilar, C. (2010) : On the family of estimators of population mean in stratified random sampling. Pak. Jour. Stat., 26(2),427-443.

Singh, D. and Chudhary, F.S. (1986): Theory and analysis of sample survey designs. Wiley Eastern Limited, New Delhi.

Singh, R., Chauhan, P. and Sawan, N.(2008): On linear combination of Ratio-product type exponential estimator for estimating finite population mean. Statistics in Transition,9(1),105-115.

Singh, R., Kumar, M., Chaudhary, M. K., Kadilar, C. (2009) : Improved Exponential Estimator in Stratified Random Sampling. Pak. J. Stat. Oper. Res. 5(2), pp 67-82.

Singh, R. and Smarandache, F. (2013): On improvement in estimating population parameter(s) using auxiliary information. Educational Publishing & Journal of Matter Regularity (Beijing) pg 25-41.

TWO-PHASE SAMPLING IN ESTIMATION OF POPULATION MEAN IN THE PRESENCE OF NON-RESPONSE

[1]Manoj Kr. Chaudhary, [1]Anil Prajapati ,[†1]Rajesh Singh and [2]Florentin Smarandache

[1]Department of Statistics, Banaras Hindu University, Varanasi-221005

[2]Department of Mathematics, University of New Mexico, Gallup, USA

† Corresponding author, rsinghstat@yahoo.com

Abstract

The present paper presents the detail discussion on estimation of population mean in simple random sampling in the presence of non-response. Motivated by Gupta and Shabbir (2008), we have suggested the class of estimators of population mean using an auxiliary variable under non-response. A theoretical study is carried out using two-phase sampling scheme when the population mean of auxiliary variable is not known. An empirical study has also been done in the support of theoretical results.

Keywords: Two-phase sampling, class of estimators, optimum estimator, non-response, numerical illustrations.

1. Introduction

The auxiliary information is generally used to improve the efficiency of the estimators. Cochran (1940) proposed the ratio estimator for estimating the population mean whenever study variable is positively correlated with auxiliary variable. Contrary to the situation of ratio estimator, if the study and auxiliary variables are negatively correlated, Murthy (1964) suggested the product estimator to estimate the population mean. Hansen et al. (1953) proposed the difference estimator which was subsequently modified to provide the linear regression estimator for the population mean or total. Mohanty (1967) suggested an estimator by combining the ratio and regression methods for estimating the population parameters. In order to estimate the population mean or population total of the study character utilizing auxiliary information, several other authors including Srivastava (1971), Reddy (1974), Ray and Sahai (1980), Srivenkataramana (1980), Srivastava and Jhajj (1981) and Singh and Kumar (2008, 2011) have proposed estimators which lead improvements over usual per unit estimator.

It is observed that the non-response is a common problem in any type of survey. Hansen and Hurwitz (1946) were the first to contract the problem of non-response while conducting mail surveys. They suggested a technique, known as 'sub-sampling of non-respondents', to deal with the problem of non-response and its adjustments. In fact they developed an unbiased estimator for population mean in the presence of non-response by dividing the population into two groups, viz. response group and non-response group. To avoid bias due to non-response, they suggested for taking a sub-sample of the non-responding units.

Let us consider a population consists of N units and a sample of size n is selected from the population using simple random sampling without replacement (SRSWOR) scheme. Let us assume that Y and X be the study and auxiliary variables with respective population means $\overline{Y}$ and $\overline{X}$. Let us consider the situation in which study variable is subjected to non-response and auxiliary variable is free from the non-response. It is observed that there are n_1 respondent and n_2 non-respondent units in the sample of n units for the study variable. Using the technique of sub sampling of non-respondents suggested by Hansen and Hurwitz (1946), we select a sub-sample of h_2 non-respondent units from n_2 units such that $h_2 = n_2/k, k \geq 1$ and collect the information on sub-sample by personal interview method. The usual sample mean, ratio and regression estimators for estimating the population mean $\overline{Y}$ under non-response are respectively represented by

$$\overline{y}^* = \frac{n_1 \overline{y}_{n1} + n_2 \overline{y}_{h2}}{n} \tag{1.1}$$

$$\bar{y}_R^* = \frac{\bar{y}^*}{\bar{x}} \bar{X} \tag{1.2}$$

$$\bar{y}_{lr}^* = \bar{y}^* + b(\bar{X} - \bar{x}) \tag{1.3}$$

where $\bar{y}_{n1}$ and $\bar{y}_{h2}$ are the means based on n_1 respondent and h_2 non-respondent units respectively. $\bar{x}$ is the sample mean estimator of population mean $\bar{X}$, based on sample of size n and b is the sample regression coefficient of Y on X.

The variance and mean square errors (MSE) of the above estimators $\bar{y}^*$, $\bar{y}_R^*$ and $\bar{y}_{lr}^*$ are respectively given by

$$V(\bar{y}^*) = \left(\frac{1}{n} - \frac{1}{N}\right) S_Y^2 + \frac{(k-1)}{n} W_2 S_{Y2}^2 \tag{1.4}$$

$$MSE(\bar{y}_R^*) = \left(\frac{1}{n} - \frac{1}{N}\right) \bar{Y}^2 \left(C_Y^2 + C_X^2 - 2\rho C_X C_Y\right) + \frac{(k-1)}{n} W_2 S_{Y2}^2 \tag{1.5}$$

$$MSE(\bar{y}_{lr}^*) = \left(\frac{1}{n} - \frac{1}{N}\right) \bar{Y}^2 C_Y^2 \left(1 - \rho^2\right) + \frac{(k-1)}{n} W_2 S_{Y2}^2 \tag{1.6}$$

where S_Y^2 and S_X^2 are respectively the mean squares of Y and X in the population. $C_Y \left(= S_Y / \bar{Y}\right)$ and $C_X \left(= S_X / \bar{X}\right)$ are the coefficients of variation of Y and X respectively. S_{Y2}^2 and W_2 are respectively the mean square and non-response rate of the non-response group in the population for the study variable Y. ρ is the population correlation coefficient between Y and X.

When the information on population mean of auxiliary variable is not available, one can use the two-phase sampling scheme in obtaining the improved estimator rather than the previous ones. Neyman (1938) was the first who gave concept of two-phase sampling in estimating the population parameters. Two-phase sampling is cost effective as well as easier. This sampling scheme is used to obtain the information about auxiliary variable cheaply from a bigger sample at first phase and relatively small sample at the second stage. Sukhatme (1962) used two-phase sampling scheme to propose a general ratio-type estimator. Rao (1973) used two-phase sampling to stratification, non-response problems and investigative comparisons. Cochran (1977) supplied some basic information for two-phase sampling. Sahoo et al. (1993) provided regression approach in estimation by using two auxiliary variables for two-phase sampling. In the sequence of improving the efficiency of the estimators, Singh and Upadhyaya (1995) suggested a generalized estimator to estimate population mean using two auxiliary variables in two-phase sampling.

In estimating the population mean $\overline{Y}$, if $\overline{X}$ is unknown, first, we obtain the estimate of it using two-phase sampling scheme and then estimate $\overline{Y}$. Under two-phase sampling scheme, first we select a larger sample of n' units from the population of size N with the help of SRSWOR scheme. Secondly, we select a small sample of size n from n' units. Let us again assume that the situation in which the non-response is observed on study variable only and auxiliary variable is free from the non-response. The usual ratio and regression estimators of population mean $\overline{Y}$ under two-phase sampling in the presence of non-response are respectively given by

$$\overline{y}_R^{**} = \frac{\overline{y}^*}{\overline{x}}\overline{x}' \tag{1.7}$$

and
$$\overline{y}_{lr}^{**} = \overline{y}^* + b\left(\overline{x}' - \overline{x}\right) \tag{1.8}$$

where $\overline{x}'$ is the mean based on n' units for the auxiliary variable.

The MSE's of the estimators $\overline{y}_R^{**}$ and $\overline{y}_{lr}^{**}$ are respectively represented by the following expressions

$$MSE\left(\overline{y}_R^{**}\right) = \overline{Y}^2\left[\left(\frac{1}{n'} - \frac{1}{N}\right)C_Y^2 + \left(\frac{1}{n} - \frac{1}{n'}\right)\left(C_Y^2 + C_X^2 - 2\rho C_X C_Y\right)\right] + \frac{(k-1)}{n}W_2 S_{Y2}^2 \tag{1.9}$$

and

$$\mathrm{MSE}\left(\bar{y}_{lr}^{**}\right)=\bar{Y}^2\left[\left(\frac{1}{n'}-\frac{1}{N}\right)C_Y^2+\left(\frac{1}{n}-\frac{1}{n'}\right)C_Y^2\left(1-\rho^2\right)\right]+\frac{(k-1)}{n}W_2S_{Y2}^2 \tag{1.10}$$

In the present paper, we have discussed the study of non-response of a general class of estimators using an auxiliary variable. We have suggested the class of estimators in two-phase sampling when the population mean of auxiliary variable is unknown. The optimum property of the class is also discussed and it is compared to ratio and regression estimators under non-response. The theoretical study is also supported with the numerical illustrations.

2. Suggested Class of Estimators

Let us assume that the non-response is observed on the study variable and auxiliary variable provides complete response on the units. Motivated by Gupta and Shabbir (2008), we suggest a class of estimators of population mean $\bar{Y}$ under non-response as

$$\bar{y}_t^* = \left[\alpha_1\bar{y}^* + \alpha_2\left(\bar{X}-\bar{x}\right)\right]\left(\frac{\eta\bar{X}+\lambda}{\eta\bar{x}+\lambda}\right) \tag{2.1}$$

where α_1 and α_2 are the constants and whose values are to be determined. λ and $\eta(\neq 0)$are either constants or functions of the known parameters.

In order to obtain the bias and MSE of $\bar{y}_t^*$, we use the large sample approximation. Let us assume that

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

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

$$E\left(e_1^2\right)=\frac{V\left(\bar{y}^*\right)}{\bar{Y}^2}=\left(\frac{1}{n}-\frac{1}{N}\right)C_Y^2+\frac{(k-1)}{n}W_2\frac{S_{Y2}^2}{\bar{Y}^2},$$

$$E\left(e_2^2\right)=\frac{V\left(\bar{x}\right)}{\bar{X}^2}=\left(\frac{1}{n}-\frac{1}{N}\right)C_X^2$$

and $E(e_1 e_2) = \frac{Cov\left(\bar{y}^*, \bar{x}\right)}{\bar{Y}\bar{X}} = \left(\frac{1}{n} - \frac{1}{N}\right)\rho C_X C_Y$.

Putting the values of $\bar{y}^*$ and $\bar{x}$ form the above assumptions in the equation (2.1), we get

$$\bar{y}_t^* - \bar{Y} \cong \bar{Y}(\alpha_1 - 1) + \alpha_1 \bar{Y}\left(e_1 - \tau e_2 - \tau e_1 e_2 + \tau^2 e_2^2\right) - \alpha_2 \bar{X}\left(e_2 - \tau e_2^2\right) \qquad (2.2)$$

On taking expectation of the equation (2.2), the bias of $\bar{y}_t^*$ to the first order of approximation is given by

$$B\left(\bar{y}_t^*\right) = E\left(\bar{y}_t^* - \bar{Y}\right) = \bar{Y}(\alpha_1 - 1) + \left(\frac{1}{n} - \frac{1}{N}\right)\left[\alpha_1 \bar{Y}\left(\tau^2 C_X^2 - \tau\rho C_X C_Y\right) + \alpha_2 \bar{X}\tau C_X^2\right] \qquad (2.3)$$

Squaring both the sides of the equation (2.2) and taking expectation, we can obtain the MSE of $\bar{y}_t^*$ to the first order of approximation as

$$MSE\left(\bar{y}_t^*\right) = \bar{Y}^2(\alpha_1 - 1)^2 + \left(\frac{1}{n} - \frac{1}{N}\right)\Big[\alpha_1^2 \bar{Y}^2\left(C_Y^2 + \tau^2 C_X^2 - 2\tau\rho C_X C_Y\right) + \alpha_2^2 \bar{X}^2 C_X^2$$

$$- 2\alpha_1\alpha_2 \bar{Y}\bar{X} C_X\left(\rho C_Y - \tau C_X\right)\Big] + \alpha_1^2 \frac{(k-1)}{n} W_2 S_{Y2}^2 \qquad (2.4)$$

In the sequence of obtaining the best estimator within the suggested class with respect to α_1 and α_2, we obtain the optimum values of α_1 and α_2. On differentiating $MSE\left(\bar{y}_t^*\right)$ with respect to α_1 and α_2 and equating the derivatives to zero, we have

$$\frac{\partial MSE\left(\bar{y}_t^*\right)}{\partial \alpha_1} = \bar{Y}^2(\alpha_1 - 1) + \left(\frac{1}{n} - \frac{1}{N}\right)\left[\bar{Y}^2\alpha_1\left(C_Y^2 + \tau^2 C_X^2 - 2\tau\rho C_X C_Y\right) - \alpha_2 \bar{X}\bar{Y} C_X\left(\rho C_Y - \tau C_X\right)\right]$$

$$+ \alpha_1 \frac{(k-1)}{n} W_2 S_{Y2}^2 = 0 \qquad (2.5)$$

$$\frac{\partial MSE\left(\bar{y}_t^*\right)}{\partial \alpha_2} = \left(\frac{1}{n} - \frac{1}{N}\right)\left[\alpha_2 \bar{X}^2 C_X^2 - \alpha_1 \bar{X}\bar{Y} C_X\left(\rho C_Y - \tau C_X\right)\right] = 0 \qquad (2.6)$$

Solving the equations (2.4) and (2.5), we get

$$\alpha_1(\text{opt}) = \frac{1}{1+\left(\frac{1}{n}-\frac{1}{N}\right)C_Y^2\left(1-\rho^2\right)+\frac{(k-1)}{n}W_2\frac{S_{Y2}^2}{\overline{Y}^2}} \tag{2.7}$$

and $$\alpha_2(\text{opt}) = \frac{\alpha_1(\text{opt})\overline{Y}(\rho C_Y - \tau C_X)}{\overline{X}C_X} \tag{2.8}$$

Substituting the values of $\alpha_1(\text{opt})$ and $\alpha_2(\text{opt})$ from equations (2.7) and (2.8) into the equation (2.4), the MSE of $\overline{y}_t^*$ is given by the following expression.

$$\text{MSE}\left(\overline{y}_t^*\right)_{\min} = \frac{\text{MSE}\left(\overline{y}_{lr}^*\right)}{1+\left(\frac{1}{n}-\frac{1}{N}\right)C_Y^2\left(1-\rho^2\right)+\frac{(k-1)}{n}W_2\frac{S_{Y2}^2}{\overline{Y}^2}} \tag{2.9}$$

3. Suggested Class in Two-Phase Sampling

It is generally seen that the population mean of auxiliary variable, $\overline{X}$ is not known. In this situation, we may use the two-phase sampling scheme to find out the estimate of $\overline{X}$. Using two-phase sampling, we now suggest a class of estimators of population mean $\overline{Y}$ in the presence of non-response when $\overline{X}$ is unknown, as

$$\overline{y}_t^{**} = \left[\alpha_1\overline{y}^* + \alpha_2\left(\overline{x}' - \overline{x}\right)\right]\left(\frac{\eta\overline{x}' + \lambda}{\eta\overline{x} + \lambda}\right) \tag{3.1}$$

3.1 Bias and MSE of $\overline{y}_t^{**}$

By applying the large sample approximation, we can obtain the bias and mean square error of $\overline{y}_t^{**}$. Let us assume that

$$\overline{y}^* = \overline{Y}(1+e_1),\ \overline{x} = \overline{X}(1+e_2) \text{ and } \overline{x}' = \overline{X}(1+e_3)$$

such that $E(e_1) = E(e_2) = E(e_3) = 0$,

$$E(e_1^2)=\left(\frac{1}{n}-\frac{1}{N}\right)C_Y^2+\frac{(k-1)}{n}W_2\frac{S_{Y2}^2}{\overline{Y}^2} \quad , \ E(e_2^2)=\left(\frac{1}{n}-\frac{1}{N}\right)C_X^2,$$

$$E(e_3^2)=\left(\frac{1}{n'}-\frac{1}{N}\right)C_X^2 \ , \ \mathrm{E}(e_1e_2)=\left(\frac{1}{n}-\frac{1}{N}\right)\rho C_X C_Y,$$

$$\mathrm{E}(e_1e_3)=\left(\frac{1}{n'}-\frac{1}{N}\right)\rho C_X C_Y \text{ and } \mathrm{E}(e_2e_3)=\left(\frac{1}{n'}-\frac{1}{N}\right)C_X^2.$$

Under the above assumption, the equation (3.1) gives

$$\overline{y}_t^{**}-\overline{Y}=\overline{Y}(\alpha_1-1)+\alpha_1\overline{Y}\left(e_1+\tau^2e_2^2-\tau e_2-\tau e_1e_2+\tau e_3+\tau e_1e_3-\tau^2e_2e_3\right)$$
$$+\alpha_2\overline{X}\left(e_3-e_2\tau e_2e_3+\tau e_2^2+\tau e_3^2-\tau e_2e_3\right) \tag{3.2}$$

Taking expectation of both the sides of equation (3.2), we get the bias of $\overline{y}_t^{**}$ up to the first order of approximation as

$$B\left(\overline{y}_t^{**}\right)=\overline{Y}(\alpha_1-1)+\left(\frac{1}{n}-\frac{1}{n'}\right)\tau\left[\alpha_1\overline{Y}\left(C_X^2-\rho C_X C_Y\right)+\alpha_2\overline{X}C_X^2\right] \tag{3.3}$$

The MSE of $\overline{y}_t^{**}$ up to the first order of approximation can be obtained by the following expression

$$\mathrm{MSE}\left(\overline{y}_t^{**}\right)=\mathrm{E}\left(\overline{y}_t^{**}-\overline{Y}\right)^2=\overline{Y}^2(\alpha_1-1)^2$$

$$+\alpha_1^2\overline{Y}^2\left[\left(\frac{1}{n}-\frac{1}{N}\right)C_Y^2+\left(\frac{1}{n}-\frac{1}{n'}\right)\left(\tau^2C_X^2-2\tau\rho C_X C_Y\right)+\frac{(k-1)}{n}W_2\frac{S_{Y2}^2}{\overline{Y}^2}\right]$$
$$+\left(\frac{1}{n}-\frac{1}{n'}\right)\left[\alpha_2^2\overline{X}^2C_X^2+2\alpha_1\alpha_2\overline{X}\overline{Y}\left(\tau C_X^2-\rho C_X C_Y\right)\right] \tag{3.4}$$

3.2 Optimum Values of α_1 and α_2

On differentiating $\mathrm{MSE}\left(\overline{y}_t^{**}\right)$ with respect to α_1 and α_2 and equating the derivatives to zero, we get the normal equations

$$\frac{\partial MSE\left(\bar{y}_t^{**}\right)}{\partial \alpha_1} = \bar{Y}^2(\alpha_1 - 1) + \alpha_1 \bar{Y}^2\left[\left(\frac{1}{n} - \frac{1}{N}\right)C_Y^2 + \left(\frac{1}{n} - \frac{1}{n'}\right)\left(\tau^2 C_X^2 - 2\tau\rho C_X C_Y\right) + \frac{(k-1)}{n} W_2 \frac{S_{Y2}^2}{\bar{Y}^2}\right]$$

$$+\left(\frac{1}{n} - \frac{1}{n'}\right)\alpha_2 \bar{X}\bar{Y}\left(\tau C_X^2 - \rho C_X C_Y\right) = 0 \qquad (3.5)$$

and $$\frac{\partial MSE\left(\bar{y}_t^{**}\right)}{\partial \alpha_2} = \left(\frac{1}{n} - \frac{1}{n'}\right)\left[\alpha_2 \bar{X}^2 C_X^2 + \alpha_1 \bar{X}\bar{Y}\left(\tau C_X^2 - \rho C_X C_Y\right)\right] = 0 \qquad (3.6)$$

From equations (3.5) and (3.6), we get the optimum values of α_1 and α_2 as

$$\alpha_1(opt) = \frac{1}{1 + \left(\frac{1}{n} - \frac{1}{N}\right)C_Y^2 - \left(\frac{1}{n} - \frac{1}{n'}\right)\rho^2 C_Y^2 + \frac{(k-1)}{n} W_2 \frac{S_{Y2}^2}{\bar{Y}^2}} \qquad (3.7)$$

and $$\alpha_2(opt) = \frac{\alpha_1(opt)\bar{Y}\left(\rho C_Y - \tau C_X\right)}{\bar{X} C_X} \qquad (3.8)$$

On substituting the optimum values of α_1 and α_2, the equation (3.4) provides minimum MSE of $\bar{y}_t^{**}$

$$MSE\left(\bar{y}_t^{**}\right)_{min} = \frac{MSE\left(\bar{y}_{lr}^{**}\right)}{1 + \left(\frac{1}{n} - \frac{1}{N}\right)C_Y^2 - \left(\frac{1}{n} - \frac{1}{n'}\right)\rho^2 C_Y^2 + \frac{(k-1)}{n} W_2 \frac{S_{Y2}^2}{\bar{Y}^2}} \qquad (3.9)$$

4. Empirical Study

In the support of theoretical results, some numerical illustrations are given below:

4.1 In this section, we have illustrated the relative efficiency of the estimators $\bar{y}_R^*$, $\bar{y}_{lr}^*$ and $\bar{y}_t^*(opt)$ with respect to $\bar{y}^*$. For this purpose, we have considered the data used by Kadilar and Cingi (2006). The details of the population are given below:

$N = 200$, $n = 50$, $\bar{Y} = 500$, $\bar{X} = 25$, $C_Y = 15$, $C_X = 2$, $\rho = 0.90$

$k = 1.5$, $S_{Y2}^2 = \frac{4}{5} S_Y^2$

Table 1. Percentage Relative Efficiency (PRE) with respect to $\bar{y}^*$

W_2	Estimator		
	$\bar{y}_R^*$	$\bar{y}_{lr}^*$	$\bar{y}_t^*(opt)$
0.1	126.74	432.88	788.38
0.2	125.13	373.03	746.53
0.3	123.70	331.43	722.93
0.4	122.42	300.83	710.33
0.5	121.28	277.37	704.87

4.2 The present section presents the relative efficiency of the estimators $\bar{y}_R^{**}$, $\bar{y}_{lr}^{**}$ and $\bar{y}_t^{**}(opt)$ with respect to $\bar{y}^*$. There are two data sets which have been considered to illustrate the theoretical results.

Data Set 1:

The population considered by Srivastava (1993) is used to give the numerical interpretation of the present study. The population of seventy villages in a Tehsil of India along with their cultivated area (in acres) in 1981 is considered. The cultivated area (in acres) is taken as study variable and the population is assumed to be auxiliary variable. The population parameters are given below:

$N = 70$, $n' = 40$, $n = 25$, $\bar{Y} = 981.29$, $\bar{X} = 1755.53$, $S_Y = 613.66$,

$S_X = 1406.13$, $S_{Y2} = 244.11$, $\rho = 0.778$, $k = 1.5$

Table 2: Percentage Relative Efficiency with respect to $\bar{y}^*$

W_2	Estimator		
	$\bar{y}^*_R$	$\bar{y}^*_{lr}$	$\bar{y}^*_t(opt)$
0.1	125.48657	153.56020	154.57983
0.2	125.10358	152.57858	153.60848
0.3	124.73193	151.63228	152.67552
0.4	124.37111	150.71945	151.77449
0.5	124.02068	149.83834	150.90579

Data Set 2:

Now, we have used another population considered by Khare and Sinha (2004). The data are based on the physical growth of upper-socio-economic group of 95 school children of Varanasi district under an ICMR study, Department of Paediatrics, Banaras Hindu University, India during 1983-84. The details are given below:

$N = 95$, $n' = 70$, $n = 35$, $\bar{Y} = 19.4968$, $\bar{X} = 55.8611$, $S_Y = 3.0435$, $S_X = 3.2735$,

$S_{Y2} = 2.3552$, $\rho = 0.8460$, $k = 1.5$.

Table 3: Percentage Relative Efficiency with respect to $\bar{y}^*$

W_2	Estimator		
	$\bar{y}^*_R$	$\bar{y}^*_{lr}$	$\bar{y}^*_t(opt)$
0.1	159.61889	217.83004	217.99278
0.2	155.61224	207.27149	207.43596
0.3	152.10325	198.44091	198.58540
0.4	149.01829	190.94488	190.94488
0.5	146.26158	184.51722	184.66554

5. Conclusion

The study of a general class of estimators of population mean under non-response has been presented. We have also suggested a class of estimators of population mean in the presence of non-response using two-phase sampling when population mean of auxiliary variable is not known. The optimum property of the suggested class has been discussed. We have compared the optimum estimator with some existing estimators through numerical study. The Tables 1, 2 and 3 represent the percentage relative efficiency of the optimum estimator of suggested class, linear regression estimator and ratio estimator with respect to sample mean estimator. In the above tables, we have observed that the percentage relative efficiency of the optimum estimator is higher than the linear regression and ratio estimators. It is also observed that the percentage relative efficiency decreases with increase in non-response.

References

1. Cochran, W. G. (1940) : The estimation of the yields of cereal experiments by

sampling for the ratio of grain in total produce, *Journal of The Agricultural Sciences*, 30, 262-275.

2. Cochran, W. G. (1977) : *Sampling Techniques*, 3rd *ed.*, John Wiley and Sons, New York.
3. Gupta, S. and Shabbir, J. (2008) : On improvement in estimating the population mean in simple random sampling, *Journal of Applied Statistics*, 35(5), 559-566.
4. Hansen, M. H. and Hurwitz, W. N. (1946) : The problem of non-response in sample surveys, *Journal of The American Statistical Association*, 41, 517-529.
5. Hansen, M. H., Hurwitz, W. N. and Madow,, W. G. (1953): *Sample Survey Methods and Theory*, Volume I, John Wiley and Sons, Inc., New York.
6. Kadilar, C. and Cingi, H. (2006) : New ratio estimators using correlation coefficient, *Interstat* 4, 1–11.
7. Khare, B. B. and Sinha, R. R. (2004) : Estimation of finite population ratio using two phase sampling scheme in the presence of non-response, *Aligarh Journal of Statistics* 24, 43–56.
8. Mohanty, S. (1967) : Combination of regression and ratio estimates, *Journal of the Indian Statistical Association* 5, 1–14.
9. Murthy, M. N. (1964) : Product method of estimation, *Sankhya*, 26A, 69-74.
10. Neyman, J. (1938). Contribution to the theory of sampling human populations, *Journal of American Statistical Association*, 33, 101-116.
11. Rao, J.N.K. (1973) : On double sampling for stratification and analytic surveys, *Biometrika*, 60, 125-133.
12. Ray, S. K. and Sahai, A. (1980) : Efficient families of ratio and product-type estimators, *Biometrika*, 67, 215-217.
13. Reddy, V. N. (1974) : On a transformed ratio method of estimation, *Sankhya*, 36C, 59-70.
14. Sahoo, J., Sahoo, L. N., Mohanty, S. (1993) : A regression approach to estimation in two phase sampling using two auxiliary variables. *Curr. Sci.* 65(1), 73–75.

15. Singh, G.N. and Upadhyaya, L.N. (1995) : A class of modified chain-type estimators using two auxiliary variables in two-phase sampling, *Metron*, Vol. LIII, No. 3-4, 117-125.
16. Singh, R., Kumar, M. and Smarandache, F. (2008): Almost Unbiased Estimator for Estimating Population Mean Using Known Value of Some Population Parameter(s). Pak. J. Stat. Oper. Res., 4(2) pp63-76.
17. Singh, R. and Kumar, M. (2011): A note on transformations on auxiliary variable in survey sampling. MASA, 6:1, 17-19. doi 10.3233/MASA-2011-0154
18. Srivastava, S. (1993) Some problems on the estimation of population mean using auxiliary character in presence of non-response in sample surveys. Ph. D. Thesis, *Banaras Hindu University, Varanasi, India*.
19. Srivastava, S.K. (1971) : A generalized estimator for the mean of a finite population using multi-auxiliary information, *Journal of The American Statistical Associatio*n, 66 (334), 404-407.
20. Srivastava, S.K. and Jhajj, H. S. (1981) : A class of estimators of the population mean in survey sampling using auxiliary information, *Biometrika*, 68 (1), 341-343.
21. Srivenkataramana, T. (1980) : A dual to ratio estimator in sample surveys, *Biometrika*, 67 (1), 199-204.
22. Sukhatme, B. V. (1962) : Some ratio-type estimators in two-phase sampling, *Journal of the American Statistical Association*, 57, 628–632.

A Family of Median Based Estimators in Simple Random Sampling

[1]Hemant K. Verma, [†1]Rajesh Singh and [2]Florentin Smarandache

Department of Statistics, Banaras Hindu University

Varanasi-221005, India

[2]Department of Mathematics, University of New Mexico, Gallup, USA

† Corresponding author, rsinghstat@yahoo.com

Abstract

In this paper we have proposed a median based estimator using known value of some population parameter(s) in simple random sampling. Various existing estimators are shown particular members of the proposed estimator. The bias and mean squared error of the proposed estimator is obtained up to the first order of approximation under simple random sampling without replacement. An empirical study is carried out to judge the superiority of proposed estimator over others.

Keywords: Bias, mean squared error, simple random sampling, median, ratio estimator.

1. Introduction

Consider a finite population $U = \{U_1, U_2, ..., U_N\}$ of N distinct and identifiable units. Let Y be the study variable with value Y_i measured on $U_i, i = 1,2,3..., N$. The problem is to estimate the population mean $\overline{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$. The simplest estimator of a finite population mean is the sample mean obtained from the simple random sampling without replacement, when there is no auxiliary information available. Sometimes there exists an auxiliary variable X which is positively correlated with the study variable Y. The information available on the auxiliary variable X may be utilized to obtain an efficient estimator of the population mean. The sampling theory describes a wide variety of techniques for using auxiliary information to obtain more efficient estimators. The ratio estimator and the regression estimator are the two important estimators available in the literature which are using the auxiliary information. To know more about the ratio and regression estimators and other related results one may refer to [1-13].

When the population parameters of the auxiliary variable X such as population mean, coefficient of variation, kurtosis, skewness and median are known, a number of modified ratio estimators are proposed in the literature, by extending the usual ratio and Exponential- ratio type estimators.

Before discussing further about the modified ratio estimators and the proposed median based modified ratio estimators the notations and formulae to be used in this paper are described below:

- N - Population size
- n - Sample size
- Y - Study variable
- X - Auxiliary variable
- $\beta_1 = \frac{\mu_3^2}{\mu_2^3}$ Where $\mu_r = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{X})^r$, Coefficient of skewness of the auxiliary variable
- ρ - Correlation Co-efficient between X and Y
- $\overline{X}, \overline{Y}$ - Population means
- $\overline{x}, \overline{y}$ - Sample means
- $\overline{M}$, - Average of sample medians of Y
- m - Sample median of Y
- β - Regression coefficient of Y on X
- B (.) - Bias of the estimator
- V (.) - Variance of the estimator
- MSE (.) - Mean squared error of the estimator
- $PRE(e,p) = \frac{MSE(e)}{MSE(e)} * 100$ - Percentage relative efficiency of the proposed estimator p with respect to the existing estimator e.

The formulae for computing various measures including the variance and the covariance of the SRSWOR sample mean and sample median are as follows:

$$V(\overline{y}) = \frac{1}{{}^{N}C_n}\sum_{i=1}^{{}^{N}C_n}(y_i - \overline{Y})^2 = \frac{1-f}{n}S_y^2, V(\overline{x}) = \frac{1}{{}^{N}C_n}\sum_{i=1}^{{}^{N}C_n}(x_i - \overline{X})^2 = \frac{1-f}{n}S_x^2, V(m) = \frac{1}{{}^{N}C_n}\sum_{i=1}^{{}^{N}C_n}(m_i - \overline{M})^2$$

,

$$Cov(\overline{y}, \overline{x}) = \frac{1}{{}^{N}C_n}\sum_{i=1}^{{}^{N}C_n}(x_i - \overline{X})(y_i - \overline{Y}) = \frac{1-f}{n}\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{X})(y_i - \overline{Y}),$$

$$Cov(\bar{y}, m) = \frac{1}{^{N}C_n} \sum_{i=1}^{^{N}C_n} (m_i - \bar{M})(y_i - \bar{Y}),$$

$$C'_{xx} = \frac{V(\bar{x})}{\bar{X}^2}, C'_{mm} = \frac{V(m)}{\bar{M}^2}, C'_{ym} = \frac{Cov(\bar{y}, m)}{\bar{M}\bar{Y}}, C'_{yx} = \frac{Cov(\bar{y}, \bar{x})}{\bar{X}\bar{Y}}$$

$$\text{Where } f = \frac{n}{N}; S_y^2 = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{Y})^2, S_x^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{X})^2 ,$$

In the case of simple random sampling without replacement (SRSWOR), the sample mean $\bar{y}$ is used to estimate the population mean $\bar{Y}$. That is the estimator of $\bar{Y} = \bar{Y}_r = \bar{y}$ with the variance

$$V(\bar{Y}_r) = \frac{1-f}{n} S_y^2 \tag{1.1}$$

The classical Ratio estimator for estimating the population mean $\bar{Y}$ of the study variable Y is defined as $\bar{Y}_R = \frac{\bar{y}}{\bar{x}} \bar{X}$. The bias and mean squared error of $\bar{Y}_R$ are given as below:

$$B(\bar{Y}_R) = \bar{Y}\left\{C'_{xx} - C'_{yx}\right\} \tag{1.2}$$

$$MSE(\bar{Y}_R) = V(\bar{y}) + R'^2 V(\bar{x}) - 2R' Cov(\bar{y}, \bar{x}) \qquad \text{where } R' = \frac{\bar{Y}}{\bar{X}} \tag{1.3}$$

2. Proposed estimator

Suppose

$$t_0 = \bar{y}, \quad t_1 = \bar{y}\left[\frac{\bar{M}^*}{\alpha m^* + (1-\alpha)\bar{M}^*}\right]^g, \quad t_2 = \bar{y} \exp\left[\frac{\delta(\bar{M}^* - m^*)}{\bar{M}^* + m^*}\right] \quad \text{where} \quad \bar{M}^* = a\bar{M} + b, \ m^* = am + b$$

Such that $t_0, t_1, t_2 \in w$, where w denotes the set of all possible ratio type estimators for estimating the population mean $\bar{Y}$. By definition the set w is a linear variety, if

$$t = w_0 \bar{y} + w_1 t_1 + w_2 t_2 \qquad \in W, \tag{2.1}$$

$$\text{for} \sum_{i=0}^{2} w_i = 1 \qquad w_i \in R \tag{2.2}$$

where w_i (i=0, 1, 2) denotes the statistical constants and R denotes the set of real numbers.

$$\text{Also, } t_1 = \bar{y}\left[\frac{\bar{M}^*}{\alpha m^* + (1-\alpha)\bar{M}^*}\right]^g, \quad t_2 = \bar{y} \exp\left[\frac{\delta(\bar{M}^* - m^*)}{\bar{M}^* + m^*}\right]$$

and $\overline{M}^* = a\overline{M} + b,\ m^* = am + b.$

To obtain the bias and MSE expressions of the estimator t, we write

$$\bar{y} = \bar{Y}(1+e_0), \qquad m = \overline{M}(1+e_1)$$

such that

$$E(e_0) = E(e_1) = 0,$$

$$E(e_0^2) = \frac{V(\bar{y})}{\bar{Y}^2},\ \ E(e_1^2) = \frac{V(m)}{\overline{M}^2} = C_{mm},\ \ E(e_0e_1) = \frac{Cov(\bar{y}, m)}{\bar{Y}\overline{M}} = C_{ym}$$

Expressing the estimator t in terms of e's, we have

$$t = \bar{Y}(1+e_0)\left[w_0 + w_1(1+\upsilon\alpha e_1)^{-g} + w_2 \exp\left\{\left(-\frac{\upsilon\delta e_1}{2}\right)\left(1+\frac{\upsilon e_1}{2}\right)^{-1}\right\}\right] \tag{2.3}$$

where $\upsilon = \dfrac{a\overline{M}}{a\overline{M}+b}$.

Expanding the right hand side of equation(2.3) up to the first order of approximation, we get

$$t \cong \bar{Y}\left[1 - \upsilon w e_1 + e_0 + \upsilon^2\left(w_1\frac{g(g+1)}{2}\alpha^2 + \left(\frac{\delta}{4} - \frac{\delta^2}{8}\right)w_2\right)e_1^2 - \upsilon w e_0 e_1\right] \tag{2.4}$$

where $\quad w = \alpha g w_1 + \dfrac{\delta}{2}w_2.$ (2.5)

Taking expectations of both sides of (2.4) and then subtracting $\bar{Y}$ from both sides, we get the biases of the estimators, up to the first order of approximation as

$$B(t) = \bar{Y}\left[\upsilon^2\left\{w_1\frac{g(g+1)}{2}\alpha^2 + \left(\frac{\delta}{4} - \frac{\delta^2}{8}\right)w_2\right\}C_{mm} - \upsilon w C_{ym}\right] \tag{2.6}$$

$$B(t_1) = \bar{Y}g\alpha\upsilon\left[\frac{\alpha\upsilon(g+1)}{2}C_{mm} - C_{ym}\right] \tag{2.7}$$

$$B(t_2) = \bar{Y}\left[\left(\frac{\delta\upsilon^2}{4} + \frac{\delta^2\upsilon^2}{8}\right)C_{mm} - \frac{\delta\upsilon}{2}C_{ym}\right] \tag{2.8}$$

From (2.4), we have

$$t - \bar{Y} \cong \bar{Y}(e_0 + \upsilon w e_1) \tag{2.9}$$

Squaring both sides of (2.9) and then taking expectations, we get the MSE of the estimator t, up to the first order of approximation as

$$MSE(t) = V(\bar{y}) + \upsilon^2 R^2 w^2 V(m) - 2\upsilon R w Cov(\bar{y}, m) \tag{2.10}$$

where $R = \dfrac{\bar{Y}}{\bar{M}}$.

MSE(t) will be minimum, when

$$w = \frac{1}{\upsilon R} \frac{Cov(\bar{y}, m)}{V(m)} = k(say) \tag{2.11}$$

Putting the value of w(=k) in (2.10), we get the minimum MSE of the estimator t, as

$$\min . MSE(t) = V(\bar{y})(1 - \rho^2) \tag{2.12}$$

The minimum MSE of the estimator t is same as that of traditional linear regression estimator.

From (2.5) and (2.11), we have

$$\alpha g w_1 + \frac{\delta}{2} w_2 = k \tag{2.13}$$

From (2.2) and (2.13), we have only two equations in three unknowns. It is not possible to find the unique values of $w_{i's}$ (i=0, 1, 2). In order to get unique values for $w_{i's}$, we shall impose the linear restriction

$$w_0 B(\bar{y}) + w_1 B(t_1) + w_2 B(t_2) = 0 \tag{2.14}$$

Equations (2.2), (2.11) and (2.14) can be written in matrix form as

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & \alpha g & \frac{\delta}{2} \\ 0 & \mathbf{B}(t_1) & B(t_2) \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 1 \\ k \\ 0 \end{bmatrix} \tag{2.15}$$

Using (2.15) we get the unique value of $w_{i's}$ (i=0, 1, 2) as

$$\left.\begin{aligned} w_0 &= \frac{\Delta_0}{\Delta_r} \\ w_1 &= \frac{\Delta_1}{\Delta_r} \\ w_2 &= \frac{\Delta_2}{\Delta_r} \end{aligned}\right\} \quad \text{where} \quad \begin{aligned} \Delta_r &= \alpha g B(t_2) - \frac{\delta}{2} B(t_1) \\ \Delta_0 &= B(t_2)(\alpha g - k) + \frac{1}{2} B(t_1)\left(k - \frac{\delta}{2}\right) \\ \Delta_1 &= k B(t_2) \\ \Delta_2 &= -k B(t_1) \end{aligned} \qquad (2.16)$$

Table 2.1: Some members of the proposed estimator

w_0	w_1	w_2	a	b	α	g	δ	Estimators
1	0	0	-	-	-	-	-	$q_1 = \bar{y}$
0	1	0	1	0	1	1	-	$q_2 = \bar{y}\frac{\bar{M}}{m}$
0	1	0	β_1	ρ	1	1	-	$q_3 = \bar{y}\left[\frac{\beta_1\bar{M} + \rho}{\beta_1 m + \rho}\right]$
0	1	0	ρ	β_1	1	1	-	$q_4 = \bar{y}\left[\frac{\rho\bar{M} + \beta_1}{\rho m + \beta_1}\right]$
0	0	1	1	0	-	-	1	$q_5 = \bar{y}\exp\left[\frac{(\bar{M} - m)}{\bar{M} + m}\right]$
0	0	1	β_1	ρ	-	-	1	$q_6 = \bar{y}\exp\left[\frac{\beta_1(\bar{M} - m)}{\beta_1(\bar{M} + m) + 2\rho}\right]$
0	0	1	ρ	β_1	-	-	1	$q_7 = \bar{y}\exp\left[\frac{\rho(\bar{M} - m)}{\rho(\bar{M} + m) + 2\beta_1}\right]$
0	1	1	β_1	ρ	1	1	1	$q_8 = \bar{y}\left[\frac{\beta_1\bar{M} + \rho}{\beta_1 m + \rho}\right] + \bar{y}\exp\left[\frac{\beta_1(\bar{M} - m)}{\beta_1(\bar{M} + m) + 2\rho}\right]$
0	1	1	ρ	β_1	1	1	1	$q_9 = \bar{y}\left[\frac{\rho\bar{M} + \beta_1}{\rho m + \beta_1}\right] + \bar{y}\exp\left[\frac{\rho(\bar{M} - m)}{\rho(\bar{M} + m) + 2\beta_1}\right]$
0	1	1	1	0	1	1	1	$q_{10} = \bar{y}\frac{\bar{M}}{m} + \bar{y}\exp\left[\frac{(\bar{M} - m)}{\bar{M} + m}\right]$

3. Empirical Study

For numerical illustration we consider: the population 1 and 2 taken from [14] pageno.177, the population 3 is taken from [15] page no.104. The parameter values and constants computed for the above populations are given in the Table 3.1. MSE for the proposed and existing estimators computed for the three populations are given in the Table 3.2 whereas the PRE for the proposed and existing estimators computed for the three populations are given in the Table 3.3.

Table: 3.1 Parameter values and constants for 3 different populations

Parameters	For sample size n=3			For sample size n=5		
	Popln-1	**Popln-2**	**Popln-3**	**Popln-1**	**Popln-2**	**Popln-3**
N	34	34	20	34	34	20
n	3	3	3	5	5	5
${}^{N}C_{n}$	5984	5984	1140	278256	278256	15504
$\overline{Y}$	856.4118	856.4118	41.5	856.4118	856.4118	41.5
$\overline{M}$	747.7223	747.7223	40.2351	736.9811	736.9811	40.0552
$\overline{X}$	208.8824	199.4412	441.95	208.8824	199.4412	441.95
β_1	0.8732	1.2758	1.0694	0.8732	1.2758	1.0694
R	1.1453	1.1453	1.0314	1.1621	1.1621	1.0361
$V(\overline{y})$	163356.4086	163356.4086	27.1254	91690.3713	91690.3713	14.3605
$V(\overline{x})$	6884.4455	6857.8555	2894.3089	3864.1726	3849.248	1532.2812
V(m)	101127.6164	101127.6164	26.0605	58464.8803	58464.8803	10.6370
$Cov(\overline{y}, m)$	90236.2939	90236.2939	21.0918	48074.9542	48074.9542	9.0665
$Cov(\overline{y}, \overline{x})$	15061.4011	14905.0488	182.7425	8453.8187	8366.0597	96.7461
ρ	0.4491	0.4453	0.6522	0.4491	0.4453	0.6522

Table: 3.2. Variance / Mean squared error of the existing and proposed estimators

Estimators	For sample size n=3			For sample size n=5		
	Population-1	Population-2	Population-3	Population-1	Population-2	Population-3
q_1	163356.41	163356.41	27.13	91690.37	91690.37	14.36
q_2	89314.58	89314.58	11.34	58908.17	58908.17	6.99
q_3	89274.35	89287.26	11.17	58876.02	58886.34	6.93
q_4	89163.43	89092.75	10.92	58787.24	58730.58	6.85
q_5	93169.40	93169.40	12.30	55561.98	55561.98	7.82
q_6	93194.86	93186.68	12.42	55573.42	55569.74	7.88
q_7	93265.64	93311.19	12.62	55605.24	55625.75	7.97
q_8	113764.16	113810.72	21.52	76860.57	76891.47	10.66
q_9	151049.79	150701.09	22.00	101236.37	101004.87	10.99
q_{10}	151791.97	151791.97	24.24	101728.97	101728.97	11.87
t(opt)	82838.45	82838.45	10.05	52158.93	52158.93	6.63

Table: 3.3. Percentage Relative Efficiency of estimators with respect to $\bar{y}$

Estimators	For sample size n=3			For sample size n=5		
	Population-1	Population-2	Population-3	Population-1	Population-2	Population-3
q_1	100	100	100	100	100	100
q_2	182.90	182.90	239.191236	155.65	155.65	205.40
q_3	182.98	182.96	242.877047	155.73	155.71	207.12
q_4	183.21	183.36	248.504702	155.97	156.12	209.64
q_5	175.33	175.33	220.500742	165.02	165.02	183.60
q_6	175.28	175.30	218.381298	164.99	165.00	182.30
q_7	175.15	175.07	214.915968	164.90	164.83	180.16
q_8	143.59	143.53	126.034732	119.29	119.25	134.70

q_9	108.15	108.40	123.254986	90.57	90.78	130.57
q_{10}	107.62	107.62	111.896010	90.13	90.13	120.97
t(opt)	197.20	197.20	269.771157	175.79	175.79	216.51

4. Conclusion

From empirical study we conclude that the proposed estimator under optimum conditions perform better than other estimators considered in this paper. The relative efficiencies and MSE of various estimators are listed in Table 3.2 and 3.3.

References

1. Murthy M.N. (1967). Sampling theory and methods. Statistical Publishing Society, Calcutta, India.
2. Cochran, W. G. (1977): Sampling Techniques. Wiley Eastern Limited.
3. Khare B.B. and Srivastava S.R. (1981): A general regression ratio estimator for the population mean using two auxiliary variables. Alig. J. Statist.,1: 43-51.
4. Sisodia, B.V.S. and Dwivedi, V.K. (1981): A modified ratio estimator using co-efficient of variation of auxiliary variable. Journal of the Indian Society of Agricultural Statistics 33(1), 13-18.
5. Singh G.N. (2003): On the improvement of product method of estimation in sample surveys. Journal of the Indian Society of Agricultural Statistics 56 (3), 267–265.
6. Singh H.P. and Tailor R. (2003): Use of known correlation co-efficient in estimating the finite population means. Statistics in Transition 6 (4), 555-560.
7. Singh H.P., Tailor R., Tailor R. and Kakran M.S. (2004): An improved estimator of population mean using Power transformation. Journal of the Indian Society of Agricultural Statistics 58(2), 223-230.
8. Singh, H.P. and Tailor, R. (2005): Estimation of finite population mean with known co-efficient of variation of an auxiliary. STATISTICA, anno LXV, n.3, pp 301-313.
9. Kadilar C. and Cingi H. (2004): Ratio estimators in simple random sampling. Applied Mathematics and Computation 151, 893-902.
10. Koyuncu N. and Kadilar C. (2009): Efficient Estimators for the Population mean. Hacettepe Journal of Mathematics and Statistics, Volume 38(2), 217-225.
11. Singh R., Kumar M. and Smarandache F. (2008): Almost unbiased estimator for estimating population mean using known value of some population parameter(s). Pak.j.stat.oper.res., Vol.IV, No.2, pp 63-76.

12. Singh, R. and Kumar, M. (2011): A note on transformations on auxiliary variable in survey sampling. Mod. Assis. Stat. Appl., 6:1, 17-19. doi 10.3233/MAS-2011-0154.
13. Singh R., Malik S., Chaudhary M.K., Verma H.K., and Adewara A.A. (2012): A general family of ratio-type estimators in systematic sampling. Jour. Reliab. Stat. Ssci., 5(1):73-82.
14. Singh, D. and Chaudhary, F. S. (1986): Theory and analysis of survey designs. Wiley Eastern Limited.
15. Mukhopadhyay, P. (1998): Theory and methods of survey sampling. Prentice Hall.

DIFFRENCE-TYPE ESTIMATORS FOR ESTIMATION OF MEAN IN THE PRESENCE OF MEASUREMENT ERROR

[1]Viplav Kr. Singh, †[1]Rajesh Singh and [2]Florentin Smarandache

Department of Statistics, Banaras Hindu University

Varanasi-221005, India

[2]Department of Mathematics, University of New Mexico, Gallup, USA

† Corresponding author, rsinghstat@yahoo.com

Abstract

In this paper we have suggested difference-type estimator for estimation of population mean of the study variable y in the presence of measurement error using auxiliary information. The optimum estimator in the suggested estimator has been identified along with its mean square error formula. It has been shown that the suggested estimator performs more efficient then other existing estimators. An empirical study is also carried out to illustrate the merits of proposed method over other traditional methods.

Key Words: Study variable, Auxiliary variable, Measurement error, Simple random Sampling, Bias, Mean Square error.

1. PERFORMANCE OF SUGGESTED METHOD USING SIMPLE RANDOM SAMPLING

INTRODUCTION

The present study deals with the impact of measurement errors on estimating population mean of study variable (y) in simple random sampling using auxiliary information. In theory of survey sampling, the properties of estimators based on data are usually presupposed that the observations are the correct measurement on the characteristic being studied. When the measurement errors are negligible small, the statistical inference based on observed data continue to remain valid.

An important source of measurement error in survey data is the nature of variables (study and auxiliary). Here nature of variable signifies that the exact measurement on variables is not available. This may be due to the following three reasons:

1. The variable is clearly defined but it is hard to take correct observation at least with the currently available techniques or because of other types of practical difficulties. Eg: The level of blood sugar in a human being.

2. The variable is conceptually well defined but observation can obtain only on some closely related substitutes known as Surrogates. Eg: The measurement of economic status of a person.

3. The variable is fully comprehensible and well understood but it is not intrinsically defined. Eg: Intelligence, aggressiveness etc.

Some authors including Singh and Karpe (2008, 2009), Shalabh(1997), Allen et al. (2003), Manisha and Singh (2001, 2002), Srivastava and Shalabh (2001), Kumar et al. (2011 a,b), Malik and Singh (2013), Malik et al. (2013) have paid their attention towards the estimation of population mean μ_y of study variable using auxiliary information in the presence of measurement errors. Fuller (1995) examined the importance of measurement errors in estimating parameters in sample surveys. His major concerns are estimation of population mean or total and its standard error, quartile estimation and estimation through regression model.

SYMBOLS AND SETUP

Let, for a SRS scheme (x_i, y_i) be the observed values instead of true values (X_i, Y_i) on two characteristics (x, y), respectively for all i=(1,2,…n) and the observational or measurement errors are defined as

$$u_i = (y_i - Y_i) \quad (1)$$

$$v_i = (x_i - X_i) \quad (2)$$

where u_i and v_i are stochastic in nature with mean 0 and variance σ_u^2 and σ_v^2 respectively. For the sake of convenience, we assume that u_i's and v_i's are uncorrelated although X_i's and Y_i's are correlated .Such a specification can be, however, relaxed at the cost of some algebraic complexity. Also assume that finite population correction can be ignored.

Further, let the population means and variances of (x, y) be (μ_x, μ_y) and (σ_x^2, σ_y^2). σ_{xy} and ρ be the population covariance and the population correlation coefficient between x and y respectively. Also let $C_y = \frac{\sigma_y}{\mu_y}$ and $C_x = \frac{\sigma_x}{\mu_x}$ are the population coefficient of variation and C_{yx} is the population coefficients of covariance in y and x.

LARGE SAMPLE APPROXIMATION

Define:

$$e_0 = \frac{\bar{y} - \mu_y}{\mu_y} \text{ and } e_1 = \frac{\bar{x} - \mu_x}{\mu_x}$$

where, e_0 and e_1 are very small numbers and $|e_i| < 1 \ \ (i = 0,1)$.

Also, $E(e_i) = 0 \, (i = 0,1)$

and, $E(e_0^2) = \theta C_y^2 \left(1 + \frac{\sigma_u^2}{\sigma_y^2}\right) = \delta_0$,

$$E(e_1^2) = \theta C_x^2 \left(1 + \frac{\sigma_v^2}{\sigma_x^2}\right) = \delta_1, \; E(e_0 e_1) = \theta \rho C_x C_y, \text{ where } \theta = \frac{1}{n}.$$

2. EXISTING ESTIMATORS AND THEIR PROPERTIES

Usual mean estimator is given by

$$\bar{y} = \sum_{i=1}^{n} \frac{y_i}{n} \tag{3}$$

Up to the first order of approximation the variance of $\bar{y}$ is given by

$$Var(\bar{y}) = \theta \mu_y^2 \left(1 + \frac{\sigma_u^2}{\sigma_y^2}\right) C_y^2 \tag{4}$$

The usual ratio estimator is given by

$$\bar{y}_R = \bar{y}\left(\frac{\mu_x}{\bar{x}}\right) \tag{5}$$

where μ_x is known population mean of x.

The bias and MSE ($\bar{y}_R$), to the first order of approximation, are respectively, given

$$B(\bar{y}_R) = \theta \mu_y \left[\left(1 + \frac{\sigma_v^2}{\sigma_x^2}\right) C_x^2 - \rho C_y C_x\right] \tag{6}$$

$$MSE(\bar{y}_R) = \theta \mu_y^2 \left[\left(1 + \frac{\sigma_u^2}{\sigma_y^2}\right) C_y^2 + \left(1 + \frac{\sigma_v^2}{\sigma_x^2}\right) C_x^2 - 2\rho C_y C_x\right] \tag{7}$$

The traditional difference estimator is given by

$$\bar{y}_d = \bar{y} + k(\mu_x - \bar{x}) \tag{8}$$

where, k is the constant whose value is to be determined.

Minimum mean square error of $\bar{y}_d$ at optimum value of

$$k = \frac{\mu_y \rho C_y}{\mu_x \left(1 + \frac{\sigma_v^2}{\sigma_x^2}\right) C_x}, \quad \text{is given by}$$

$$MSE(\bar{y}_d) = \mu_y^2 \, \theta \left(1 + \frac{\sigma_u^2}{\sigma_y^2}\right) C_y^2 \left[1 - \frac{\rho^2}{\left(1 + \frac{\sigma_u^2}{\sigma_y^2}\right)\left(1 + \frac{\sigma_v^2}{\sigma_x^2}\right)}\right] \tag{9}$$

Srivastava (1967) suggested an estimator

$$\bar{y}_S = \bar{y}\left(\frac{\mu_x}{\bar{x}}\right)^{\ell_1} \tag{10}$$

where, ℓ_1 is an arbitrary constant.

Up to the first of approximation, the bias and minimum mean square error of $\bar{y}_S$ at optimum value of $\ell_1 = \frac{\rho C_y}{\left(1 + \frac{\sigma_v^2}{\sigma_x^2}\right) C_x}$ are respectively, given by

$$B(\bar{y}_S) = \mu_y \left[\frac{\ell_1(\ell_1 + 1)}{2} \theta \left(1 + \frac{\sigma_v^2}{\sigma_x^2}\right) C_x^2 - \ell_1 \theta \rho C_y C_x\right] \tag{11}$$

$$MSE(\bar{y}_S) = \mu_y^2 \, \theta \left(1 + \frac{\sigma_u^2}{\sigma_y^2}\right) C_y^2 \left[1 - \frac{\rho^2}{\left(1 + \frac{\sigma_u^2}{\sigma_y^2}\right)\left(1 + \frac{\sigma_v^2}{\sigma_x^2}\right)}\right] \tag{12}$$

Walsh (1970) suggested an estimator $\bar{y}_w$

$$\bar{y}_w = \bar{y}\left[\frac{\mu_x}{\ell_2 \bar{x} + (1 - \ell_2)\mu_x}\right]$$

(13)

where, ℓ_2 is an arbitrary constant.

Up to the first order of approximation, the bias and minimum mean square error of $\bar{y}_w$ at optimum value of $\ell_2 = \frac{\rho C_y}{\left(1+\frac{\sigma_v^2}{\sigma_x^2}\right)C_x}$, are respectively, given by

$$B(\bar{y}_w) = \mu_y \theta \left[\ell_2^2 C_x^2 \left(1+\frac{\sigma_v^2}{\sigma_x^2}\right) - \ell_2 \rho C_y C_x \right] \quad (14)$$

$$MSE(\bar{y}_w) = \mu_y^2 \theta \left(1+\frac{\sigma_u^2}{\sigma_y^2}\right) C_y^2 \left[1 - \frac{\rho^2}{\left(1+\frac{\sigma_u^2}{\sigma_y^2}\right)\left(1+\frac{\sigma_v^2}{\sigma_x^2}\right)}\right] \quad (15)$$

Ray and Sahai (1979) suggested the following estimator

$$\bar{y}_{RS} = (1-\ell_3)\bar{y} + \ell_3 \bar{y}\left(\frac{\bar{x}}{\mu_x}\right) \quad (16)$$

where, ℓ_3 is an arbitrary constant.

Up to the first order of approximation, the bias and mean square of $\bar{y}_{RS}$ at optimum value of $\ell_3 = -\frac{\rho C_y}{\left(1+\frac{\sigma_v^2}{\sigma_x^2}\right)}$ are respectively, given by

$$B(\bar{y}_{RS}) = \theta \ell_3 \mu_y \rho C_y C_x \quad (17)$$

$$MSE(\bar{y}_{RS}) = \mu_y^2 \theta \left(1+\frac{\sigma_u^2}{\sigma_y^2}\right) C_y^2 \left[1 - \frac{\rho^2}{\left(1+\frac{\sigma_u^2}{\sigma_y^2}\right)\left(1+\frac{\sigma_v^2}{\sigma_x^2}\right)}\right] \quad (18)$$

3. SUGGESTED ESTIMATOR

Following Singh and Solanki (2013), we suggest the following difference-type class of estimators for estimating population mean $\bar{Y}$ of study variable y as

$$t_p = \left[\alpha_1 \bar{y} + \alpha_2 \bar{x}^* + (1-\alpha_1-\alpha_2)\mu_x^*\right]\left[\frac{\mu_x^*}{\bar{x}^*}\right]^{\alpha} \tag{19}$$

where (α_1, α_2) are suitably chosen scalars such that MSE of the proposed estimator is minimum, $\bar{x}^* (= \eta\bar{x} + \lambda)$, $\mu_x^* (= \eta\mu_x + \lambda)$ with (n, λ) are either constants or function of some known population parameters. Here it is interesting to note that some existing estimators have been shown as the members of proposed class of estimators t_p for different values of $(\alpha_1, \alpha_2, \alpha, \eta, \lambda)$, which is summarized in Table 1.

Table 1: Members of suggested class of estimators

	Values of Constants				
Estimators	α_1	α_2	α	η	λ
$\bar{y}$ [Usual unbiased]	1	0	0	-	-
$\bar{y}_R$ [Usual ratio]	1	0	1	1	0
$\bar{y}_d$ [Usual difference]	1	α_2	0	-1	μ_x
$\bar{y}_S$ [Srivastava (1967)]	1	0	α	1	0
$\bar{y}_{DS}$ [Dubey and Singh]	α_1	α_2	0	1	0

The properties of suggested estimator are derived in the following theorems.

Theorem 1.1: Estimator t_p in terms of $e_i; i = 0,1$ expressed as:

$$t_p = \left[\mu_x^* - \alpha A e_1 \mu_x^* + B\mu_x^* e_1^2 + \alpha_1\left\{C - \alpha A C e_1 + B C e_1^2 + e_0\mu_y - \alpha A \mu_y e_0 e_1\right\}\right.$$
$$\left. + \alpha_2 \eta \mu_x \left\{e_1 - \alpha A e_1^2\right\}\right]$$

ignoring the terms $E(e_i^r e_j^s)$ for (r+s)>2,where r,s=0,1,2... and $i=0,1;\, j=1$ (first order of approximation).

where, $A=\dfrac{\eta\mu_x}{\eta\mu_x+\lambda}$, $B=\dfrac{\alpha(\alpha+1)}{2}A^2$ and $C=\mu_y-\mu_x^*$.

Proof

$$t_p=\left[\alpha_1\bar{y}+\alpha_2\bar{x}^*+(1-\alpha_1-\alpha_2)\mu_x^*\right]\left[\frac{\mu_x^*}{\bar{x}^*}\right]^{\alpha}$$

Or

$$t_p=\left[\alpha_1(1+e_0)+\alpha_2\eta\,\mu_x e_1+(1-\alpha_1)\mu_x^*\right]\left[1+Ae_1\right]^{-\alpha} \tag{20}$$

We assume $|A e_1|<1$, so that the term $(1+Ae_1)^{-\alpha}$ is expandable. Expanding the right hand side (20) and neglecting the terms of e's having power greater than two, we have

$$t_p=\mu_x^*-\alpha Ae_1\mu_x^*+B\mu_x^*e_1^2+\alpha_1\left\{C-\alpha ACe_1+BCe_1^2+e_0\mu_y-\alpha A\mu_y e_0 e_1\right\}$$
$$+\alpha_2\eta\mu_x\left\{e_1-\alpha Ae_1^2\right\}$$

Theorem: 1.2 Bias of the estimator t_p is given by

$$B(t_p)=\left[B\mu_x^*\delta_1+\alpha_1\left\{BC\delta_1-\alpha A\mu_y\delta_{01}\right\}-\alpha_2\eta\mu_x A\alpha\delta_1\right] \tag{21}$$

Proof:

$$B(t_p)=E(t_p-\mu_y)$$

$$=E\left[\mu_x^*-\mu_y-\alpha Ae_1\mu_x^*+B\mu_x^*e_1^2+\alpha_1\left\{C-\alpha ACe_1+BCe_1^2+e_0\mu_y-\alpha A\mu_y e_0 e_1\right\}\right.$$
$$\left.+\alpha_2\eta\mu_x\left\{e_1-\alpha Ae_1^2\right\}\right]$$

$$=\left[B\mu_x^*\delta_1+\alpha_1\left\{BC\delta_1-\alpha A\mu_y\delta_{01}\right\}-\alpha_2\eta\mu_x A\alpha\delta_1\right]$$

where, δ_0, δ_1 and δ_{01} are already defined in section 3.

Theorem 1.3: MSE of the estimator t_p, up to the first order of approximation is

$$MSE(t_p) = \alpha_1^2\left\{C^2 + \mu_y^2\delta_0 + \delta_1\left(\alpha^2A^2C^2 + 2BC^2\right) - 4\alpha AC\mu_y\delta_{01}\right\} + \alpha_2^2\eta^2\mu_x^2\delta_1$$

$$+ \left\{C^2 + \delta_1\left(\alpha^2A^2\mu_x^2 - 2BC\mu_x^*\right)\right\} - 2\alpha_1\left\{C^2 + \delta_1\left(BC^2 - BC\mu_x^* - \alpha^2A^2C\mu_x^*\right) + \delta_{01}\alpha A\mu_y\left(\mu_x^* - C\right)\right\}$$

$$- 2\alpha_2\eta\mu_x\alpha A\delta_1\left(\mu_x^* - C\right) + 2\alpha_1\alpha_2\eta\mu_x\left(\mu_y\delta_{01} - 2A\alpha C\delta_1\right) \quad (22)$$

Proof:

$$MSE(t_p) = E(t_p - \mu_y)^2$$

$$= E\left[\alpha_1\left\{C - A\alpha Ce_1 + e_0\mu_y + BCe_1^2 - \alpha A\mu_y e_0e_1\right\} + \alpha_2\eta\mu_x\left\{e_1 - A\alpha e_1^2\right\}\right.$$

$$\left. - C + \alpha Ae_1\mu_x^* - B\mu_x^*e_1^2\right]^2$$

Squaring and then taking expectations of both sides, we get the MSE of the suggested estimator up to the first order of approximation as

$$MSE(t_p) = \alpha_1^2\left\{C^2 + \mu_y^2\delta_0 + \delta_1\left(\alpha^2A^2C^2 + 2BC^2\right) - 4\alpha AC\mu_y\delta_{01}\right\} + \alpha_2^2\eta^2\mu_x^2\delta_1$$

$$+ \left\{C^2 + \delta_1\left(\alpha^2A^2\mu_x^2 - 2BC\mu_x^*\right)\right\} - 2\alpha_1\left\{C^2 + \delta_1\left(BC^2 - BC\mu_x^* - \alpha^2A^2C\mu_x^*\right) + \delta_{01}\alpha A\mu_y\left(\mu_x^* - C\right)\right\}$$

$$- 2\alpha_2\eta\mu_x\alpha A\delta_1\left(\mu_x^* - C\right) + 2\alpha_1\alpha_2\eta\mu_x\left(\mu_y\delta_{01} - 2A\alpha C\delta_1\right)$$

Equation (22) can be written as:

$$MSE(t_p) = \alpha_1^2\varphi_1 + \alpha_2^2\varphi_2 - 2\alpha_1\varphi_3 - 2\alpha_2\varphi_4 + 2\alpha_1\alpha_2\varphi_5 + \varphi \quad (23)$$

Differentiating (23) with respect to (α_1, α_2) and equating them to zero, we get the optimum values of (α_1, α_2) as

$$\alpha_{1(opt)} = \frac{\varphi_2\varphi_3 - \varphi_4\varphi_5}{\varphi_1\varphi_2 - \varphi_5^2} \text{ and } \alpha_{2(opt)} = \frac{\varphi_1\varphi_4 - \varphi_3\varphi_5}{\varphi_1\varphi_2 - \varphi_5^2}$$

where, $\varphi_1 = C^2 + \mu_y^2\delta_0 + \delta_1\left(\alpha^2A^2C^2 + 2BC^2\right) - 4\alpha AC\mu_y\delta_{01}$

$\varphi_2 = \eta^2\mu_x^2\delta_1$

$$\varphi_3 = C^2 + \delta_1\left(BC^2 - BC\mu_x^* - \alpha^2 A^2 C\mu_x^*\right) + \delta_{01}\alpha A\mu_y\left(\mu_x^* - C\right)$$

$$\varphi_4 = \eta\mu_x \alpha A\delta_1\left(\mu_x^* - C\right)$$

$$\varphi_5 = \eta\mu_x\left(\mu_y\delta_{01} - 2A\alpha C\delta_1\right).$$

$$\varphi = C^2 + \delta_1\left(\alpha^2 A^2\mu_x^2 - 2BC\mu_x^*\right)$$

In the Table 2 some estimators are listed which are particular members of the suggested class of estimators t_p for different values of (α, η, λ).

Table 2: Particular members of the suggested class of estimators t_p

Estimators	**Values of constants**		
	α	η	λ
$t_1 = [\alpha_1\bar{y} + \alpha_2\bar{x} + (1-\alpha_1-\alpha_2)\mu_x]\left[\dfrac{\mu_x}{\bar{x}}\right]$	-1	1	0
$t_2 = [\alpha_1\bar{y} + \alpha_2(\bar{x}+1) + (1-\alpha_1-\alpha_2)(\mu_x+1)]\left[\dfrac{\mu_x+1}{\bar{x}+1}\right]$	1	1	1
$t_3 = [\alpha_1\bar{y} + \alpha_2(\bar{x}+1) + (1-\alpha_1-\alpha_2)(\mu_x+1)]\left[\dfrac{\mu_x}{\bar{x}}\right]^{-1}$	-1	1	1
$t_4 = [\alpha_1\bar{y} + \alpha_2(\bar{x}+\rho) + (1-\alpha_1-\alpha_2)(\mu_x+\rho)]\left[\dfrac{\mu_x+\rho}{\bar{x}+\rho}\right]^{-1}$	-1	1	ρ
$t_5 = [\alpha_1\bar{y} + \alpha_2(\bar{x}+C_x) + (1-\alpha_1-\alpha_2)(\mu_x+C_x)]\left[\dfrac{\mu_x+C_x}{\bar{x}+C_x}\right]^{-1}$	-1	1	C_x

$t_6 = \left[\alpha_1\bar{y} + \alpha_2(\bar{x} - C_x) + (1 - \alpha_1 - \alpha_2)(\mu_x - C_x)\right]\left[\dfrac{\mu_x - C_x}{\bar{x} - C_x}\right]$	-1	1	$-C_x$
$t_7 = \left[\alpha_1\bar{y} - \alpha_2(\bar{x} + 1) - (1 - \alpha_1 - \alpha_2)(\mu_x + C_x)\right]\left[\dfrac{\mu_x + C_x}{\bar{x} + C_x}\right]^{-1}$	-1	-1	-1

4. EMPIRICAL STUDY

Data statistics: The data used for empirical study has been taken from Gujarati (2007)

Where, Y_i =True consumption expenditure,

X_i =True income,

y_i =Measured consumption expenditure,

x_i = Measured income.

n	μ_y	μ_x	σ_y^2	σ_x^2	ρ	σ_u^2	σ_v^2
10	127	170	1278	3300	0.964	36	36

The percentage relative efficiencies (PRE) of various estimators with respect to the mean per unit estimator of $\bar{Y}$, that is $\bar{y}$,can be obtained as

$$PRE(.) = \frac{Var(\bar{y})}{MSE(.)} * 100$$

Table 3: MSE and PRE of estimators with respect to $\bar{y}$

Estimators	Mean Square Error	Percent Relative Efficiency
$\bar{y}$	131.4	100
$\bar{y}_R$	21.7906	603.0118
$\bar{y}_d$	13.916	944.1285

$\bar{y}_S$	13.916	944.1285
$\bar{y}_{DS}$	13.916	944.1285
t_1	10.0625	1236.648
t_2	9.92677	1323.693
t_3	**6.82471**	1925.356
t_4	6.9604	1887.818
t_5	9.3338	1407. 774
t_6	11.9246	1101.923
t_7	7.9917	1644.194

5. PERFORMANCE OF SUGGESTED ESTIMATOR IN STRATIFIED RANDOM SAMPLING

SYMBOLS AND SETUP

Consider a finite population $U = (u_1, u_2, ..., u_N)$ of size N and let X and Y respectively be the auxiliary and study variables associated with each unit $u_j = (j = 1,2,......,N)$ of population. Let the population of N be stratified in to L strata with the h^{th} stratum containing N_h units, where h=,1,2,3,….,L such that $\sum_{i=1}^{L} N_h = N$. A simple random size n^h is drown without replacement from the h^{th} stratum such that $\sum_{i=1}^{L} n_h = n$. Let (y_{hi}, X_{hi}) of two characteristics (Y,X) on i^{th} unit of the h^{th} stratum, where i=,1,2,…,N_h. In addition let

$$(\bar{y}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} y_{hi}, \bar{x}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} x_{hi}),$$

$$(\bar{y}_{st} = \sum_{i=1}^{n_h} W_h \bar{y}_h, \bar{x}_{st} = \sum_{i=1}^{n_h} W_h \bar{x}_h),$$

$(\mu_{Yh} = \frac{1}{N_h}\sum_{i=1}^{N_h} y_{hi}, \mu_{Xh} = \frac{1}{N_h}\sum_{i=1}^{N_h} x_{hi}$),

And ($\mu_Y = \sum_{h=1}^{L} W_h \mu_{Yh}, \mu_X = \sum_{i=1}^{N_h} W_h \mu_{Xh}$) be the samples means and population means of(Y,X) respectively, where $W_h = \frac{N_h}{N}$ is the stratum weight. Let the observational or measurement errors be

$$u_{hi} = y_{hi} - Y_{hi} \quad (24)$$

$$v_{hi} = x_{hi} - X_{hi} \quad (25)$$

Where u_{hi} and v_{hi} are stochastic in nature and are uncorrelated with mean zero and variances $\sigma_{Vh}{}^2$ and $\sigma_{Uh}{}^2$ respectively. Further let ρ_h be the population correlation coefficient between Y and X in the hth stratum. It is also assumed that the finite population correction terms $(1-f_h)$ and $(1-f)$ can be ignored where $f_h = \frac{n_h}{N_h}$ and $f = \frac{n}{N}$.

LARGE SAMPLE APPROXIMATION

Let,

$\bar{y}_{st} = \mu_Y(1+e_{0h})$, and $\bar{x}_{st} = \mu_X(1+e_{1h})$

such that, $E(e_{0h}) = E(e_{1h}) = 0$,

$$E(e_{0h}^2) = \frac{C_{Yh}^2}{n_h}\left(1+\frac{\sigma_{Uh}^2}{\sigma_{Yh}^2}\right) = \frac{C_{Yh}^2}{n_h\theta_{Yh}} = \nabla_0,$$

$$E(e_{1h}^2) = \frac{C_{Xh}^2}{n_h}\left(1+\frac{\sigma_{Vh}^2}{\sigma_{Xh}^2}\right) = \frac{C_{Xh}^2}{n_h\theta_{Xh}} = \nabla_1,$$

$$E(e_{0h}e_{1h}) = \frac{1}{n_h}\rho_h C_{Yh} C_{Xh} = \nabla_{01}.$$

where, $C_{Yh} = \frac{\sigma_{Yh}}{\mu_{Yh}}$, $C_{Xh} = \frac{\sigma_{Xh}}{\mu_{Xh}}$, $\theta_{Yh} = \frac{\sigma_{Uh}^2}{\sigma_{Uh}^2 + \sigma_{Yh}^2}$ and $\theta_{Xh} = \frac{\sigma_{Vh}^2}{\sigma_{Vh}^2 + \sigma_{Xh}^2}$.

EXISTING ESTIMATORS AND THEIR PROPERTIES

$\bar{y}_{st}$ is usual unbiased estimator in stratified random sampling scheme.

The usual combined ratio estimator in stratified random sampling in the presence of measurement error is defined as-

$$T_R = \bar{y}_{st} \frac{\mu_x}{\bar{x}_{st}} \tag{26}$$

The usual combined product estimator in the presence of measurement error is defined as-

$$T_{PR} = \bar{y}_{st} \frac{\bar{x}_{st}}{\mu_x} \tag{27}$$

Combined difference estimator in stratified random sampling is defined in the presence of measurement errors for a population mean, as

$$T_D = \bar{y}_{st} + d(\mu_x - \bar{x}_{st}) \tag{28}$$

The variance and mean square term of above estimators, up to the first order of approximation, are respectively given by

$$Var(\bar{y}_{st}) = \frac{C_{Xh}^2}{n_h}\left(1 + \frac{\sigma_{Uh}^2}{\sigma_{Yh}^2}\right) \tag{29}$$

$$MSE(T_R) = \sum_{h=1}^{L} \frac{W_h^2}{n_h}\left[\frac{\sigma_{Yh}^2}{\theta_{Yh}} + R\left(\frac{\sigma_{Xh}^2}{\theta_{Xh}}\right)\right]\left(R - 2\beta_{YXh}\theta_{Xh}\right) \tag{30}$$

$$MSE(T_P) = \sum_{h=1}^{L} \frac{W_h^2}{n_h}\left[\frac{\sigma_{Yh}^2}{\theta_{Yh}} + R\left(\frac{\sigma_{Xh}^2}{\theta_{Xh}}\right)\right]\left(R + 2\beta_{YXh}\theta_{Xh}\right) \tag{31}$$

$$MSE(T_D) = \sum_{h=1}^{L} \frac{W_h^2}{n_h}\left(\frac{\sigma_{Yh}^2}{\theta_{Yh}}\right) + d^2 \sum_{h=1}^{L} \frac{W_h^2}{n_h}\left(\frac{\sigma_{Xh}^2}{\theta_{Xh}}\right) - 2d\sum_{h=1}^{L} \frac{W_h^2}{n_h}\beta_{XYh}\sigma_{Xh}^2 \tag{32}$$

$$\text{where, } d_{opt} = \frac{\sum_{h=1}^{L} \frac{W_h^2}{n_h} \beta_{XYh} \sigma_{Xh}^2}{\sum_{h=1}^{L} \frac{W_h^2}{n_h} \left(\frac{\sigma_{Xh}^2}{\theta_{Xh}} \right)}$$

6. SUUGESTED ESTIMATOR AND ITS PROPERTIES

Let B(.) and M(.) denote the bias and mean square error (M.S.E) of an estimator under given sampling design. Estimator t_p defined in equation (19) can be written in stratified random sampling as

$$T_P = \left[\beta_1 \bar{y}_{st} + \beta_2 \bar{x}_{st}^* + (1 - \beta_1 - \beta_2)\mu_x^*\right]\left[\frac{\mu_x^*}{\bar{x}_{st}^*}\right]^{\beta} \quad (33)$$

where (α_1, α_2) are suitably chosen scalars such that MSE of proposed estimator is minimum, $\bar{x}_{st}^* (= \eta \bar{x}_{st} + \lambda)$, $\mu_x^* (= \eta \mu_x + \lambda)$ with (n, λ) are either constants or functions of some known population parameters. Here it is interesting to note that some existing estimators have been found particular members of proposed class of estimators T_p for different values of $(\alpha_1, \alpha_2, \alpha, \eta, \lambda)$, which are summarized in Table 4.

Table 4: Members of proposed class of estimators T_p

	Values of Constants				
Estimators	α_1	α_2	α	η	λ
$\bar{y}_{st}$ [Usual unbiased]	1	0	0	-	-
T_R [Usual ratio]	1	0	1	1	0
T_{PR} [Usual product]	1	0	-1	1	0
T_D [Usual difference]	1	α_2	0	-1	μ_x

Theorem 2.1: Estimator T_P in terms of $e_i; i = 0,1$ by ignoring the terms $E(e_{ih}^r e_{jh}^s)$ for (r+s)>2,where r,s=0,1,2... and $i = 0,1$; $j = 1$, can be written as

$$T_P = \left[\mu_x^* - \beta A' e_{1h}\mu_x^* + B'\mu_x^* e_{1h}^2 + \beta_1\left\{C' - \beta A' C e_{1h} + B'C' e_{1h}^2 + e_{0h}\mu_y - \beta A'\mu_y e_{0h} e_{1h}\right\}\right.$$
$$\left. + \beta_2 \eta\mu_x\left\{e_{1h} - \beta A' e_{1h}^2\right\}\right]$$

where, $A' = \dfrac{\eta\mu_x}{\eta\mu_x + \lambda}$, $B' = \dfrac{\beta(\beta+1)}{2} A'^2$ and $C' = \mu_y - \mu_x^*$.

Proof

$$T_P = \left[\beta_1 \bar{y}_{st} + \beta_2 \bar{x}_{st}^* + (1 - \beta_1 - \beta_2)\mu_x^*\right]\left[\frac{\mu_x^*}{\bar{x}_{st}^*}\right]^\beta$$

$$= \left[\beta_1(1 + e_{0h}) + \beta_2 \eta\, \mu_x e_{1h} + (1 - \beta_1)\mu_x^*\right]\left[1 + A' e_{1h}\right]^{-\beta} \quad (34)$$

We assume $|A' e_{1h}| < 1$, so that the term $(1 + A' e_{1h})^{-\beta}$ is expandable. Thus by expanding the right hand side (20) and neglecting the terms of e's having power greater than two, we have

$$T_P = \left[\mu_x^* - \beta A' e_{1h}\mu_x^* + B'\mu_x^* e_{1h}^2 + \beta_1\left\{C' - \beta A' C' e_{1h} + B'C' e_{1h}^2 + e_{0h}\mu_y - \beta A'\mu_y e_{0h} e_{1h}\right\}\right.$$
$$\left. + \beta_2 \eta\mu_x\left\{e_{1h} - \beta A' e_{1h}^2\right\}\right]$$

Theorem: 2.2 Bias of T_p is given by

$$B(T_P) = \left[B'\mu_x^* \nabla_1 + \beta_1\left\{B'C'\nabla_1 - \beta A'\mu_y \nabla_{01}\right\} - \beta_2 \eta\mu_x A' \beta \nabla_1\right] \quad (35)$$

Proof:

$$B(T_P) = E(T_P - \mu_y)$$

$$= E\left[\mu_x^* - \mu_y - \beta A' e_{1h}\mu_x^* + B'\mu_x^* e_{1h}^2 + \beta_1\left\{C' - \beta A' C' e_{1h} + B'C' e_{1h}^2 + e_{0h}\mu_y - \beta A'\mu_y e_{0h} e_{1h}\right\}\right.$$
$$\left. + \beta_2 \eta\mu_x\left\{e_{1h} - \beta A' e_{1h}^2\right\}\right]$$

$$= \left[B'\mu_x^* \nabla_1 + \beta_1 \left\{ B'C'\nabla_1 - \beta A'\mu_y \nabla_{01} \right\} - \beta_2 \beta \eta \mu_x A'\nabla_1 \right]$$

where, ∇_0, ∇_1 and ∇_{01} are already defined in section 3.

Theorem: 2.3 Mean square error of T_P, up to the first order of approximation is given by

$$MSE(T_P) = \beta_1^2 \left\{ C'^2 + \mu_y^2 \nabla_0 + \nabla_1 \left(\beta^2 A'^2 C'^2 + 2B'C'^2 \right) - 4\beta A'C'\mu_y \nabla_{01} \right\} + \beta_2^2 \eta^2 \mu_x^2 \nabla_1$$

$$+ \left\{ C'^2 + \nabla_1 \left(\beta^2 A'^2 \mu_x^2 - 2B'C'\mu_x^* \right) \right\} - 2\beta_1 \left\{ C'^2 + \nabla_1 \left(B'C'^2 - B'C'\mu_x^* - \beta^2 A'^2 C'\mu_x^* \right) + \nabla_{01} \beta A \mu_y \left(\mu_x^* - C' \right) \right\}$$
$$- 2\beta_2 \eta \mu_x \beta A \nabla_1 \left(\mu_x^* - C' \right) + 2\beta_1 \beta_2 \eta \mu_x \left(\mu_y \nabla_{01} - 2A'\beta C'\nabla_1 \right) \qquad \textbf{(36)}$$

Proof:

$$MSE(T_P) = E(T_P - \mu_y)^2$$

$$MSE(T_P) = \beta_1^2 \left\{ C'^2 + \mu_y^2 \nabla_0 + \nabla_1 \left(\beta^2 A'^2 C'^2 + 2B'C'^2 \right) - 4\beta A'C'\mu_y \nabla_{01} \right\} + \beta_2^2 \eta^2 \mu_x^2 \nabla_1$$

$$+ \left\{ C'^2 + \nabla_1 \left(\beta^2 A'^2 \mu_x^2 - 2B'C'\mu_x^* \right) \right\} - 2\beta_1 \left\{ C'^2 + \nabla_1 \left(B'C'^2 - B'C'\mu_x^* - \beta^2 A'^2 C'\mu_x^* \right) + \nabla_{01} \beta A \mu_y \left(\mu_x^* - C' \right) \right\}$$
$$- 2\beta_2 \eta \mu_x \beta A \nabla_1 \left(\mu_x^* - C' \right) + 2\beta_1 \beta_2 \eta \mu_x \left(\mu_y \nabla_{01} - 2A'\beta C'\nabla_1 \right)$$

$MSE(T_P)$ can also be written as

$$MSE(T_P) = \beta_1^2 \chi_1 + \beta_2^2 \chi_2 - 2\beta_1 \chi_3 - 2\beta_2 \chi_4 + 2\beta_1 \beta_2 \chi_5 + \chi \qquad (37)$$

Differentiating equation (37) with respect to (β_1, β_2) and equating it to zero, we get the optimum values of (β_1, β_2) respectively, as

$$\beta_{1(opt)} = \frac{\chi_2 \chi_3 - \chi_4 \chi_5}{\chi_1 \chi_2 - \chi_5^2} \text{ and } \beta_{2(opt)} = \frac{\chi_1 \chi_4 - \chi_3 \chi_5}{\chi_1 \chi_2 - \chi_5^2}$$

where, $\chi_1 = C'^2 + \mu_y^2 \nabla_0 + \nabla_1 \left(\beta^2 A'^2 C'^2 + 2B'C'^2 \right) - 4\beta A'C'\mu_y \nabla_{01}$

$$\chi_2 = \eta^2 \mu_x^2 \nabla_1$$

$$\chi_3 = C'^2 + \nabla_1 \left(BC^2 - B'C'\mu_x^* - \beta^2 A'^2 C'\mu_x^* \right) + \nabla_{01} \beta A'\mu_y \left(\mu_x^* - C' \right)$$

$$\chi_4 = \eta \mu_x \beta A'\nabla_1 \left(\mu_x^* - C' \right)$$

$$\chi_5 = \eta\mu_x\left(\mu_y\nabla_{01} - 2A'\beta C'\nabla_1\right).$$

$$\chi = C'^2 + \nabla_1\left(\beta^2 A'^2 \mu_x^2 - 2B'C'\mu_x^*\right)$$

With the help of these values, we get the minimum MSE of the suggested estimator T_p.

7. DISCUSSION AND CONCLUSION

In the present study, we have proposed difference-type class of estimators of the population mean of a study variable when information on an auxiliary variable is known in advance. The asymptotic bias and mean square error formulae of suggested class of estimators have been obtained. The asymptotic optimum estimator in the suggested class has been identified with its properties. We have also studied some traditional methods of estimation of population mean in the presence of measurement error such as usual unbiased, ratio, usual difference estimators suggested by Srivastava(1967), dubey and singh(2001), which are found to be particular members of suggested class of estimators. In addition, some new members of suggested class of estimators have also been generated in simple random sampling case. An empirical study is carried to throw light on the performance of suggested estimators over other existing estimators using simple random sampling scheme. From the Table 3, we observe that suggested estimator t_3 performs better than the other estimators considered in the present study and which reflects the usefulness of suggested method in practice.

REFERENCES

Allen, J., Singh, H. P. and Smarandache, F. (2003): A family of estimators of population mean using mutliauxiliary information in presence of measurement errors. International Journal of Social Economics 30 (7), 837–849.

A.K. Srivastava and Shalabh (2001). Effect of measurement errors on the regression method of estimation in survey sampling. **Journal of Statistical Research**, Vol. **35**, No. 2, pp. 35-44.

Bahl, S. and Tuteja, R. K. (1991): Ratio and product type exponential estimator. Information and optimization sciences12 (1), 159-163.

Chandhok, P.K., & Han, C.P.(1990):On the efficiency of the ratio estimator under Midzuno scheme with measurement errors. Journal of the Indian statistical Association,28,31-39.

Dubey, V. and Singh, S.K. (2001). An improved regression estimator for estimating population mean, J. Ind. Soc. Agri. Statist., 54, p. 179-183.

Gujarati, D. N. and Sangeetha (2007): Basic econometrics. Tata McGraw – Hill.

Koyuncu, N. and Kadilar, C. (2010): On the family of estimators of population mean in stratified sampling. Pakistan Journal of Statistics. Pak. J. Stat. 2010 vol 26 (2), 427-443.

Kumar, M. , Singh, R., Singh, A.K. and Smarandache, F. (2011 a): Some ratio type estimators under measurement errors. WASJ 14(2) :272-276.

Kumar, M., Singh, R., Sawan, N. and Chauhan, P. (2011b): Exponential ratio method of estimators in the presence of measurement errors. Int. J. Agricult. Stat. Sci. 7(2): 457-461.

Malik, S. and Singh, R. (2013) : An improved class of exponential ratio- type estimator in the presence of measurement errors. OCTOGON Mathematical Magazine, 21,1, 50-58.

Malik, S., Singh, J. and Singh, R. (2013) : A family of estimators for estimating the population mean in simple random sampling under measurement errors. JRSA, 2(1), 94-101.

Manisha and Singh, R. K. (2001): An estimation of population mean in the presence of measurement errors. Journal of Indian Society of Agricultural Statistics 54(1), 13–18.

Manisha and Singh, R. K. (2002): Role of regression estimator involving measurement errors. Brazilian journal of probability Statistics 16, 39- 46.

Shalabh (1997): Ratio method of estimation in the presence of measurement errors. Journal of Indian Society of Agricultural Statistics 50(2):150– 155.

Singh, H. P. and Karpe, N. (2008): Ratio-product estimator for population mean in presence of measurement errors. Journal of Applied Statistical Sciences 16, 49–64.

Singh, H. P. and Karpe, N. (2009): On the estimation of ratio and product of two populations means using supplementary information in presence of measurement errors. Department of Statistics, University of Bologna, 69(1), 27-47.

Singh, H. P. and Vishwakarma, G. K. (2005): Combined Ratio-Product Estimator of Finite Population Mean in Stratified Sampling. Metodologia de Encuestas 8: 35- 44.

Singh, H. P., Rathour, A., Solanki, R. S.(2013): An improvement over difference method of estimation of population mean. JRSS, 6(1):35-46.

Singh H. P., Rathour A., Solanki R.S. (2013): an improvement over difference method of estimation of population mean. JRSS, 6(1): 35-46.

Srivastava, S. K. (1967): An estimator using auxiliary information in the sample surveys. Calcutta statistical Association Bulletin 16,121-132.

Walsh, J.E.(1970). Generalisation of ratio estimate for population total. Sankhya A.32, 99-106.

The purpose of writing this book is to suggest some improved estimators using auxiliary information in sampling schemes like simple random sampling, systematic sampling and stratified random sampling.

This volume is a collection of five papers, written by nine co-authors (listed in the order of the papers): Rajesh Singh, Mukesh Kumar, Manoj Kr. Chaudhary, Cem Kadilar, Prayas Sharma, Florentin Smarandache, Anil Prajapati, Hemant Verma, and Viplav Kr. Singh.

In first paper dual to ratio-cum-product estimator is suggested and its properties are studied. In second paper an exponential ratio-product type estimator in stratified random sampling is proposed and its properties are studied under second order approximation. In third paper some estimators are proposed in two-phase sampling and their properties are studied in the presence of non-response.

In fourth chapter a family of median based estimator is proposed in simple random sampling. In fifth paper some difference type estimators are suggested in simple random sampling and stratified random sampling and their properties are studied in presence of measurement error.

ISBN 978-1-59973-275-6
53500>
9 781599 732756