STATISTICA, anno LXIX, n. 1, 2009

LINEAR COMBINATION OF ESTIMATORS IN PROBABILITY PROPORTIONAL TO SIZES SAMPLING TO ESTIMATE THE POPULATION MEAN AND ITS ROBUSTNESS TO OPTIMUM VALUE

S.K. Agarwal, M. Al Mannai

1. INTRODUCTION

To reduce the cost of the survey, most large scale sample surveys are conducted to estimate the parameters of several characteristics simultaneously. In multistage sampling the primary units are selected with probability proportional to size (pps) with or without replacement. It is well known that the selection of primaries with pps of an auxiliary variable, highly positively correlated with characteristics of interest, may improve the efficiency of the estimate of population mean/total. A vast literature (Agarwal *et al.*, 1978); (Chaudhuri and Vos, 1988); is available on this topic. (Rao and Bayless, 1969); carried out an empirical study for a wide variety of populations for pps estimators to develop the confidence of survey practitioners. For those characteristics of interest having low or very low correlation with size measure, (Rao, 1966a,b); introduced certain biased alternative estimators. (Singh, 1978); (Bansal, 1985); (Amahia *et al.*, 1989); (Bansal, 1990); and others have extended these alternative estimators by considering other situations which one may come across in real life. The variance expressions of the alternative estimators are very complicated and the direct comparison is not possible. Therefore, to see the relative efficiency for alternative biased estimators, an empirical study is conducted by (Agarwal and Kumar, 1998); and (Agarwal *et al.*, 2003); for a wide variety of populations.

In the last four decades several linear weighted estimators are suggested for estimating population mean/total [see (Singh, 1967); (Murthy, 1967); (Rubin and Weisberg, 1975); (Agarwal and Kumar, 1980); (Pandey and Singh, 1984); (Amahia *et al.*, 1989); (Bansal, 1990); (Agarwal *et al.*, 2003)]. However, a glance at the reports of various large scale sample surveys carried out by various organizations shows that sample survey practitioners have not shown much interest in these developments. This is a sort of paradox and there appears to be a gap in the theory developed on linear weighted estimators and its applications in actual surveys. One of the reasons appears to be that the weighted estimator involves unknown parameter values in the estimator, and the other reason might be non availability of computational facilities till the eighties of the last century. Therefore, the performance of these estimators over conventional estimators could not be studied for a wide variety of populations.

In this paper, an attempt has been made to show the importance of linear weighted estimators under ppswr sampling over the alternative estimator and also over the conventional estimators through an empirical study for a wide variety of populations available in literature. This study is more useful for primary unit populations. The linear weighted estimators involve unknown weights, and normally a guess value based on the past experience or an expert's opinion is used for unknown weights in the linear weighted estimators. The survey practitioners are not sure how close these guess values are from the optimum values. Therefore, an empirical study is also carried out to see the robustness of the linear weighted estimators by deviating the optimum value of the unknown weights up to 50% on either side. This study may help to develop the confidence of survey practitioners, on the use of linear weighted estimators.

2. STATEMENT OF THE PROBLEM

Consider a finite population U of size N identifiable, distinct units u_1, u_2, u_i, ..., u_N. It is assumed that several study variables y are defined on U. The information on auxiliary variables x_2 is also defined on U, besides the selection probabilities based on x_1. Some y's have high positive correlation with x_1, some y's have low positive correlation or have low negative correlation or have very low correlation with x_1. The problem is to estimate the population mean/total of those study variables y's which do not have high linear relationship with x_1. In such cases the conventional pps estimator doesn't estimate mean/total of y's with higher precision. Therefore, there is a need to use an additional auxiliary variable to estimate mean /total of these y's with higher precision. One of the approaches is by using linear weighted estimator. The selection probabilities are

$$p_{1i}\ (= x_{1i}/X_1;\ X_1 = \sum_{i=1}^{N} x_{1i}\ );\ i=1,2,\ ...,\ N.$$

There are three main situations using linear weighted estimators:

(i) The variables y and x_1 are uncorrelated or have very low correlation, and y and x_2 may have negative or positive correlation.

(ii) The variables y and x_1 are linearly, but not highly positively correlated, and y and x_2 have positive correlation.

(iii) The variables y and x_1 are linearly, but not having high correlation and y and x_2 have negative correlation.

For situation (i), Rao (1966) suggested alternative biased estimators. Since the direct comparison of the mean square error of Rao's estimator is not possible with the mean square error of the linear weighted estimator due to complicated expressions, hence we have made an empirical study for wide variety of populations to see the performance of linear weighted estimators over Rao's alternative biased estimator. Several other authors have also suggested alternative biased estimators, but we preferred Rao's biased estimator for comparison because it doesn't require any other additional information. Moreover Agarwal and Kumar (1998) when compared all seventeen biased alternative estimators, didn't find significant gain of other estimators over Rao (1966) estimator.

A linear weighted estimator under ppswr is suggested by Agarwal and Kumar (1980) for situation (ii). However, they have not carried out an extensive empirical study to demonstrate the usefulness of their estimator in real life surveys. In section 4, we have carried out an empirical study, to show its usefulness in practice, by deviating the unknown parameters up to 50% on either side from its optimum value.

To meet situation (iii), we extend product method of estimation to pps sampling in the next section.

3. LINEAR WEIGHTED ESTIMATOR WHEN STUDY VARIABLE y AND x_2 HAVE EITHER POSITIVE OR NEGATIVE LINEAR RELATIONSHIP

Let,

$$u_i = \frac{y_i}{Np_{1i}}\,;\; v_i = \frac{x_{2i}}{Np_{1i}}\,;$$

$$\bar{u} = n^{-1}\sum_{i=1}^{n} u_i = \bar{y}_{pps}\,;\; \bar{v} = n^{-1}\sum_{i=1}^{n} v_i = \bar{x}_{2pps}$$

$$nV(\bar{y}_{pps}) = \sigma_u^2 = \sum_{i=1}^{N} p_{1i}(u_i - \bar{Y})^2 \tag{1}$$

$$\sigma_v^2 = \sum_{i=1}^{N} p_{1i}(v_i - \bar{X}_2)^2 \tag{2}$$

$$C_u = \frac{\sigma_u}{\bar{Y}}\,;\; C_v = \frac{\sigma_v}{\bar{X}_2} \tag{3}$$

$$\bar{y}_R = \frac{\bar{u}}{\bar{v}}\bar{X}_2\,;\; \bar{y}_p = \frac{\overline{uv}}{\bar{X}_2}\,;$$

$$\rho_{uv} = \sum_{i=1}^{N} \frac{p_{1i}(u_i - \bar{Y})(v_i - \bar{X}_2)}{\sigma_u \sigma_v} \tag{4}$$

$$M(\bar{y}_R) \cong n^{-1}\bar{Y}^2(C_u^2 + C_v^2 - 2\rho_{uv}C_uC_v) \tag{5}$$

$$M(\overline{y}_p) \cong n^{-1}\overline{Y}^2(C_u^2 + C_v^2 + 2\rho_{uv}C_uC_v) \tag{6}$$

$$Cov(\overline{y}_R, \overline{y}_{pps}) \cong n^{-1}\overline{Y}^2(C_u^2 - \rho_{uv}C_uC_v) \tag{7}$$

$$Cov(\overline{y}_p, \overline{y}_{pps}) \cong n^{-1}\overline{Y}^2(C_u^2 + \rho_{uv}C_uC_v) \tag{8}$$

The proposed estimator $\overline{y}_1$ of the population mean $\overline{Y}$ is a linear weighted estimator, when the study variable y and x_2 are linearly but negatively correlated:

$$\overline{y}_1 = w\overline{y}_p + (1-w)\overline{y}_{pps} \tag{9}$$

where w is the weight to be determined. The bias and mean squared error of $\overline{y}_1$ are respectively

$$B(\overline{y}_1) = wB(\overline{y}_p) \tag{10}$$

and

$$M(\overline{y}_1) = w^2M(\overline{y}_p) + (1-w)^2V(\overline{y}_{pps}) + 2w(1-w)Cov(\overline{y}_p, \overline{y}_{pps}) \tag{11}$$

The value of w which minimizes eqn (11) is

$$w_{opt} = \frac{V(\overline{y}_{pps}) - Cov(\overline{y}_p, \overline{y}_{pps})}{M(\overline{y}_p) + V(\overline{y}_{pps}) - 2Cov(\overline{y}_p, \overline{y}_{pps})} \tag{12}$$

On substitution of w_{opt} from (12) into eqn (11) we get

$$M(\overline{y}_1)_{\min} = \frac{V(\overline{y}_{pps})M(\overline{y}_p) - Cov^2(\overline{y}_p, \overline{y}_{pps})}{M(\overline{y}_p) + V(\overline{y}_{pps}) - 2Cov(\overline{y}_p, \overline{y}_{pps})} \tag{13}$$

Substituting the values from eqns (1), (3), (6) and (8), into eqn (10), eqn (12) and eqn (13), we get w_{opt}, the bias and the mean square error of $\overline{y}_1$, to the first degree of approximations, as follows:

$$w_{opt} = -\left(\rho_{uv}\frac{C_u}{C_v}\right) \tag{14}$$

$$B(\overline{y}_1) = -(n^{-1}\rho_{uv}^2\sigma_uC_u) \tag{15}$$

and

$$M(\overline{y}_1) = n^{-1}(1-\rho_{uv}^2)\sigma_u^2 \tag{16}$$

It can be noted that $M(\bar{y}_1)$ in equation (16) is of the same form as the mean square error of the regression estimator under ppswr sampling. Thus, it can be estimated approximately the same way as we do in the regression method of estimation under ppswr sampling.

When the study variables y and x_2 are linearly and low positively correlated (Agarwal and Kumar, 1980); defined the linear weighted estimator as follows:

$$\bar{y}_0 = k\bar{y}_{\mathrm{R}} + (1-k)\bar{y}_{pps} \tag{17}$$

The value of k which minimizes $M(\bar{y}_0)$ is

$$k_{opt} = \rho_{uv}\frac{C_u}{C_v}$$

The bias and mean square error of $\bar{y}_0$ for k_{opt}, to the first degree of approximations are:

$$B(\bar{y}_0) = n^{-1}\rho_{uv}\sigma_u(C_v - \rho_{uv}C_u) \tag{18}$$

and

$$M(\bar{y}_0) = n^{-1}(1-\rho_{uv}^2)\sigma_u^2 \tag{19}$$

Rao (1966) defined alternative estimators for situation (i) by modifying the conventional ppswr estimator of the population mean by replacing y_i' by $Ny_i p_i$ in the following equations.

$$\hat{\bar{Y}}' = \frac{1}{n\mathrm{N}}\sum_{i=1}^{n}\frac{y_i'}{p_i} \tag{20}$$

with

$$V(\hat{\bar{Y}}') = \frac{1}{n\mathrm{N}^2}\left(\sum_{i=1}^{n}\frac{y_i'^2}{p_i} - \bar{Y}'^2\right) \tag{21}$$

where y_i' is a prime characteristic under study and y_i' and p_i are high positively correlated. The estimator defined by equation (20) is unbiased while Rao's (1966) following alternative estimator is biased.

$$\bar{y}_2 = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{22}$$

The bias of the estimator is

$$B(\bar{y}_2) = \sum_{i=1}^{n} y_i p_i - \bar{Y} \tag{23}$$

$$M(\bar{y}_2) = \frac{1}{n}\left(\sum_{i=1}^{N} p_i y_i^2 - \bar{Y}^2\right) \tag{24}$$

In the next section we will study the respective percentage relative bias to its standard error for Rao's alternative estimator and that of linear weighted estimators. We will also study the gain of efficiency of linear weighted estimators over Rao's alternative estimator.

4. GAIN IN EFFICIENCY

Let the guess value of weights "w" in eqn (9) or "k" in eqn (17) be d and the optimum value of the weight is d_{opt}. The percentage gain in efficiency of $\bar{y}_0$ or $\bar{y}_1$ over $\bar{y}_{pps}$ is

$$\left(\frac{d_{opt}^2 C_v^2}{C_u^2 - d_{opt}^2 C_v^2}\right)100 \tag{25}$$

The estimators $\bar{y}_0$ or $\bar{y}_1$ will be more efficient than $\bar{y}_{pps}$ if $d < 2\left|d_{opt}\right|$. It means that even if the value of d departs from d_{opt} by 100%, the estimators $\bar{y}_0$ or $\bar{y}_1$ will remain more efficient than $\bar{y}_{pps}$.

The percentage gain in efficiency of $\bar{y}_0$ over $\bar{y}_R$ or $\bar{y}_1$ over $\bar{y}_p$ is

$$\left[\frac{C_v^2 (1 - d_{opt})^2}{C_u^2 - d_{opt}^2 C_v^2}\right]100 \tag{26}$$

The estimators $\bar{y}_0$ will be more efficient than $\bar{y}_R$, if the condition

$$\left|d - d_{opt}\right| < \left|1 - d_{opt}\right| \tag{27}$$

holds.

5. EMPIRICAL STUDY

To study the relative efficiency and the relative bias of linear weighted estimators over conventional estimator/s under probability proportional to size with replacement (ppswr) sampling we have considered a wide variety of populations that cover most of the practical situations we come across in real life surveys.

These populations are taken from the available literature ((Freund and Perles, 1999); (Hines and Montgomery, 1990); (Mendenhall *et al.*, 2003); (Milton and Arnold, 2003); (Neter *et al.*, 1985) and (Ott, 1984)).

5.1. *Description of populations*

Table 1 gives the characteristics of the populations such as population size N, coefficients of variation of the study variable (y), and of the auxiliary variables x_1 and x_2, the correlation coefficients between (y, x_1) and (y, x_2). The population size varies from 14 to 113, the coefficient of variation of y from 13.44 % to 49.67%, the coefficient of variation of x_1 from 17.87% to 80.39%, the coefficient of variation of x_2 from 2.41% to 75.75%. The correlation coefficient between (y, x_1) varies from -0.002 to 0.587, while the correlation coefficient between (y, x_2) varies from -0.833 to 0.988. The above described populations thus represent a variety of situations and we further divide these into three categories for comparison purpose. Category (i) represents those populations for which the study variable y and x_1 either uncorrelated or have very low correlation, and y and x_2 have negative or positive correlation with y. Category (ii) represents those populations for which the study variables y and x_1 are linearly, but not highly positively correlated, and y and x_2 have positive correlation. Category (iii) represents those populations for which the study variables y and x_1 not having high correlation and y and x_2 have negative correlation.

TABLE 1

Description of the populations with five number summary

S.No	N	C_y	C_{x1}	C_{x2}	ρ_{yx1}	$\lvert\rho_{yx2}\rvert$	Situation
1	113	19.81	80.39	8.38	0.34	0.19	ii
2	16	14.01	34.43	32.99	0.41	0.89	ii
3	50	13.44	31.27	39.17	0.59	0.29	ii
4	107	19.81	64.81	19.42	0.33	0.30	iii
5	63	19.51	33.46	41.65	0.52	0.48	iii
6	22	14.19	55.96	2.41	0.00	0.03	i
7	97	18.09	63.10	19.04	0.35	0.22	iii
8	43	13.67	51.11	30.38	0.45	0.01	iii
9	113	19.81	30.79	19.42	0.53	0.30	iii
10	25	21.07	19.64	16.20	0.51	0.50	ii
11	26	49.67	33.88	28.32	0.16	0.83	i
12	33	36.52	17.87	31.73	0.35	0.67	iii
13	48	27.12	27.80	73.69	0.10	0.37	i
14	54	26.92	39.00	71.69	0.50	0.37	ii
15	18	31.09	42.76	75.75	0.25	0.99	i
16	14	27.02	19.78	41.24	0.45	0.88	ii
17	54	27.72	39.00	73.69	0.50	0.37	ii
18	52	27.56	39.65	75.52	0.51	0.36	ii
19	47	27.45	72.65	12.32	0.36	0.32	ii
20	27	49.67	33.88	27.42	0.16	0.78	i
21	22	48.56	27.31	25.86	0.76	0.80	iii
22	20	48.64	24.56	35.03	0.74	0.18	ii
23	18	31.09	42.76	75.75	0.25	0.99	i
24	19	6.64	28.87	6.56	0.69	0.24	iii
25	17	21.84	22.60	20.99	0.62	0.35	ii
26	22	21.31	8.28	10.84	0.88	0.89	ii
min	**14**	**6.64**	**8.28**	**2.41**	**0.00**	**0.01**	
max	**113**	**49.67**	**80.39**	**75.75**	**0.88**	**0.99**	
q1	**24**	**17.12**	**30.04**	**19.32**	**0.31**	**0.28**	
q2	**46**	**19.81**	**34.15**	**31.06**	**0.38**	**0.37**	
q3	**72**	**27.04**	**52.32**	**41.34**	**0.51**	**0.71**	

Table 2 gives the bias relative to the standard error for the estimators $\overline{y}_2$, $\overline{y}_0$ and $\overline{y}_1$ for a sample of size 10% of the population. The relative bias of $\overline{y}_2$ over $\overline{y}_0$ and $\overline{y}_1$ along with five number summary is also given. It can be noted that the bias of $\overline{y}_0$ and $\overline{y}_1$ relative to their respective standard errors are more than that of Rao's estimator $\overline{y}_2$. The reason appears to be that the standard error decreases considerably for the estimators $\overline{y}_0$ and $\overline{y}_1$. When we look into the relative bias of $\overline{y}_2$ over $\overline{y}_0$ and $\overline{y}_1$, we find that for approximately 70% of the populations, the bias of Rao's estimator is more than that of the bias of $\overline{y}_0$ while for 50% of the populations, the bias of Rao's estimator is more than that of the bias of $\overline{y}_1$. The range of relative bias of $\overline{y}_2$ over $\overline{y}_0$ is more than 20 while the relative bias of $\overline{y}_2$ over $\overline{y}_1$ is more than 137. It indicates that there are significant fluctuations in relative bias of Rao's estimator over $\overline{y}_0$ or $\overline{y}_1$.

TABLE 2

Bias relative to standard error and the bias relative to the bias of Rao's estimator with five number summary

S.No.	$\frac{B(\overline{y}_2)}{S.E(\overline{y}_2)}$	$\frac{B(\overline{y}_0)}{S.E(\overline{y}_0)}$	$\frac{B(\overline{y}_1)}{S.E(\overline{y}_1)}$	$\frac{B(\overline{y}_2)}{B(\overline{y}_0)}$	$\frac{B(\overline{y}_2)}{B(\overline{y}_1)}$	Situation
1	0.1976	1.1623	1.5909	4.3360	0.8530	ii
2	0.0827	2.1426	0.0889	0.5083	0.3039	ii
3	0.0623	2.4959	0.0061	2.2463	4.7161	ii
4	0.1148	1.1072	0.4930	3.5006	0.9786	iii
5	0.0616	0.8073	0.0285	3.0431	9.1014	iii
6	0.0009	0.2791	3.6211	0.0125	0.0012	i
7	0.1158	1.0513	0.4944	3.3379	0.9641	iii
8	0.0868	3.7835	0.0337	1.0715	0.5905	iii
9	0.1039	0.7264	0.1353	6.6641	6.5586	iii
10	0.1056	1.1638	0.0071	3.4766	3.3751	ii
11	0.0373	1.6869	0.0000	3.9975	39.8656	i
12	0.0347	3.1574	0.0065	0.6175	1.6993	iii
13	0.0102	0.5490	0.0112	0.6035	2.6886	i
14	0.0851	0.3701	0.0029	12.1542	137.4944	ii
15	0.0571	4.4281	0.1708	0.1468	0.0891	i
16	0.1036	0.6421	0.5696	0.7070	0.5773	ii
17	0.0851	0.3701	0.0029	12.1542	137.4944	ii
18	0.0847	0.4110	0.0041	9.6165	93.9787	ii
19	0.2210	0.4541	0.6364	9.2680	0.9191	ii
20	0.0840	1.6327	0.0138	4.4314	0.3194	i
21	0.1376	5.3049	0.0104	1.2436	3.3332	iii
22	0.1459	2.5332	0.0004	4.3132	25.7247	ii
23	0.0571	4.4281	0.1708	0.1468	0.0891	i
24	0.1251	0.5447	0.2277	3.9045	0.2216	iii
25	0.1170	1.3267	0.0060	2.3331	4.0920	ii
26	0.1195	0.3698	0.0041	20.3573	6.2104	ii
min	**0.0009**	**0.2791**	**0.0000**	**0.0125**	**0.0012**	
max	**0.2210**	**5.3049**	**3.6211**	**20.3573**	**137.4944**	
q1	**0.0521**	**0.7053**	**0.0069**	**0.6140**	**0.5872**	
q2	**0.0839**	**1.1348**	**0.0613**	**2.6447**	**1.3389**	
q3	**0.1043**	**2.2309**	**0.4934**	**3.6249**	**5.1767**	

Table 3 gives the relative efficiency of linear weighted estimator $\overline{y}_0$ or $\overline{y}_1$ over $\overline{y}_{pps}$ and also the loss in gain of efficiency when unknown weight in the estimator is deviated from its optimum value from ± 10% to ± 50%, along with the five number summary. Normally, the guess value of the weight is obtained either from the past surveys or an opinion of the expert is taken. It can be noted that for all the populations the relative efficiency of linear weighted estimator $\overline{y}_0$ or $\overline{y}_1$ over $\overline{y}_{pps}$ is significantly high even if the unknown value of weight in the estimators is deviated by up to 50%. From the summary statistics, it is evident that for more than 50% populations the gain is more than three times even if the optimum value is deviated by up to 25%. For 50% of the populations the gain in efficiency is more than 2 times even if the optimum value is deviated up to 50%.

TABLE 3

Relative efficiency of linear weighted estimator over $\overline{y}_{pps}$ when unknown parameter is deviated from its optimum value, with five number summary

S.No.	$\frac{V(\overline{y}_{pps})}{M(\overline{y}_0)}$	$\frac{V(\overline{y}_{pps})}{M(\overline{y}_0)}$ ±10%	$\frac{V(\overline{y}_{pps})}{M(\overline{y}_0)}$ ±25%	$\frac{V(\overline{y}_{pps})}{M(\overline{y}_0)}$ ±50%	Situation
1	16.349	14.174	8.344	3.380	ii
2	8.270	7.709	5.686	2.935	ii
3	1.717	1.705	1.643	1.456	ii
4	5.757	5.495	4.437	2.630	iii
5	1.270	1.267	1.249	1.190	iii
6	13.249	11.803	7.504	3.262	i
7	5.728	5.469	4.421	2.625	iii
8	4.937	4.750	3.962	2.488	iii
9	2.305	2.276	2.131	1.738	iii
10	1.654	1.643	1.589	1.421	ii
11	1.005	1.005	1.005	1.004	i
12	1.557	1.548	1.504	1.366	iii
13	1.017	1.017	1.016	1.013	i
14	9.545	8.794	6.222	3.043	ii
15	6.329	6.009	4.748	2.714	i
16	1.084	1.084	1.079	1.062	ii
17	1.017	2.020	1.916	1.619	ii
18	1.024	1.023	1.022	1.018	ii
19	4.715	4.546	3.826	2.445	ii
20	3.347	3.271	2.919	2.109	i
21	2.699	2.654	2.440	1.894	iii
22	1.063	1.062	1.059	1.047	ii
23	9.545	8.794	6.222	3.043	i
24	11.877	10.712	7.071	3.193	iii
25	1.501	1.493	1.455	1.334	ii
26	1.351	1.346	1.322	1.242	ii
min	**1.005**	**1.005**	**1.005**	**1.004**	
max	**13.249**	**11.803**	**7.504**	**3.262**	
q1	**1.224**	**1.478**	**1.440**	**1.322**	
q2	**2.011**	**2.148**	**2.024**	**1.679**	
q3	**5.900**	**5.624**	**4.515**	**2.651**	

Table 4 gives the relative efficiency of the linear weighted estimator $\overline{y}_0$ or $\overline{y}_1$ over $\overline{y}_R$ and also the loss in gain of efficiency when unknown weight is deviated from its optimum value from ± 10% to ± 50%, along with the five number summary. From the summary statistics, it can be seen that, for approximately 75% populations the gain is more or less 1.5 times even if the optimum value is deviated by up to 25%. For a few populations such as serial numbers 6, 19, 20, 24 and 26, the gain is not significant. One of the reasons might be that the big difference between the values of the coefficients of variation of the study variable (y) and the auxiliary variable (x_1).

TABLE 4

Relative efficiency of linear weighted estimator over $\bar{y}_R$ when unknown parameter is deviated from its optimum value, with five number summary

S.No.	$\frac{V(\bar{y}_R)}{M(\bar{y}_0)}$	$\frac{V(\bar{y}_R)}{M(\bar{y}_0)}$ ±10%	$\frac{V(\bar{y}_R)}{M(\bar{y}_0)}$ ±25%	$\frac{V(\bar{y}_R)}{M(\bar{y}_0)}$ ±50%	Situation
1	1.594	1.382	0.814	0.330	ii
2	3.599	3.355	2.475	1.277	ii
3	4.160	4.130	3.982	3.528	ii
4	1.372	1.309	1.057	0.627	iii
5	3.417	3.408	3.360	3.201	iii
6	1.121	0.999	0.635	0.276	i
7	1.394	1.331	1.076	0.639	iii
8	2.196	2.113	1.762	1.107	iii
9	2.264	2.235	2.094	1.707	iii
10	1.616	1.605	1.553	1.389	ii
11	1.541	1.541	1.540	1.539	i
12	5.215	5.186	5.039	4.578	iii
13	3.145	3.145	3.142	3.132	i
14	4.152	3.826	2.707	1.324	ii
15	4.553	4.322	3.415	1.952	i
16	4.751	4.747	4.726	4.652	ii
17	3.145	3.145	3.142	3.132	ii
18	3.253	3.252	3.248	3.234	ii
19	1.037	0.999	0.841	0.537	ii
20	1.012	0.989	0.883	0.638	i
21	13.206	12.986	11.938	9.269	iii
22	3.243	3.241	3.230	3.192	ii
23	4.152	3.826	2.707	1.324	i
24	1.035	0.933	0.616	0.278	iii
25	2.541	2.528	2.464	2.258	ii
26	1.033	1.029	1.010	0.949	ii
min	**1.012**	**0.933**	**0.616**	**0.276**	
max	**13.206**	**12.986**	**11.938**	**9.269**	
q1	**1.581**	**1.501**	**1.424**	**0.990**	
q2	**2.705**	**2.690**	**2.284**	**1.464**	
q3	**4.154**	**3.902**	**3.374**	**3.149**	

Table 5 gives the relative efficiency of linear weighted estimator $\bar{y}_0$ or $\bar{y}_1$ over Rao's estimator $\bar{y}_2$ and also the loss in gain of efficiency when the unknown weight is deviated from its optimum value from ± 10% to ± 50%, along with the five number summary. From the summary statistics, it can be seen that for approximately 50% of the populations the gain is more than1.5 times even if the optimum value is deviated by up to 25%. While the gain is more than 1.3 times for more than 50% populations even if the optimum value is deviated up to 50%.

TABLE 5

Relative efficiency of linear weighted estimator over Rao's estimator when unknown parameter is deviated from its optimum value, with five number summary

S.No.	$\frac{V(\bar{y}_2)}{M(\bar{y}_0)}$	$\frac{V(\bar{y}_2)}{M(\bar{y}_0)}$ ±10%	$\frac{V(\bar{y}_2)}{M(\bar{y}_0)}$ ±25%	$\frac{V(\bar{y}_2)}{M(\bar{y}_0)}$ ±50%	Situation
1	3.304	2.864	1.686	0.683	ii
2	4.098	3.820	2.818	1.455	ii
3	1.865	1.852	1.785	1.581	ii
4	1.348	1.287	1.039	0.616	iii
5	1.314	1.311	1.292	1.231	iii
6	1.234	1.100	0.699	0.304	i
7	1.374	1.312	1.060	0.630	iii
8	1.354	1.303	1.086	0.682	iii
9	2.627	2.593	2.429	1.981	iii
10	3.706	3.682	3.561	3.186	ii
11	1.084	1.084	1.083	1.082	i
12	2.236	2.224	2.161	1.963	iii
13	1.589	1.588	1.587	1.582	i
14	2.058	1.896	1.342	0.656	ii
15	11.175	10.610	8.383	4.792	i
16	1.525	1.524	1.517	1.494	ii
17	1.589	1.588	1.587	1.582	ii
18	1.574	1.573	1.571	1.564	ii
19	2.538	2.447	2.060	1.316	ii
20	3.608	3.526	3.147	2.274	i
21	2.266	11.261	10.353	8.038	iii
22	3.880	3.878	3.865	3.820	ii
23	2.058	1.896	1.342	0.656	i
24	2.033	1.833	1.210	0.547	iii
25	4.357	4.335	4.224	3.872	ii
26	4.835	4.819	4.732	4.446	ii
min	**1.084**	**1.084**	**0.699**	**0.304**	
max	**11.175**	**11.261**	**10.353**	**8.038**	
q1	**1.352**	**1.309**	**1.086**	**0.676**	
q2	**1.727**	**1.720**	**1.552**	**1.343**	
q3	**2.796**	**2.661**	**2.228**	**1.677**	

6. CONCLUDING REMARKS

The weighted linear estimator involves unknown parameter values in the estimator, hence the practitioner have some reservations of using these estimators with a guess value of the weight which is not close to the optimum value. The present study, for a wide variety of populations which we normally come across in real life situations, shows that the relative efficiency and the relative bias of linear weighted estimators over conventional estimator/s, are quite satisfactory even if the guess value of the weights in the weighted linear estimator departs by 50% from the optimum value. Therefore, our recommendations are that the proposed estimators are reasonably satisfactory from bias and efficiency point of view to estimate the population mean or total. To reduce the sampling error, the survey practitioners can use weighted linear estimators in the multi purpose surveys.

ACKNOWLEDGMENTS

The authors are thankful to the referee for his constructive and valuable comments, which helped to improve the original manuscript.

Department of Mathematics SATISH KUMAR AGARWAL
College of Science MARIAM AL MANNAI
University of Bahrain
Kingdom of Bahrain

REFERENCES

S.K. AGARWAL, R.S. KUSHWAHA, B.B.P.S. GOEL (1978), *On unequal probability sampling WOR: Theory and Practice*, "National Academy of Sciences", Sp Vol, pp. 251-266.

S.K. AGARWAL, P. KUMAR (1980), *Combination of Ratio and PPS estimators*, "Journal of Indian Society of Agricultural Statistics", 32, pp. 81-86.

S.K. AGARWAL, P. KUMAR (1998), *On relative efficiency of estimators of population total in PPS sampling when study variable has weak relationship with size variable.* "Computational Statistics and Data Analysis", 28, pp. 271-281.

S.K. AGARWAL, J.A. AL-SALEH, D.K. AL-MUTAIRI (2003), *An estimator of population total in multiple characteristics and its robustness to product moment correlation coefficient*, "Journal of Statistical Computation and Simulation.", 73, pp. 377-386.

G.N. AMAHIA, Y.P. CHAUBEY, T.J. RAO (1989), *Efficiency of a new estimator in PPS sampling for multiple characteristics*, "Journal of Statistical Planning and Inference", 21, pp. 75-84.

M.L. BANSAL, R. SINGH (1985), *An alternative estimator for multiple characteristics in PPS Sampling,* "Journal of Statistical Planning andInference", 11, pp. 313-320.

M.L. BANSAL (1990), *An alternative estimator for multiple characteristics in RHC Scheme,* "Communications in Statistics – Theory and Methods", 29, pp. 1777-1784.

A. CHAUDHURI, J.W.E. VOS (1988), *Unified Theory and Strategies of Survey Sampling.* Elsevier Science Publishers, North Holland.

J.E. FREUND, B.M. PERLES (1999), *Statistics: A first course,* 7th Ed. Prentice Hall, USA.

W.W. HINES, D.C. MONTGOMERY (1990), *Probability and Statistics in Engineering and Management Sciences, 3rd Ed.*, Prentice Hall.

P. KUMAR, S.K. AGARWAL (1997), *Alternative estimators for the population totals in Multiple characteristics surveys.* "Communications in Statistics – Theory and Methods", 26, pp. 2527-2537.

W. MENDENHALL, W. SINCICH, L. TERRY (2003), *A second course in statistics: Regression Analysis, 6th Ed.* Prentice Hall, New Jersey.

J.S. MILTON, J.C. ARNOLD (2003), *Introduction to probability and Statistics for engineers and computing Sciences*, McGraw Hill, New York.

M.N. MURTHY (1967), *Sampling Theory and Methods.* Statistical Publication Society, Calcutta, India.

J. NETER, W. WASSERMANN, M.H. KUTNER (1985), *Applied Linear Statistical Models, 2nd Ed.* Richard D. Irwin Inc., Illinois.

L. OTT (1984), *An introduction to Statistical Methods and Data Analysis, 2nd Ed.* Duxbury, Boston.

S. K. PANDEY, R. KARAN SINGH (1984), *On Combination of Ratio and PPS estimators*, "Biometrical Journal", 26-3, pp. 333-336.

J.N.K. RAO (1966a), *Alternative estimators in PPS sampling for multiple characteristics*, "Sankhya", A 28, pp. 47-60.

J.N.K. RAO (1966b), *On the relative efficiency of some estimators in PPS sampling for multiple characteristics,* "Sankhya", A 28, pp. 61-70.

J.N.K. RAO, D.L. BAYLESS (1969), *An empirical study of stabilities of estimators and variance estimators in unequal probability sampling.* "Journal of American Statistical Association", 64, pp. 540-549.

D.B. RUBIN, S. WEISBERG (1975), *The variance of a linear combination of independent estimators using estimated weights*, "Biometrika", 62, pp. 708 -709.

M.P. SINGH (1978), *Alternative estimators in PPS sampling,* "Survey Methodology", 4, pp. 263-271.

M.P. SINGH (1967), *Ratio cum product method of estimation.* "Metrika", 12, pp. 21-28.

SUMMARY

Linear combination of estimators in probability proportional to sizes sampling to estimate the population mean and its robustness to optimum value

In this paper we have studied the gain of efficiency and the relative bias of linear weighted estimators over conventional estimators under probability proportional to size with replacement (ppswr) sampling for a wide variety of populations. The five number summary statistics for the relative bias and the relative efficiency over conventional estimators is given for different magnitude of correlation coefficients. The computational study shows that there is a considerable gain in the efficiency of linear weighted estimators over conventional estimators. To develop the confidence of survey practitioners on linear weighted estimator, the computational study is extended to see the robustness of the linear weighted estimator by deviating the optimum value of the weight up to 50% on either side.