



The Efficiency of Cluster Sampling with Ratio Estimators Using Single and Two Auxiliary Variables Over simple Random Sampling

Mariam Al-Mannai

Department of Mathematics, College of Science, University of Bahrain, Kingdom of Bahrain

Received September 3, 2014; Revised October 27, 2014; Accepted October 31, 2014, Published 1 Nov. 2014

Abstract: In this paper we have studied empirically the role of single or two auxiliary variables under which cluster sampling in conjunction with ratio method of estimation is more efficient (or less efficient) than using conventional estimator in simple random sampling (SRS), if intra class correlation of the elements within the cluster has positive correlation. We have considered some extreme populations which may help the survey practitioners to know the magnitude of gain or loss in efficiency of cluster sampling using single or two auxiliary variables in conjunction with ratio method of estimation. To have more insight, the clusters are formed for sample sizes 2, 3 and 4.

Keywords: Auxiliary variables; Cluster sampling; Intra class correlation; Ratio method of estimation; Simple Random Sampling

1. Introduction

In large scale sample surveys cluster sampling is the most preferable sampling because it not only reduces the cost of the surveys, but also the researcher can have a larger sample size than if he or she was using simple random sampling (SRS). Besides, it may be used when it is either difficult or costly or impractical to prepare sampling frame of the units, to draw a sample. Moreover, sometimes the population elements are naturally grouped into subpopulations and lists of those subpopulations either already exist or easy and cheap to obtain. For a fixed budget, the main motivation to use cluster sampling, in large scale surveys, is to reduce the cost of data collection, data collection is faster and it increases sample size which helps to reduce the sampling error [see [4],[5], [9], [10] and [12]].

One of the main disadvantage of cluster sampling is that its efficiency is less than SRS for given sample size because the intra class correlation of the elements within the cluster has positive correlation. Zarkovic [13] illustrated that correlation between two characteristics of cluster means is directly proportional to the size of the clusters. Mishro and Sukhatme [8] utilized this information and gave the conditions under which cluster sampling in conjunction with ratio or regression method is more efficient than SRS even if intra class correlation

of the elements within the cluster has positive correlation. Due to lack of computational facilities till early seventies, the conditions were not tested for large or actual populations.

In this paper we have tested the conditions, for a wide variety of known populations, and studied whether the cluster sampling in conjunction with ratio estimator is more efficient than simple random sampling (SRS) or not if intra class correlation of the elements within the cluster has positive correlation. The study will help the survey practitioners to know the magnitude of gain or loss in efficiency of cluster sampling over SRS in conjunction with ratio or regression method. To have more insight, the clusters are formed for sample sizes 2, 3 and 4. The study is also extended to see how much gain (or loss) can be obtained in cluster sampling in conjunction with ratio estimator using two auxiliary variables.

2. Notations and Preliminaries

Consider a finite population U of size N identifiable, distinct units $u_1, u_2, u_i, \dots, u_N$. It is assumed that study variables y , and auxiliary variables x_j , $j = 1, 2$ are defined on U . A simple random sample without replacement of size n is selected from finite population of size N . Let



$$\bar{y} = n^{-1} \sum_{i=1}^n y_i ; \bar{Y}_j = N^{-1} \sum_{i=1}^N y_i ; V(y) = \sigma_y^2 = N^{-1} \sum_{i=1}^N (y_i - \bar{Y})^2 ; j = 1, 2.$$

$$\bar{x}_j = n^{-1} \sum_{i=1}^n x_i ; \bar{X}_j = N^{-1} \sum_{i=1}^N x_i ; V(x_j) = \sigma_{x_j}^2 = N^{-1} \sum_{i=1}^N (x_{ji} - \bar{X}_j)^2 ; j=1, 2.$$

$$\bar{y}_{R_j} = \frac{\bar{y}}{\bar{x}_j} \bar{X}_j ; Cov(y, x_j) = \sum_{j=1}^N \sum_{i=1}^N (y_i - \bar{Y})(x_{ji} - \bar{X}_j) ; j = 1, 2$$

$$\rho_j = \frac{Cov(y, x_j)}{\sqrt{V(y)V(x_j)}} ; j = 1, 2. ; \rho_{12} = \frac{Cov(x_k, x_l)}{\sqrt{V(x_k)V(x_l)}} ; k \neq l$$

Where

ρ_j is the correlation coefficient between y and x_j .

ρ_{12} is the correlation coefficient between x_1 and x_2 .

Let M be the size of the N_1 clusters; $N=N_1M$.

$$\bar{y}_{cl} = \frac{1}{mn_1} \sum_{t=1}^{n_1} \sum_{s=1}^m y_{ts}$$

$$\hat{V}(\bar{y}_{cl}) = \left(\frac{N_1 - n_1}{N_1} \right) \left(\frac{1}{mn_1^2} \right) \left(\frac{1}{n_1 - 1} \right) \sum_{t=1}^{n_1} (y_t - m\bar{y}_{cl})^2$$

ρ_{bj} is the correlation coefficient between cluster means of the characteristics under study

(y) and the auxiliary characteristics (x_j) for clusters of size M .

ρ'_{bj} is the multiple correlation coefficient between cluster means of the characteristics under study (y) and the auxiliary characteristics (x_1 and x_2) for clusters of size M .

ρ' is the intra correlation coefficient for y .

R^2 is the multiple correlation coefficient between y and x_1 and x_2 .



3. Conditions for cluster sampling in conjunction with ratio or regression method to be more efficient than SRS

Mishro and Sukhatme [9] gave the following conditions under which cluster sampling in conjunction with ratio or regression method is more efficient than SRS even if intra class correlation of the elements within the cluster has positive correlation.

$$\rho_{bj} > \rho_j + \frac{(M-1)\rho(1-\rho_j)}{1+(M-1)\rho} ; j=1, 2 \text{ for ratio method} \quad (1)$$

$$\rho_{bj}^2 > \rho_j^2 + \frac{(M-1)\rho^2(1-\rho_j^2)}{1+(M-1)\rho^2} ; j=1, 2 \text{ for regression method} \quad (2)$$

The efficiencies E_1 and E_2 of cluster sampling in conjunction with ratio or regression estimators relative to SRS, for given budget are:

$$E_1 = \frac{(1-\rho_j)}{(1-\rho_{bj})\{1+(M-1)\rho\}} * \frac{n^*}{n} \quad (3)$$

$$E_2 = \frac{(1-\rho_j^2)}{(1-\rho_{bj}^2)\{1+(M-1)\rho^2\}} * \frac{n^*}{n} \quad (4)$$

where fixed budget permits n^*M ($n^* > n$) units for cluster sampling and nM units for SRS.

3.1 Cluster sampling in conjunction with ratio or regression method based on two auxiliary variables to be more efficient than SRS:

Agarwal and Goel [1] used two auxiliary variables to obtain ratio and regression estimators in conjunction with cluster sampling to increase the efficiency of cluster sampling further. Al-Mannai [2] suggested a ratio estimator based on two auxiliary variables to estimate the mean of a survey variable when the means as well as the coefficients of variation of two auxiliary variables are known.

$$\bar{y}_R = \bar{y} \left[\frac{k_1}{k_2} \bar{X}_1 + \bar{X}_2 \right] / \left(\frac{k_1}{k_2} \bar{x}_1 + \bar{x}_2 \right)$$

where $\frac{k_1}{k_2} = \left(\frac{\sigma_{x_1}}{\sigma_{x_2}} \right) \left(\frac{d - \rho_{12}}{1 - d \rho_{12}} \right)$ and $d = \frac{\rho_1}{\rho_2}$

The estimator suggested by [2] is an improved estimator over [1] ratio estimator. Hence, we will consider the estimator in [2] in conjunction with cluster sampling. The

efficiencies E_3 and E_4 of cluster sampling in conjunction with ratio estimator using two auxiliary variables are:

$$E_3 = \frac{(1-R)}{(1-\rho_b)\{1+(M-1)\rho\}} * \frac{n^*}{n} \quad (5)$$

$$E_4 = \frac{(1-R^2)}{(1-\rho_b^2)\{1+(M-1)\rho^2\}} * \frac{n^*}{n}$$

4. Illustration

Agarwal and Goel [1] considered only a single population in the illustration, which is not sufficient to develop the confidence of survey practitioners. Therefore, in this paper we are considering 11 well known populations taken from [4], [5], [7] and [11] to see the gain or loss of cluster sampling in conjunction with ratio estimator using single and also two auxiliary variables.

Table-1 gives the characteristics of the populations such as population size N , coefficients of variation of the study variable (y), the auxiliary variables x_1 and x_2 , the correlation coefficients between (y, x_1) and (y, x_2). In table -1, the population size varies from 45 to 705, the coefficient of variation of y from 19.81 % to 159.5 %, the coefficient of variation of x_1 from 4% to 124.62 %, the coefficient of variation of x_2 from 0.90 % to 116.69 %. The correlation coefficient between (y, x_1) varies from 0.189 to 0.842, while the correlation coefficient between (y, x_2) varies from 0.025 to 0.843. It can be noted that in the above described populations, four have $N \geq 100$. A few populations have very high values of coefficient of variation of either y , or x_1 or x_2 ; more than 50% populations have correlation coefficient between (y, x_1) less than 0.5; while for most of the populations the correlation coefficient between (y, x_2) is either close to 0.5 or less than 0.5. Thus most of the populations represent an extreme situation with respect to one characteristics or the other. We considered these extreme populations in our study because sometimes the population elements are naturally grouped into subpopulations and lists of those subpopulations either already exist or easy and cheap to obtain and hence the survey practitioner has no other choice but to go for cluster sampling.

Table 2 represents the correlation coefficients between cluster means for sizes $M = 2, 3$ and 4. It is worth and important to mention that the correlation coefficient between cluster means for $M = 2, 3$ and 4 decreases for population numbers 2, 4, 6 and 8 of Table 1. Besides, in many cases the correlation between two characteristics of cluster means is not directly proportional to the size of the clusters and hence it contradicts the illustrations obtained by Zarkovic [13].



Table 3 and 4 represent the efficiency of cluster sampling for cluster sizes $M = 2, 3$ and 4 with ratio estimator using x_1 or x_2 or both auxiliary variables over SRS if intra class correlation of the elements within the cluster has positive correlation. For those populations for which the condition of equation 1, is satisfied, the gain of efficiency of cluster sampling for cluster sizes $M = 2, 3$ and 4 with ratio estimator using x_1 or x_2 or both auxiliary variables over SRS is from marginal to substantial even if intra class correlation of the elements within the cluster has positive correlation. In table 4, we considered the situations when fixed budget permits n^*M ($n^* > n$) units for cluster sampling and nM units for SRS. From table 3, it can be noted that for more than 50% populations, the cluster sampling in conjunction with ratio estimator using x_1 or x_2 or both auxiliary variables over SRS is less efficient. Therefore, the practitioners must verify the condition of equation 1, before using cluster sampling in conjunction with ratio estimator.

The similar observations can be made from Table 5 and 6. For those populations for which condition of equation

2, is satisfied, the gain in efficiency of cluster sampling for cluster sizes $M = 2, 3$ and 4 with regression estimator using x_1 or x_2 or both auxiliary variables over SRS is from marginal to substantial even if intra class correlation of the elements within the cluster has positive correlation. In table 6, we considered the situations when fixed budget permits n^*M ($n^* > n$) units for cluster sampling and nM units for SRS.

Table 7: Percentage gain or loss in efficiency $[= \frac{E_3 - E_1}{E_1} * 100]$ of cluster sampling with ratio estimator

using two auxiliary variables over cluster sampling with ratio estimators using auxiliary variable x_1 or x_2 for cluster sizes $M = 2, 3$ and 4. It is interesting to note that it is not necessary that cluster sampling with ratio estimator using two auxiliary variables would be better than cluster sampling with ratio estimator using auxiliary variable x_1 or x_2 for cluster sizes $M = 2, 3$ and 4.

Table1: Characteristics of the populations

S. No.	N	C_y	C_{x_1}	C_{x_2}	ρ_{01}	ρ_{02}	ρ_{12}
1	52	97.17	124.62	116.69	0.340	0.505	0.306
2	100	28.97	34.85	58.16	0.380	0.502	0.456
3	73	78.35	28.67	52.20	0.365	0.446	0.763
4	522	49.63	31.45	0.90	0.820	0.556	0.441
5	81	67.92	11.36	26.66	0.535	0.441	0.414
6	705	21.31	16.35	24.21	0.366	0.399	0.443
7	72	159.52	4.00	6.54	0.609	0.370	0.548
8	76	30.22	9.42	22.21	0.202	0.025	0.053
9	113	19.81	8.38	30.79	0.189	0.533	0.001
10	93	53.76	74.30	74.23	0.571	0.548	0.174
11	45	37.24	47.50	45.10	0.842	0.843	0.974
min	45	19.81	4.00	0.90	0.189	0.025	0.053
max	705	159.52	124.62	116.69	0.842	0.843	0.974
Q_1	72.5	29.59	10.39	23.21	0.353	0.420	0.240
median	81	49.63	28.67	30.79	0.380	0.502	0.441
Q_3	106.5	73.13	41.17	55.18	0.590	0.541	0.502



Table2: Correlation coefficients between cluster means for sizes M =2, 3 and 4

Number	M=2 ρ_{b1}	M=2 ρ_{b2}	M=3 ρ_{b1}	M=3 ρ_{b2}	M=4 ρ_{b1}	M=4 ρ_{b2}	M=2 ρ_b	M=3 ρ_b	M=4 ρ_b
1	0.5662	0.6457	0.1694	0.5979	0.7455	0.5846	0.7066	0.5981	0.7801
2	0.3403	0.3898	0.3417	0.1934	0.3870	0.5769	0.4286	0.3625	0.5802
3	0.5651	0.6049	0.1178	0.4468	0.6235	0.6036	0.6121	0.4966	0.6756
4	0.7275	0.4587	0.6947	0.3369	0.6905	0.3969	0.7722	0.7319	0.7262
5	0.5090	0.5932	0.5016	0.1545	0.6991	0.4859	0.6337	0.5016	0.7400
6	0.4586	0.4673	0.2532	0.2547	0.4204	0.3636	0.5403	0.3011	0.4705
7	0.5642	0.4204	0.7032	0.4208	0.5934	0.5153	0.5750	0.7073	0.6351
8	0.1604	0.0068	0.0838	0.1516	0.1860	0.0620	0.1604	0.1717	0.2154
9	0.2483	0.5186	0.4259	0.5426	0.0314	0.5757	0.5729	0.7024	0.5830
10	0.4694	0.6053	0.3850	0.5580	0.3354	0.6893	0.6806	0.6409	0.7305
11	0.8850	0.8996	0.8574	0.8596	0.8636	0.8698	0.8999	0.8608	0.8717
min	0.1604	0.0068	0.0838	0.1516	0.0314	0.0620	0.1604	0.1717	0.2154
max	0.8850	0.8996	0.8574	0.8596	0.8636	0.8698	0.8999	0.8608	0.8717
Q_1	0.3995	0.4395	0.2113	0.2240	0.3612	0.4414	0.5566	0.4296	0.5816
median	0.5090	0.5186	0.3850	0.4208	0.5934	0.5757	0.6121	0.5981	0.6756
Q_3	0.5656	0.6051	0.5981	0.5503	0.6948	0.5941	0.6936	0.7048	0.7353

Table 3: Efficiency of cluster sampling for cluster sizes M =2, 3 and 4 with ratio estimator using x_1 or x_2 or both auxiliary variables over SRS if intra class correlation of the elements within the cluster has positive correlation.

Number	M=2 using x_1	M=2 using x_2	M=3 using x_1	M=3 using x_2	M=4 using x_1	M=4 using x_2	M=2 using both	M=3 using both	M=4 using both
1	1.64	1.77	0.93	1.45	3.34	1.53	1.69	0.58	2.17
2	0.83	0.67	0.75	0.49	0.73	0.85	0.73	0.34	0.09
3	1.21	0.92	0.51	0.71	1.04	0.87	1.18	0.37	0.15
4	0.37	0.33	0.23	0.26	0.17	0.21	0.37	0.20	0.10
5	0.67	0.68	0.51	0.36	0.69	0.49	0.80	0.36	0.01
6	0.63	0.43	0.31	0.30	0.31	0.27	0.64	0.76	0.76
7	0.64	0.64	0.73	0.60	0.43	0.58	0.65	0.15	0.25
8	1.11	1.15	1.22	1.61	1.71	1.60	1.10	0.65	0.54
9	1.12	1.05	1.52	1.10	0.94	1.23	1.05	1.05	0.52
10	0.73	0.90	0.57	0.84	0.49	1.10	0.76	0.67	1.04
11	1.31	1.05	1.01	1.02	1.01	1.05	1.45	4.53	4.74
min	0.37	0.33	0.23	0.26	0.17	0.21	0.37	0.15	0.01
max	1.64	1.77	1.52	1.61	3.34	1.60	1.69	4.53	4.74
Q_1	0.65	0.66	0.51	0.43	0.46	0.54	0.69	0.35	0.125
median	0.83	0.90	0.73	0.71	0.73	0.87	0.80	0.58	0.52
Q_3	1.16	1.05	0.97	1.06	1.03	1.16	1.14	0.715	0.90



Table 4: Efficiency of cluster sampling for cluster sizes M =2, 3 and 4 with ratio estimator using x_1 or x_2 or both auxiliary variables over SRS if intra class correlation of the elements within the cluster has positive correlation and fixed budget permits $\frac{n^*}{n} = 1.2$.

Number	M=2 using x_1	M=2 using x_2	M=3 using x_1	M=3 using x_2	M=4 using x_1	M=4 using x_2	M=2 using both	M=3 using both	M=4 using both
1	1.97	2.12	1.12	1.74	4.01	1.84	2.03	0.70	2.61
2	1.00	0.81	0.90	0.59	0.88	1.02	0.87	0.41	0.11
3	1.45	1.11	0.61	0.85	1.25	1.04	1.42	0.44	0.18
4	0.44	0.39	0.27	0.31	0.20	0.26	0.44	0.25	0.12
5	0.80	0.82	0.61	0.43	0.83	0.58	0.96	0.43	0.01
6	0.76	0.52	0.38	0.36	0.37	0.32	0.77	0.91	0.91
7	0.77	0.77	0.87	0.72	0.52	0.70	0.78	0.18	0.30
8	1.33	1.38	1.46	1.93	2.05	1.92	1.32	0.79	0.65
9	1.34	1.26	1.82	1.32	1.12	1.48	1.26	1.25	0.62
10	0.88	1.08	0.69	1.01	0.58	1.32	0.91	0.80	1.25
11	1.57	1.26	1.21	1.22	1.21	1.26	1.74	5.44	5.69
min	0.44	0.39	0.27	0.31	0.20	0.26	0.44	0.18	0.01
max	1.97	2.12	1.82	1.93	4.01	1.92	2.03	5.44	5.69
Q_1	0.78	0.79	0.61	0.51	0.55	0.64	0.83	0.42	0.15
median	1.00	1.08	0.87	0.85	0.88	1.04	0.96	0.70	0.62
Q_3	1.40	1.26	1.16	1.27	1.23	1.40	1.37	0.855	1.08

Table 5: Efficiency of cluster sampling for cluster sizes M =2, 3 and 4 with regression estimator using x_1 or x_2 or both auxiliary variables over SRS if intra class correlation of the elements within the cluster has positive correlation.

Number	M=2 using x_1	M=2 using x_2	M=3 using x_1	M=3 using x_2	M=4 using x_1	M=4 using x_2	M=2 using both	M=3 using both	M=4 using both
1	1.21	1.38	1.07	1.49	2.56	1.46	1.52	1.52	2.32
2	1.11	0.78	0.77	0.56	0.73	0.81	0.78	0.78	0.78
3	1.60	1.05	0.62	0.62	0.88	0.78	1.06	1.06	0.91
4	3.72	0.48	0.24	0.23	0.18	0.24	0.39	0.39	0.17
5	1.64	0.88	0.52	0.37	0.62	0.47	0.78	0.78	0.65
6	7.54	0.58	0.34	0.25	0.30	0.27	0.61	0.61	0.29
7	1.56	0.74	0.69	0.47	0.44	0.53	0.67	0.67	0.47
8	0.86	1.17	1.35	1.78	1.73	1.75	1.15	1.15	1.75
9	0.99	1.01	1.27	1.13	1.08	1.20	1.05	1.05	1.15
10	0.97	1.00	0.65	0.77	0.57	1.00	0.78	0.78	0.75
11	1.42	1.45	1.00	0.97	1.00	1.04	1.41	1.41	1.02
min	0.86	0.48	0.24	0.23	0.18	0.24	0.39	0.39	0.17
max	7.54	1.45	1.35	1.78	2.56	1.75	1.52	1.52	2.32
Q_1	1.05	0.76	0.57	0.42	0.50	0.50	0.72	0.72	0.56
median	1.42	1.00	0.69	0.62	0.73	0.81	0.78	0.78	0.78
Q_3	1.62	1.11	1.03	1.05	1.04	1.12	1.10	1.10	1.09



Table 6: Efficiency of cluster sampling for cluster sizes M =2, 3 and 4 with regression estimator using x_1 or x_2 or both auxiliary variables over SRS if intra class correlation of the elements within the cluster has positive correlation and fixed budget permits $\frac{n^*}{n} = 1.2 \cdot$

Number	M=2 using x_1	M=2 using x_2	M=3 using x_1	M=3 using x_2	M=4 using x_1	M=4 using x_2	M=2 using both	M=3 using both	M=4 using both
1	1.45	1.66	1.28	1.79	3.08	1.75	1.83	1.83	2.79
2	1.33	0.94	0.93	0.67	0.87	0.97	0.94	0.94	0.94
3	1.92	1.26	0.75	0.74	1.05	0.94	1.27	1.27	1.09
4	4.46	0.58	0.29	0.27	0.22	0.29	0.46	0.46	0.21
5	1.97	1.06	0.63	0.44	0.75	0.57	0.93	0.93	0.78
6	9.04	0.70	0.41	0.30	0.35	0.33	0.73	0.73	0.34
7	1.87	0.89	0.82	0.57	0.52	0.63	0.80	0.80	0.57
8	1.03	1.40	1.62	2.14	2.08	2.10	1.38	1.38	2.10
9	1.19	1.22	1.52	1.36	1.30	1.44	1.26	1.26	1.38
10	1.16	1.19	0.78	0.92	0.69	1.21	0.94	0.94	0.90
11	1.70	1.74	1.20	1.16	1.20	1.24	1.69	1.69	1.22
min	1.03	0.58	0.29	0.27	0.22	0.29	0.46	0.46	0.21
max	9.04	1.74	1.62	2.14	3.08	2.10	1.83	1.83	2.79
Q_1	1.26	0.92	0.69	0.50	0.61	0.60	0.86	0.86	0.67
median	1.70	1.19	0.82	0.74	0.87	0.97	0.94	0.94	0.94
Q_3	1.95	1.33	1.24	1.26	1.25	1.34	1.32	1.32	1.30

Table 7: Percentage gain or loss in efficiency $[= \frac{E_3 - E_1}{E_1} * 100]$ of cluster sampling with ratio estimator using two auxiliary variables over cluster sampling with ratio estimators using auxiliary variable x_1 or x_2 for cluster sizes M =2, 3 and 4

Number	M=2 using x_1	M=2 using x_2	M=3 using x_1	M=3 using x_2	M=4 using x_1	M=4 using x_2
1	-2.73	-4.55	162.55	140.37	-165.07	-241.62
2	12.48	8.14	54.84	31.10	-87.86	-89.57
3	2.41	27.98	27.56	47.97	-86.00	-83.12
4	-1.04	12.50	190.67	179.98	-41.49	-53.70
5	-19.14	16.97	30.45	1.97	-101.20	-101.70
6	-1.93	48.53	-142.92	-155.70	147.72	186.91
7	-2.13	2.08	120.22	124.50	-42.46	-57.42
8	0.39	-3.68	46.28	59.29	-68.19	-66.04
9	5.76	0.79	168.79	195.27	-155.26	-142.08
10	-4.30	-15.15	216.93	179.88	-314.15	-195.16
11	-10.68	38.64	549.71	544.62	-570.10	-550.83
min	-19.14	-15.15	-142.92	-155.70	-570.10	-550.83
max	12.48	48.53	549.71	544.62	147.72	186.91
Q_1	-3.51	-1.44	38.37	39.54	-160.16	-168.62
median	-1.93	8.14	120.22	124.50	-87.86	-89.57
Q_3	1.40	22.48	179.73	179.93	-55.33	-61.73



5. Concluding Remarks

Sometimes the population elements are naturally grouped into subpopulations and lists of those subpopulations either already exist or easy and cheap to obtain. In such cases cluster sampling is the obvious choice of survey practitioners. However, one of the main disadvantage of cluster sampling is that its efficiency is less than SRS for given sample size because the intra class correlation of the elements within the cluster has positive correlation. In this paper we have studied empirically the gain or loss of cluster sampling in conjunction with ratio method of estimation using single auxiliary variable or using two auxiliary variables over simple random sampling (SRS), if intra class correlation of the elements within the cluster has positive correlation. As mentioned above that if clusters are already available, it is quite possible that the characteristics of such populations may be undesirable [Considerably high values of coefficient of variation of either y , or x_1 or x_2 ; or low values of correlation coefficient between (y, x_1) or between (y, x_2)]. Therefore, we have considered more than 50% extreme populations which may help the survey practitioners to know the magnitude of gain or loss in efficiency of cluster sampling using single or two auxiliary variables in conjunction with ratio method of estimation. To have more insight, the clusters are formed randomly for sample sizes 2, 3 and 4. From the condition of equations (1) and (2), it can be seen that it contains characteristics of the populations, and hence one can verify them using sample values for large sample, and may decide whether to go for cluster sampling in conjunction with ratio or regression methods of estimation.

ACKNOWLEDGMENT

The author would like to express her gratitude to the University of Bahrain in where premises this research was conducted and to the referee for the constructive comments.

REFERENCES

- [1] S.K. Agarwal and B.B.P.S. Goel, "Use of certain type of multi auxiliary information for increasing the efficiency of cluster sampling with ratio and regression estimators", *JISAS*, Vol 32, No. 2, pp 13-18, 1980.
- [2] M. Al-Mannai, "Two Auxiliary Variables in Ratio Method of Estimation and its Application", *JSAP*, vol3(1),pp.101-108,Doi: <http://dx.doi.org/10.12785/jsap/030109>, 2014.
- [3] W.W. Daniel, *Biostatistics: A foundation for analysis in the health sciences*. 8th ed. John Wiley and Sons: USA, 2005.
- [4] L. P. Galway, N. Bell, S. A. Al Shatari, A. Hagopian, and G. Burnham, "A two-stage cluster sampling method using gridded population data, a GIS, and Google EarthTM imagery in a population-based mortality survey in Iraq", *Int J Health Geogr* 11: 12. doi: 10.1186/1476-072x-11-12, 2012.
- [5] K. Himelein, S. Eckman, S. Murray, "The use of random geographic cluster sampling to survey pastoralists", Policy Research Working Paper Report Number WPS6589, Ethiopia, 2013 .
- [6] M. H. Kutner, C. J. Nachtsheim, C. J. Neter and W. Li, *Applied Linear Statistical Models*, 5th ed., McGraw-Hill International Edition, Singapore, 2005.
- [7] W. Mendenhall and T. Sincich, *A second Course ins Statistics: Regression Analysis*, 6th ed., Pearson, USA, 2003.
- [8] G. K. Mishro, and B. V. Sukhatme, "Efficiency of cluster sampling in conjunction with ratio and regression method of estimation", *J. Indian Soc. of Ag. Statistics*, vol. 24(2), pp. 81-90, 1972.
- [9] S. K. Morris, and C. K. Nguyen: "A review of the cluster survey sampling method in humanitarian emergencies" *.Public Health Nurs*, vol 25 (3), pp.70-374, 2008.
- [10] L. Roberts, R. Lafta, R. Garfield, J. Khudhairi, G. Burnham, "Mortality before and after the 2003 invasion of Iraq:cluster sample survey", *The Lancet*, vol.364 pp. 1857-1864, 2004.
- [11] S. J. Sheater, *A modern approach to regression with R*, Springer, USA, 2009.
- [12] A. G. Turner, R. J. Magnani, M. A. Shuaib," Not Quite as Quick but Much Cleaner Alternative to the Expanded Programme on Immunization (EPI) Cluster Survey Design", *International Journal of Epidemiology*, vol. 25, pp.198-203, 1996.
- [13] S. S. Zarkovic and J. Krane, "Some efficient ways of Cluster Sampling", *Proceedings of 35th session of I. S. I. , Belgrade*, 1965.