STATISTICS IN TRANSITION-new series, October 2010
Vol. 11, No. 2, pp. 253—265

ON THE ESTIMATION OF RATIO-CUM-PRODUCT ESTIMATORS USING TWO-STAGE SAMPLING

Olufadi, Yunusa[1]

ABSTRACT

In this paper, we apply two-stage sampling which is not only commonly used in survey sampling but also has great advantages over element and cluster sampling. These advantages are comprehensively described in some sampling literature. Thus, we examine two-stage ratio-cum-product estimators with unequal sub-sampling fractions and obtain their MSE equations. The optimum sampling and sub-sampling fractions were also derived for these estimators. It is shown theoretically that these two-stage estimators will be more efficient than Singh (1965, 1967) estimators if certain conditions are satisfied. Finally, a numerical illustration with discussions is carried out to show the application of this technique.

Key words: auxiliary information, efficiency, mean square error, optimum sampling fraction, ratio-cum-product estimator, two-stage sampling.

1. Introduction

The literature on survey sampling (Cochran (1977), Sukhatme et al. (1984) and references cited therein) describes a great number of techniques for using auxiliary information by means of ratio and product estimators. It is well known that Ratio estimators are efficient when there is a strong positive correlation (and preferably also a strong ratio relationship) between the variable of interest and the supplementary variable. Product estimators are only efficient when the variable of interest is negatively correlated with the supplementary variable, and preferably when it has a strong inverse relationship with it, so that the products are very similar.

However, it has been established theoretically that, in general, the linear regression estimator is more efficient than ratio and product estimators except when the regression line of y on x passes through the neighbourhood of the origin, in which case the efficiencies of these estimators are almost equal. Various

[1] Department of Statistics, University of Ilorin, Ilorin, Nigeria.

improvements of these estimators have been considered by many authors particularly in the presence of more than one auxiliary variable.

Olkin (1958) was the first author to deal with the problem of estimating the mean of a survey variable when auxiliary variables are made available. He suggested the use of more than one auxiliary variable, considering a linear combination of ratio estimators based on each auxiliary variable separately. The coefficients of the linear combination were determined so as to minimize the variance of the estimator. Analogously to Olkin, Singh (1967b) gave a multivariate expression of Murthy's (1964) product estimator, while Raj (1965) suggested a method of using multi-auxiliary variables through a linear combination of single difference estimators. Moreover, Rao and Mudholkar (1967) proposed a multivariate estimator based on a weighted sum of single ratio and product estimators.

Many other contributions are present in sample survey literature and recently some new estimators have emerged. Tracy *et al.* (1996) and Perri (2005), when two auxiliary variables are available, proposed an alternative to Singh's (1965, 1967) ratio-cum-product estimators while Kadilar and Cingi (2004, 2005) examined the combination of regression-type estimators in the presence of two auxiliary variables. Other references are: Muhammad *et al.* (2010), Abdul and Javid (2010), Tailor and Sharma (2009), Diana and Perri (2007), Samiuddin and Hanif (2007), Abu-Dayyeh *et al.* (2003), Upadhyaya and Singh (1999, 2003) and many others.

In most practical situations, the sampling frame may be unavailable and when it does, it may be expensive or not feasible to obtain the sampling units directly from a population of interest. Thus, in this case, we can make use of two-stage sampling whose efficiency could be improved by the use of auxiliary information. The use of two stage sampling in ratio-cum-product estimators when there is more than one auxiliary variable is scarcely considered and to the best of our knowledge, we can only mention Sukhatme *et al.* (1984), Sahoo and Panda (1997, 1999) and Hossain and Ahmed (2001) in the case of single auxiliary variable. Thus, in this paper, an attempt is made to apply two-stage sampling for estimating the population mean in Singh's (1965, 1967a) estimators which involve two auxiliary variables.

2. Two-stage sampling set-up

Let U be a finite population partitioned into N First Stage Units (FSU) denoted by $U_1, U_2, ..., U_I, ...U_N$ such that the number of Second Stage Units (SSU) in U_i is M_i. Let y and (x, z) be the study variable and auxiliary variates taking the values y_{ij} and (x_{ij}, z_{ij}) respectively, for the j^{th} SSU on the unit U_i $(i = 1,2,..., N; j = 1,2,..., M_i)$. Where x is assumed to be positively correlated with y and z is negatively correlated with y. Suppose it is of interest to estimate the population mean $\overline{Y}$ of y and assume that the population means $\left(\overline{X}, \overline{Y}\right)$ of (x, z) are known.

Define $\overline{P}_i = \frac{1}{M_i}\sum_{j=1}^{M_i} P_{ij}$; $\overline{P} = \frac{1}{N}\sum_{i=1}^{N} \overline{P}_i$; $P = X$, Y , Z . Assume that a sample s of n FSU's is drawn from U and then a sample s_i of m_i SSU's from the i^{th} selected FSU from U_i using simple random sampling without replacement.

Define also $\overline{p}_i = \frac{1}{m_i}\sum_{j=1}^{m_i} p_{ij}$; $\overline{p} = \frac{1}{n}\sum_{i=1}^{n} \overline{p}_i$; $p = x$, y , z .

3. Ratio-cum-product estimators

The product-ratio estimators using one dual auxiliary variable were introduced by Badhopadyaya (1980) and Srivenkataramana (1980). The estimators are defined for simple random sampling and can be effective when one auxiliary variable is positively correlated with the study variable and the other variable is negatively correlated with study variate. Recently, Singh *et al.* (2005), Singh and Espejo (2007), Singh and Upadhyaya (1995) and a host of others have used the combination of ratio and product estimators to estimate the unknown population mean $\overline{Y}$, when $\overline{X}$ and $\overline{Z}$ are known but our interest in this paper is to apply two-stage sampling with unequal FSU to the ratio-cum-product estimators developed by Singh (1965, 1967a). These estimators are as given below:

$$t_1 = \overline{y}\frac{\overline{X}}{\overline{x}}\frac{\overline{z}}{\overline{Z}}\, ; \quad t_2 = \overline{y}\frac{\overline{X}}{\overline{x}}\frac{\overline{Z}}{\overline{z}}\, ; \quad t_3 = \overline{y}\frac{\overline{x}}{\overline{X}}\frac{\overline{z}}{\overline{Z}}\, ; \quad t_4 = \overline{y}\frac{\overline{x}}{\overline{X}}\frac{\overline{Z}}{\overline{z}}\, .$$

The estimators are biased even if the bias may be considered negligible for large samples. Under suitable conditions, involving the correlation coefficients between the variables, it is easy to prove (Singh, 1965) that they are more efficient than ratio and product estimators which make use of a single auxiliary variable. Their corresponding Mean Square Errors (MSE) to the first degree of approximation is presented below:

$$MSE(t_1) = \lambda\left[S_y^2 + R_1^2 S_x^2 + R_2^2 S_z^2 - 2\left(R_1 S_{xy} - R_2 S_{yz} + R_1 R_2 S_{xz}\right)\right] \quad (1)$$

$$MSE(t_2) = \lambda\left[S_y^2 + R_1^2 S_x^2 + R_2^2 S_z^2 - 2\left(R_1 S_{xy} + R_2 S_{yz} - R_1 R_2 S_{xz}\right)\right] \quad (2)$$

$$MSE(t_3)=\lambda\left[S_y^2+R_1^2S_x^2+R_2^2S_z^2+2\left(R_1S_{xy}+R_2S_{yz}+R_1R_2S_{xz}\right)\right] \tag{3}$$

$$MSE(t_4)=\lambda\left[S_y^2+R_1^2S_x^2+R_2^2S_z^2-2\left(R_1S_{xy}-R_2S_{yz}-R_1R_2S_{xz}\right)\right] \tag{4}$$

where, $R_1=\dfrac{\overline{Y}}{\overline{X}}$; $R_2=\dfrac{\overline{Y}}{\overline{Z}}$; $\lambda=\dfrac{1-f}{n}$; $f=\dfrac{n}{N}$; $S_p^2=\dfrac{1}{N-1}\sum_{i=1}^{N}\left(P_i-\overline{P}\right)^2$; $S_{pq}=\dfrac{1}{N-1}\sum_{i=1}^{N}\left(P_i-\overline{P}\right)\left(Q_i-\overline{Q}\right)$ $P,Q=X,Y,Z$; $p,q=x,y,z$; $P\neq Q$ and $p\neq q$.

4. Suggested method of estimation

The suggested two-stage ratio-cum-product estimators are as defined below:

$t_1^*=\overline{y}^*\dfrac{\overline{X}}{\overline{x}^*}\dfrac{\overline{z}^*}{\overline{Z}}$; $t_2^*=\overline{y}^*\dfrac{\overline{X}}{\overline{x}^*}\dfrac{\overline{Z}}{\overline{z}^*}$; $t_3^*=\overline{y}^*\dfrac{\overline{x}^*}{\overline{X}}\dfrac{\overline{z}^*}{\overline{Z}}$; $t_4^*=\overline{y}^*\dfrac{\overline{x}^*}{\overline{X}}\dfrac{\overline{Z}}{\overline{z}^*}$.

Define $\overline{y}^*=\overline{Y}(1+e_0)$, $\overline{x}^*=\overline{X}(1+e_1)$, $\overline{z}^*=\overline{Z}(1+e_2)$ such that $E(e_h)=0$; $E\left(e_h^2\right)=\dfrac{V\left(\overline{P}\right)}{\overline{P}^2}$ for $h=0,1,2$; $E(e_he_k)=\dfrac{Cov\left(\overline{P},\overline{Q}\right)}{\overline{P}\overline{Q}}$ for $h,k=0,1,2$ and $h\neq k$.

Under two-stage sampling with unequal FSU,

$V\left(\overline{P}\right)=\lambda S_{1p}^2+\sum_{i=1}^{N}\lambda_{2i}S_{2pi}^2$; $Cov\left(\overline{P},\overline{Q}\right)=\lambda S_{1pq}+\sum_{i=1}^{N}\lambda_{2i}S_{2pqi}$ where

$\lambda_{2i}=\dfrac{M_i^2(1-f_{2i})}{nNm_i}$; $f_{2i}=\dfrac{m_i}{M_i}$; $S_{1p}^2=\dfrac{1}{N-1}\sum_{i=1}^{N}\left(\overline{P}_i-\overline{P}\right)^2$;

$S_{2pi}^2=\dfrac{1}{N-1}\sum_{i=1}^{N}\left(P_{ij}-\overline{P}_i\right)^2$; $S_{1pq}=\dfrac{1}{N-1}\sum_{i=1}^{N}\left(\overline{P}_i-\overline{P}\right)\left(\overline{Q}_i-\overline{Q}\right)$;

$S_{2pqi}=\dfrac{1}{N-1}\sum_{i=1}^{N}\left(P_{ij}-\overline{P}_i\right)\left(Q_{ij}-\overline{Q}_i\right)$. S_{1p}^2 and S_{2pi}^2 are the variance among FSU means and variance among subunits for the i^{th} FSU while S_{1pq} and S_{2pqi} are their corresponding covariances.

In order to evaluate the precision of these estimators with respect to Singh (1965, 1967a) estimators, their MSEs need to be obtained; and this can be found to the first degree of approximation using Taylor's linearization method described by Wolter (1985).

If we express t_1^* in terms of $e's$, we have

$$t_1^* = \overline{Y}(1+e_0)(1+e_2)(1+e_1)^{-1} \quad (5)$$

Expanding the right hand side of (5) and neglecting terms involving powers of $e's$ greater than two, $t_1^* - \overline{Y}$ yields

$$t_1^* - \overline{Y} = \overline{Y}\left(e_1^2 + e_0 + e_2 - e_1 + e_0 e_2 - e_0 e_1 - e_1 e_2\right) \quad (6)$$

Squaring both sides of (6) and neglecting terms of order greater than two, we have

$$\left(t_1^* - \overline{Y}\right)^2 = \overline{Y}^2\left(e_1^2 + e_0^2 + e_2^2 + 2e_0 e_2 - 2e_0 e_1 - 2e_1 e_2\right) \quad (7)$$

Taking expectation on both sides of (7) and simplifying results in

$$MSE\left(t_1^*\right) = \lambda S_{1R_0^*}^2 + \sum_{i=1}^{N} \lambda_{2i} S_{2R_{0i}^*}^2 \quad (8)$$

where

$$S_{1R_0^*}^2 = \left[S_{1y}^2 + R_1^2 S_{1x}^2 + R_2^2 S_{1z}^2 - 2\left(R_1 S_{1xy} - R_2 S_{1yz} + R_1 R_2 S_{1xz}\right)\right]$$
$$S_{2R_{0i}^*}^2 = \left[S_{2yi}^2 + R_1^2 S_{2xi}^2 + R_2^2 S_{2zi}^2 - 2\left(R_1 S_{2xyi} - R_2 S_{2yzi} + R_1 R_2 S_{xzi}\right)\right].$$

Following the procedure above, it is easy to verify that

$$MSE\left(t_2^*\right) = \lambda S_{1R_{00}^*}^2 + \sum_{i=1}^{N} \lambda_{2i} S_{2R_{00i}^*}^2 \quad (9)$$

$$MSE\left(t_3^*\right) = \lambda S_{1P_0^*}^2 + \sum_{i=1}^{N} \lambda_{2i} S_{2P_{0i}^*}^2 \quad (10)$$

$$MSE\left(t_4^*\right) = \lambda S_{1p_{00}^*}^2 + \sum_{i=1}^{N} \lambda_{2i} S_{2P_{00i}^*}^2 \quad (11)$$

where

$$S_{1R_{00}^*}^2 = \left[S_{1y}^2 + R_1^2 S_{1x}^2 + R_2^2 S_{1z}^2 - 2\left(R_1 S_{1xy} + R_2 S_{1yz} - R_1 R_2 S_{1xz}\right)\right]$$
$$S_{2R_{00i}^*}^2 = \left[S_{2yi}^2 + R_1^2 S_{2xi}^2 + R_2^2 S_{2zi}^2 - 2\left(R_1 S_{2xyi} + R_2 S_{2yzi} - R_1 R_2 S_{xzi}\right)\right]$$
$$S_{1P_0^*}^2 = \left[S_{1y}^2 + R_1^2 S_{1x}^2 + R_2^2 S_{1z}^2 + 2\left(R_1 S_{1xy} + R_2 S_{1yz} + R_1 R_2 S_{1xz}\right)\right]$$
$$S_{2P_{0i}^*}^2 = \left[S_{2yi}^2 + R_1^2 S_{2xi}^2 + R_2^2 S_{2zi}^2 + 2\left(R_1 S_{2xyi} + R_2 S_{2yzi} + R_1 R_2 S_{xzi}\right)\right]$$
$$S_{1P_{00}^*}^2 = \left[S_{1y}^2 + R_1^2 S_{1x}^2 + R_2^2 S_{1z}^2 + 2\left(R_1 S_{1xy} - R_2 S_{1yz} - R_1 R_2 S_{1xz}\right)\right]$$
$$S_{2P_{00i}^*}^2 = \left[S_{2yi}^2 + R_1^2 S_{2xi}^2 + R_2^2 S_{2zi}^2 + 2\left(R_1 S_{2xyi} - R_2 S_{2yzi} - R_1 R_2 S_{xzi}\right)\right]$$

REMARK 1: The MSE of t_1^* depends on the unknown population variances $S^2_{1R_0^*}$ and $S^2_{2R_{0i}^*}$ whose values may be obtained from the data from a past survey. When this is not feasible, the data from a pilot survey could be used to estimate these parameters. However, the values of $S^2_{1R_0^*}$ and $S^2_{2R_{0i}^*}$ from these estimates will depart from the actual population variances depending on how close our estimates is to the true values of $S^2_{1R_0^*}$ and $S^2_{2R_{0i}^*}$. Similar explanation goes for (9), (10) and (11).

5. Efficiency comparison

We compare the MSE of the suggested method of estimation given in (8), (9), (10) and (11) with that of Singh's (1965, 1967) estimators. We will have the conditions as follows:

(i) $$MSE(t_1^*) - MSE(t_1) \le 0$$

$$\lambda\left(S^2_{1R_0^*} - S^2_{1R}\right) + \sum_{i=1}^{N} \lambda_{2i} S^2_{2R_{0i}^*} \le 0$$

(ii) $$MSE(t_2^*) - MSE(t_2) \le 0$$

$$\lambda\left(S^2_{1R_{00}^*} - S^2_{2R}\right) + \sum_{i=1}^{N} \lambda_{2i} S^2_{2R_{00i}^*} \le 0$$

(iii) $$MSE(t_3^*) - MSE(t_3) \le 0$$

$$\lambda\left(S^2_{1P_0^*} - S^2_{1P}\right) + \sum_{i=1}^{N} \lambda_{2i} S^2_{2P_{0i}^*} \le 0$$

(iv) $$MSE(t_4^*) - MSE(t_4) \le 0$$

$$\lambda\left(S^2_{1P_{00}^*} - S^2_{2P}\right) + \sum_{i=1}^{N} \lambda_{2i} S^2_{2P_{00i}^*} \le 0$$

where

$$S^2_{1R} = \left[S_y^2 + R_1^2 S_x^2 + R_2^2 S_z^2 - 2\left(R_1 S_{xy} - R_2 S_{yz} + R_1 R_2 S_{xz}\right)\right]$$

$$S^2_{2R} = \left[S_y^2 + R_1^2 S_x^2 + R_2^2 S_z^2 - 2\left(R_1 S_{xy} + R_2 S_{yz} - R_1 R_2 S_{xz}\right)\right]$$

$$S^2_{1P} = \left[S_y^2 + R_1^2 S_x^2 + R_2^2 S_z^2 + 2\left(R_1 S_{xy} + R_2 S_{yz} + R_1 R_2 S_{xz}\right)\right]$$

$$S^2_{2P} = \left[S_y^2 + R_1^2 S_x^2 + R_2^2 S_z^2 - 2\left(R_1 S_{xy} - R_2 S_{yz} - R_1 R_2 S_{xz}\right)\right]$$

When these conditions are satisfied, the suggested method of estimation will be more efficient than Singh (1965, 1967) estimators.

6. Optimum sampling and sub-sampling fractions

In this section, we deduce the optimum sampling and sub-sampling fractions that will minimize the MSEs given by equations (8), (9), (10) and (11) for a specified cost and variance.

Consider the cost function

$$C = nc_1 + c_2 \sum_{i=1}^{n} m_i + c_3 \sum_{i=1}^{n} M_i$$

Where C is the total cost; c_1 and c_2 are the fixed cost per FSU and SSU respectively while c_3 is the cost of listing per SSU in a selected FSU.

However, the total cost C depends on the selected units which always vary from sample to sample, hence, we shall use the average cost denoted C'

$$C' = nc_1 + c_2 \frac{n}{N} \sum_{i=1}^{N} m_i + c_3 \frac{n}{N} \sum_{i=1}^{N} M_i$$

$$= n\left(c_1 + c_3 \overline{M}\right) + nc_2 \overline{m} \tag{12}$$

We shall now proceed to determine the best n and $\overline{m}$ which will minimize (8) subject to the cost constraints. To do this, MSE of t_1^* is rearranged as follows:

Let $f_{2i} = \dfrac{m_i}{M_i} = f_2$ such that $f_2 = \dfrac{\overline{m}}{\overline{M}}$ and then $m_i = \dfrac{M_i}{\overline{M}} \overline{m}$.

If we substitute the values of m_i and f_2 in (8) and rearranging, we have

$$MSE\left(t_1^*\right) = \frac{\lambda_1 S_{1R_0^*}^2 - S_{2R_0^*}^2}{n} + \frac{\overline{M} S_{2R_0^*}^2}{n\overline{m}} \tag{13}$$

Where $\lambda_1 = 1 - f$ and $S_{2R_0^*}^2 = \dfrac{1}{N} \sum_{i=1}^{N} M_i S_{2R_{0i}^*}^2$.

Using Lagrange multiplier we construct the function

$$F = \frac{\lambda_1 S_{1R_0^*}^2 - S_{2R_0^*}^2}{n} + \frac{\overline{M} S_{2R_0^*}^2}{n\overline{m}} + \theta\left[n\left(c_1 + c_3 \overline{M}\right) + c_2 n\overline{m} - C'\right]$$

Differentiating F partially with respect to n and $\overline{m}$, equating the derivatives to zero, and simplifying, we obtain the two equations below

$$\lambda_1 \overline{m} S^2_{1R_0^*} - \overline{m} S^2_{2R_0^*} + \overline{M} S^2_{2R_0^*} = \theta\left(\overline{m} c_1 n^2 + \overline{m} c_3 n^2 \overline{M} + c_2 \overline{m}^2 n^2\right) \tag{14}$$

$$\overline{M} S^2_{2R_0^*} = \theta c_2 \overline{m}^2 n^2 \tag{15}$$

Subtracting (15) from (14) and solving for θ, we obtain

$$\theta = \frac{\lambda_1 S^2_{1R_0^*} - S^2_{2R_0^*}}{c_1 n^2 + c_3 n^2 \overline{M}}$$

Substituting the value of θ in (15) and solving for $\overline{m}$, we obtain the optimum $\overline{m}$ as

$$\overline{m}_{opt} = \sqrt{\frac{S^2_{2R_0^*}\left(c_1 + c_3 \overline{M}\right)}{c_2\left(\lambda_1 S^2_{1R_0^*} - S^2_{2R_0^*}\right)}}.$$

To obtain the optimum first stage sample size, we substitute $\overline{m}_{opt}$ into (12) and solve for n, this results in

$$n_{opt} = \frac{C'}{c_1 + c_3 \overline{M} + c_2 \sqrt{\frac{S^2_{2R_0^*}\left(c_1 + c_3 \overline{M}\right)}{c_2\left(\lambda_1 S^2_{1R_0^2} - S^2_{2R_0^*}\right)}}}$$

If, however, the MSE is specified, say M_0, then $\overline{m}_{opt}$ is substituted in (13). By equating (13) to M_0 and solving for n, we get

$$n_{opt} = \frac{\left(\lambda_1 S^2_{1R_0^*} - S^2_{2R_0^*}\right)\left[c_1 + \left(c_2 + c_3\right)\overline{M}\right]}{M_0\left(c_1 + c_3 \overline{M}\right)}.$$

REMARK 2: following the approach above, it is easy to show that the optimum n and $\overline{m}$ for the estimators t_2^*, t_3^* and t_4^* would yield similar results but with $S^2_{1R_0^*}$ and $S^2_{2R_0^*}$ now being replaced by $S^2_{1R_{00}^*}$ and $S^2_{2R_{00}^*}$; $S^2_{1P_0^*}$ and $S^2_{2P_0^*}$; $S^2_{1P_{00}^*}$ and $S^2_{2P_{00}^*}$ respectively.

7. Numerical illustration and discussion

To illustrate the properties of the suggested estimators for estimating the population mean $\bar{Y}$, we consider a real data set whose detail description is given below.

One of the States in Nigeria (Kwara State) is divided into 16 Local Government Areas (LGA) and a random sample of four LGAs is selected. Within each selected LGA, a random sample of m_i (for $m_i = 3$, 5, 10 and 20; $i = 1,2,...n$) primary schools was selected from M_i. Information on the enrolment of pupils for the sessions 2008/2009 (y), 2007/2008 (x) and 2006/2007 (z) was obtained among other characteristics of interest. The summary of the data is presented below. $N = 16$, $n = 4$, $S_y^2 = 7658.25$, $S_x^2 = 4273$, $S_z^2 = 2048.667$, $S_{xy} = 5692.833$, $S_{yz} = 3941.333$, $S_{xz} = 2958$. Other details are provided in table 1.

The MSEs of the estimators t_1 through t_4 are 1408.83, 1208.42, 12087.91 and 1464.53 respectively, and that of t_1^*, t_2^*, t_3^* and t_4^* are given in Table 2.

Table 1. Summary of the data

LG	M_i	m_i	$\bar{x}_i$	$\bar{y}_i$	$\bar{z}_i$	S_{2xi}^2	S_{2yi}^2	S_{2zi}^2	S_{2xyi}	S_{2yzi}	S_{2xzi}
	K=20										
1	78	4	58.25	68.750	45.25	400.25	464.92	250.25	411.08	327.75	314.92
2	139	7	85.86	101.00	70.57	1046.48	1037.33	797.29	1033.5	884.17	894.93
3	95	5	32.00	40.600	24.00	332.50	473.30	249.50	396.50	343.25	287.25
4	55	3	95.33	109.67	83.33	424.33	364.33	241.33	393.17	295.67	319.33
	K=10										
1	78	8	53.25	62.880	42.38	297.64	311.55	235.13	301.04	268.34	261.32
2	139	14	101.5	118.14	84.93	617.35	740.75	376.23	668.69	511.55	470.5
3	95	10	45.50	54.200	35.50	311.83	394.62	223.83	347.33	289.33	261.17
4	55	6	94.50	112.83	79.83	492.30	667.77	418.57	561.50	510.97	452.10
	K=5										
1	78	16	49.060	60.130	39.44	169.66	222.52	168.66	188.53	183.94	165.84
2	139	28	97.430	113.32	81.71	384.58	553.63	384.58	493.08	438.98	413.83
3	95	20	42.900	53.850	33.00	227.26	306.56	227.26	283.51	250.79	240.89
4	55	12	106.33	122.58	92.17	413.61	563.17	413.61	519.88	455.76	454.76

Table 2. MSE for the different estimators

Estimators	MSE			
	$k = 20$	$k = 10$	$k = 5$	$k = 3$
t_1^*	86726.540	28348.6100	11569.740	1272.51
t_2^*	118966.59	38066.3400	11699.410	1096.73
t_3^*	753023.59	2702242.91	100808.02	9768.25
t_4^*	62875.180	27144.8300	8512.7200	1215.96

From table 2, we see that the MSEs of the estimators t_1^*, t_2^*, t_3^* and t_4^* is larger than that of t_1 through t_4 when $k = 20$, 10 and 5; therefore, we can say that t_1^*, t_2^*, t_3^* and t_4^* are less efficient than t_1 through t_4. However, for $k = 3$ the suggested estimator is more efficient than Singh (1965, 1967) estimators since they have smaller MSE.

This is not a surprising result, since potential sample units do tend to be more similar (show a high degree of inner homogeneity) if they are formed by putting elements, which are physically (or geographically) close to each other than if they are far apart. This argument is "almost always" true. The word "almost always" is a technical term which means that the event in question is not impossible theoretically, but that the probability of it actually happening in practice is close to zero.

Furthermore, in most practical situations, a complete list of sampling units (frame) from which drawing our sample may be impossible; even if such a frame exists, it will be uneconomical to obtain information from a sample of elements of the population scattered all over the region. Thus, to overcome these problems (absence or incomplete list; frame construction; cost consideration; organization of the survey and so on) our suggested method of estimation will be found to be more useful and rewarding than Singh's (1965, 1967) estimators.

Moreover, it could also be observed from table 2 that as k decreases i.e. increasing the sub-sampling size, the MSEs of the suggested estimators increases substantially at the rate of about 32% and 45%. Thus, we may say that the effect of increasing the sub-sampling size is justified through precision.

Finally, we can conclusively say that our suggested method of estimation is not only found to be more applicable in many situations but also more efficient than Singh's (1965, 1967) estimators when the sub-sampling size is large. It should be noted that the conclusion above on efficiency is based on the data set used for the numerical illustration.

Acknowledgement

The author is grateful to the referees for their comments and suggestions which have greatly helped in improving the manuscript.

REFERENCES

ABDUL, H. and JAVID, S. (2010): A family of ratio estimators for population mean in extreme ranked set sampling using two auxiliary variables. SORT 34 (1) January-June 2010, 45–64

ABU-DAYYEH, W.A., AHMED, M.S., AHMED, R.A., and MUTTLAK, H.A. (2003): Some estimators of a finite population mean using auxiliary information, Applied Mathematics and Computation 139: 287–298.

BANDYOPADHYAY, S. (1980): Improved ratio and product estimators. Sankhya Series C, 42(2), 45-49.

Cochran, W.G. (1963): Sampling techniques. John Wiley & Sons, New York.

DES, R. (1965): On a method of using multi-auxiliary information in sample surveys. J. Amer. Statist. Assoc. 60, 270–277

DIANA, G. and PERRI, P.F. (2007): Estimation of finite population mean using multi-auxiliary information. METRON - International Journal of Statistics, vol. LXV, n. 1, pp. 99–112.

HOSSAIN, M.I., and AHMED, M. S. (2001): A Class of Predictive Estimators in Two-Stage Sampling Using Auxiliary Information. Information and Management Sciences, Volume 12, Number 1, pp.49–55

KADILAR, C. and CINGI, H. (2004): Estimator of a population mean using two auxiliary variables in simple random sampling, International Mathematical Journal, 5, 357–367.

KADILAR, C. and CINGI, H. (2005): A new estimator using two auxiliary variables, Applied Mathematics and Computation, 162, 901–908.

KIREGYERA, B. (1980): A Chain Ratio-Type Estimator in Finite Population Double Sampling using two Auxiliary Variables. Metrika, 27: 217–223.

KIREGYERA, B. (1984): A Regression-Type Estimator using two Auxiliary Variables and Model of Double sampling from Finite Populations. Metrika, 31: 215–226.

MISHRA, G. and ROUT, K. (1997): A regression estimator in two-phase sampling in presence of two auxiliary variables, Metron 55(1–2), 177–186.

MUHAMMAD, H., NAQVI H., and MUHAMMAD, Q. S. (2010): Some New Regression Types Estimators in Two Phase Sampling. World Applied Sciences Journal 8 (7): 799–803.

MURTHY, M. N. (1964): Product method of estimation. Sankhya, A, 26, 69–74.

OLKIN, I. (1958): Multivariate ratio estimation for finite populations, Biometrika 45, 154–165.

PERRI, P. F. (2005): Combining two auxiliary variables in ratio-cum-product type estimators, Proceedings of Italian Statistical Society, Intermediate Meeting on Statistics and Environment, Messina, 21–23, 193–196.

RAO, P. S. R. S and MUDHOLKAR, G.S. (1967): Generalized multivariate estimator for the mean of finite populations, J. Amer. Statist. Assoc. 62, 1009–1012.

SAHOO, L.N. and PANDA, P. (1997): A class of estimators in two-stage samplings with varying probabilities. South African Statistics, J. 31, 151–160.

SAHOO, L.N. and PANDA, P. (1999): A class of estimators using auxiliary information in two-stage sampling. Australian and New Zealand Journal of Statistics, 41(4), 405–410

SAMIUDDIN, M. and HANIF, M. (2007): Estimation of population mean in single phase and two phase sampling with or with out additional information. Pak. J. Statist., 23 (2): 99–118.

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. 53, N 3–4, 117–125.

SINGH, H.P. and ESPEJO, M.R. (2007): Double sampling ratio-product estimator of a finite population mean in sample surveys. J. Appl. Statist., 34: 71–85.

SINGH, H. P., SINGH, R., ESPEJO, M. R., PINEDA, M. D., and NADARAJAN, S. (2005): On the efficiency of the dual to ratio-cum-product estimator. Mathematical Proceedings of the Royal Irish Academy, 105A (2), 51–56.

SINGH, M. P. (1965): On the estimation of ratio and product of the population parameters. Sankhya, B, 27, 321–328.

SINGH, M. P. (1967a): Ratio cum product method of estimation, Metrika, 12, 34–42.

SINGH, M. P. (1967b): Multivariate product method of estimation for finite populations, Journal of the Indian Society of Agricultural Statistics, 31, 375–378.

SINGH, R., CHAUHAN, P and SAWAN, N. (2007): A family of estimators for estimating population mean using known correlation coefficient in two phase sampling. Statistics in Transition, 8 (1): 89–96.

SRIVENKATARAMANA, T. (1980): A dual to ratio estimator in sample surveys. Biometrika, 67, 199–204.

SUKHATME, P.V., SUKHATME, B.V., SUKHATME, S. and ASHOK, C. (1984): Sampling theory of surveys with applications. Iowa State University Press, USA.

TAILOR R. and SHARMA, B. (2009): A Modified Ratio -Cum-Product Estimator of Finite Population Mean Using Known Coefficient of Variation and Coefficient of Kurtosis, Statistics in Transition- new series, vol. 10, pp. 15–24.

TRACY, D. S., SINGH, H. P., and SINGH, R. (1996): An alternative to the ratio-cum-product estimator in sample surveys, Journal of Statistical Planning and Inference, 53, 375–387.

UPADHYAYA, L. N., and SINGH, H. P. (1999): Use of transformed auxiliary variable in estimating the finite population mean. Biometrical Journal, 41, 627–636.

UPADHYAYA, L.N. and SINGH, H.P. (2003): A Note on the Estimation of Mean Using Auxiliary Information. Statistics in Transition, Vol. 6, No. 4, 571–575.

WOLTER, K. M. (1985): Introduction to Variance Estimation, Springer-Verlag.