Electronic Journal of Applied Statistical Analysis
EJASA (2012), Electron. J. App. Stat. Anal., Vol. 5, Issue 1, 30 – 43
e-ISSN 2070-5948, DOI 10.1285/i20705948v5n1p30

ON A MODIFIED HORVITZ-THOMPSON ESTIMATOR UNDER MIDZUNO-SEN SCHEME OF SAMPLING

R.K. Nayak[(1)], L.N. Sahoo[*] [(2)]

[(1)]*Department of Statistics, Khalikote College, Berhampur 760001, India*
[(2)]*Department of Statistics, Utkal University, Bhubaneswar 751004, India*

Received 27 October 2010; Accepted 25 August 2011
Available online 26 April 2012

Abstract: *This paper considers estimation of a finite population total using Horvitz-Thompson estimator under Midzuno-Sen scheme of sampling. Since Midzuno-Sen scheme is incapable of producing revised probabilities in certain situations, a linear transformation on the size measure is suggested to make the revised probabilities feasible. Further, using these revised probabilities, a modified Horvitz-Thompson estimator is developed employing a linear transformation on the study variable. Analytical as well as numerical studies have been undertaken in order to obtain a rough idea on the relative performance of the modified estimator*

Keywords: *Inclusion probability, joint inclusion probability, Midzuno-Sen scheme of sampling, πps sampling, super population model.*

1. Introduction

Let y_i and x_i, respectively, be the values of the survey variable y and an auxiliary variable x (used as a size measure), for the ith unit of a finite population of N units with corresponding population totals $Y = \sum_{i=1}^{N} y_i$ and $X = \sum_{i=1}^{N} x_i$. To estimate Y, assume that a sample s of n units is drawn from the population according to an unequal probability sampling without replacement scheme with π_i as the inclusion probability of ith unit, and π_{ij} as the joint inclusion probability of ith and jth units. The most commonly used estimator in this context is the Horvitz-Thompson [8] (HT) estimator defined by:

[*] E-mail: lnsahoostatuu@rediffmail.com

$$t_{HT} = \sum_{i \in s} \frac{y_i}{\pi_i},$$

with variance:

$$V(t_{HT}) = \sum_{i<j=1}^{N} \left(\pi_i \pi_j - \pi_{ij}\right)\left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2. \tag{1}$$

An unbiased estimator of $V(t_{HT})$, as suggested by Yates and Grundy [23], is given by:

$$v(t_{HT}) = \sum_i \sum_{j \in s} \frac{\pi_i \pi_j - \pi_{ij}}{\pi_{ij}}\left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2. \tag{2}$$

A sufficient condition for (2) to be always non-negative is that $\pi_{ij} < \pi_i \pi_j, i \neq j$.
It is a well known result that considerable reduction in the variance of t_{HT} can be expected if π_i 's are proportional to x_i. Such schemes are known as πps (inclusion probability proportional to size) schemes. A number of πps sampling schemes are available in the literature [*cf.*, Brewer and Hanif [3], Chaudhuri and Vos [5]]. However, Midzuno-Sen (MS) [Midzuno [9] and Sen [17]] scheme is one of the most useful schemes that greatly facilitate calculations of inclusion probabilities. The scheme with HT estimator also enjoys certain optimal properties as envisaged in Chao [4].
Under the MS scheme of sampling, the first unit in the sample s is selected with probability proportional to x and the remaining $(n-1)$ units with simple random sampling without replacement (SRSWOR) such that:

$$\pi_i = \frac{N-n}{N-1} p_i + \frac{n-1}{N-1} \tag{3}$$

and

$$\pi_{ij} = \frac{(n-1)}{(N-1)(N-2)}\left[\frac{(N-n)(x_i + x_j)}{X} + (n-2)\right], i \neq j = 1,2,\ldots,N, \tag{4}$$

where $p_i = x_i / X$ is the initial probability of selection of ith unit. The MS scheme although provides a non-negative Yates-Grundy variance estimator of the HT estimator, does not produce a πps scheme as π_i is not proportional to x_i.
In this paper, we consider a revised size measure (z) so that MS scheme becomes a πpz (π proportional to z) scheme by virtue of a linear transformation of the size measure x. Another transformed variable u is also developed from the study variable y on the basis of a regression

model. Finally, a modified HT estimator, based on u, is considered and its performance is evaluated under the MS scheme with the revised size measure z.

2. The MS Scheme with Revised Size Measure

It is true that the MS scheme of sampling yields a πps sample through revised probabilities only when:

$$X < kx_{(1)}, \tag{5}$$

where $x_{(1)}$ is the smallest size and $k = n\dfrac{N-1}{n-1}$. However, this requirement may not be fulfilled for some specified units of the sample and consequently the scheme cannot be operated. Further, it can be shown that reverse of the inequality is never true for all x. Let us now assume that (5) is not true for some $x_{(i)}$'s, where $x_{(i)}$ is the size of the unit of ith order. Then, we have an inequality of the following form:

$$kx_{(1)} < kx_{(2)} < \ldots < kx_{(m)} < X < kx_{(m+1)} < kx_{(m+2)} < \ldots < kx_{(N)}. \tag{6}$$

Let us now define a transformed size variable z with:

$$z_i = N\frac{x_i}{X} + c, i = 1,2,\ldots,N, \tag{7}$$

where c is a suitably chosen constant satisfying:

$$kz_{(1)} > Z = z_{(1)} + z_{(2)} + \ldots + z_{(N)} = N(1+c). \tag{8}$$

One can easily verify that the transformed variable z does not alter the order of $x-$ values and thus (8) holds. On the other hand, (8) is valid only when:

$$c > \frac{(n-1)N}{N-n}\left(\frac{X - kx_{(1)}}{X}\right) = c_1, \text{ say} \tag{9}$$

or equivalently,

$$c = c_1 + d, \text{ say}, \tag{10}$$

where d is a positive constant. We shall explain, in the latter part of our discussion, how one can assign a value for d. The MS scheme with objective size z has the following revised probabilities:

$$p_i' = \frac{1}{N} + f_1\left(\frac{N-1}{N-n}\right)\left(\frac{x_i - \overline{X}}{(1+c)\overline{X}}\right), i = 1,2,\ldots,N, \tag{11}$$

where $f_1 = \dfrac{n}{N}$ and $\overline{X} = \dfrac{X}{N}$. Hence, the first and second order inclusion probabilities are given by:

$$\pi_i' = f_1 \frac{x_i + c\overline{X}}{(1+c)\overline{X}} \tag{12}$$

and

$$\pi_{ij}' = f_2\left[1 + \frac{(N-1)(x_i + x_j - 2\overline{X})}{(N-2)(1+c)\overline{X}}\right], \tag{13}$$

where $f_2 = \dfrac{n(n-1)}{N(N-1)}$. The HT estimator of Y is then defined by:

$$t_{HT}' = \sum_{i \in s} \frac{y_i}{\pi_i'},$$

whose variance is given by:

$$V(t_{HT}') = \sum_{i<j=1}^{N} \left(\pi_i'\pi_j' - \pi_{ij}'\right)\left(\frac{y_i}{\pi_i'} - \frac{y_j}{\pi_j'}\right)^2. \tag{14}$$

This variance is estimated by:

$$v(t_{HT}') = \sum_i \sum_{j \in s} \frac{\pi_i'\pi_j' - \pi_{ij}'}{\pi_{ij}'}\left(\frac{y_i}{\pi_i'} - \frac{y_j}{\pi_j'}\right)^2, \tag{15}$$

and a sufficient condition for this estimate to be non-negative is that $\pi_{ij}' < \pi_i'\pi_j', i \neq j$.
It may be pointed out here that π_i' is always positive for any value of c while π_{ij}' is positive provided:

$$c > \frac{N\overline{X} - (N-1)(x_{(1)} + x_{(2)})}{(N-2)\overline{X}} = c_2, \text{ say.} \tag{16}$$

From (9) and (16), we have:

$$c_1 - c_2 = \frac{N(n-2)\{N\bar{X} - (N-1)x_{(1)}\}}{(N-2)(N-n)\bar{X}} + \left(\frac{N-1}{n-2}\right)\left(\frac{x_{(2)} - x_{(1)}}{\bar{X}}\right) + \frac{2N}{(N-1)(N-n)}, \tag{17}$$

which is always a positive quantity. This implies that π'_{ij} would be positive when the restriction (9) holds.

3. A Modified Estimator Under A Super Population Model

Let us consider the following hypothetical super population model in which:

$$y_i = \beta x_i + e_i, \tag{18}$$

with $\in(e_i / x_i) = 0$, $\in(e_i^2 / x_i) = \sigma^2 x_i^g$ for all i, and $\in(e_i e_j / x_i, x_j) = 0$, for $i \neq j$, such that $\beta > 0, 0 < \sigma^2 < \infty, 0 \leq g \leq 2$ and $\in(\cdot/\cdot)$ denotes the conditional expectation with respect to the distribution of e_i for fixed x_i.

One can easily verify from (14) that $V(t'_{HT})$ approaches zero if y_i / π'_i are nearly the same. But y_i / π'_i is not a constant even if $y_i = \beta x_i$, for all i. However, under such a situation $(y_i + \beta c\bar{X}) / \pi'_i$ is seen to be a constant. This leads to consider a transformed survey variable u defined by:

$u_i = y_i + bc, i = 1,2,\ldots,N,$

where b is a scalar to be chosen. Accordingly, we define a modified estimator as:

$$t_M = \sum_{i \in s} \frac{u_i}{\pi'_i} - Nbc,$$

with variance:

$$V(t_M) = V(t'_{HT}) + (bc)^2 d_1 + 2bcd_2, \tag{19}$$

where $d_1 = V\left(\sum_{i \in s} \frac{1}{\pi'_i}\right) = \sum_{i<j=1}^{N} (\pi'_i \pi'_j - \pi'_{ij})\left(\frac{1}{\pi'_i} - \frac{1}{\pi'_j}\right)^2$ and

$$d_2 = Cov\left(\sum_{i \in s} \frac{1}{\pi'_i}, \sum_{i \in s} \frac{y_i}{\pi'_i}\right) = \sum_{i<j=1}^{N} (\pi'_i \pi'_j - \pi'_{ij})\left(\frac{1}{\pi'_i} - \frac{1}{\pi'_j}\right)\left(\frac{y_i}{\pi'_i} - \frac{y_j}{\pi'_j}\right).$$

Now, we find that $V(t_M)$ is minimized for:

$$b = -d_2 / cd_1 = b_{opt}\text{, say,} \tag{20}$$

and the resulting minimum value of $V(t_M)$ is given by:

$$\min V(t_M) = V(t'_{HT}) - \frac{d_2^2}{d_1} = V(t_M)_{opt}, \text{ say.} \tag{21}$$

Hence, t_M is always more efficient than t'_{HT} under optimal situations. However, from (19) we see that $V(t_M) < V(t'_{HT})$ when ever $b < 2b_{opt}$.

3.1 Choices of b and c

Determination of the value of c (or ultimately the value of d) is of considerable important not only in computing revised probabilities π'_i and π'_{ij} but also in predicting efficiencies of t'_{HT} and t_M. But, the values of c or d are not pre-assigned. They have to be chosen according to the values of x_i's. Since there is no upper limit for the value of c, efficiencies of the proposed estimators cannot be predicted. However, to determine a reasonable value of d, we rewrite (20) as

$$b_{opt} = -\left(\frac{d_2}{d_1 c_1}\right)\left(1 + \frac{d}{c_1}\right)^{-1}. \tag{22}$$

The second factor in the *r.h.s.* of (22) is however expandable when $|d/c_1| < 1$. Thus, a permissible (not strictly) range of d is $(0, c_1)$.

Since all $y_i, i = 1,2,\ldots,N$ are unknown, an exact value of b_{opt} cannot be achieved. The following are some simple ways to get a reasonable value of b:

(i) When $y_i = \beta x_i \forall i$, we have $b_{opt} = \beta\overline{X}$ and a suitable value of b lies in $(0, 2\hat{\beta}\overline{X})$, where $\hat{\beta}$ is estimated by the slope of the line of best fit to the points $\left(y_i/\pi'_i, y_j/\pi'_j\right), i = 1,2,\ldots,n$.

(ii) b can also be estimated by

$$\hat{b} = \frac{\hat{d}_2}{\hat{d}_1 c}, \tag{23}$$

where $\hat{d}_1$ and $\hat{d}_2$ are respectively sample estimates of d_1 and d_2 given by:

$$\hat{d}_1 = \sum_i \sum_{j \in s} \left(\frac{\pi'_i \pi'_j - \pi'_{ij}}{\pi'_{ij}}\right)\left(\frac{y_i}{\pi'_i} - \frac{y_j}{\pi'_j}\right)^2, \text{ and } \hat{d}_2 = \sum_i \sum_{j \in s} \left(\frac{\pi'_i \pi'_j - \pi'_{ij}}{\pi'_{ij}}\right)\left(\frac{1}{\pi'_i} - \frac{1}{\pi'_j}\right)\left(\frac{y_i}{\pi'_i} - \frac{y_j}{\pi'_j}\right).$$

We see that under the condition $\pi'_{ij} < \pi'_i \pi'_j$, $\hat{d}_1$ is always non-negative whereas $\hat{d}_2$ a may not be so. Hence, the estimated value of $V(t_M)$ given by:

$$v(t_M) = v(t'_{HT}) + \left(\hat{b}c\right)^2 \hat{d}_1 + 2\hat{b}c\hat{d}_2, \tag{24}$$

may have chance of achieving negative values.

3.2 Sensitivity of the Efficiency of t_M for b

It is seen from (19) that the variance of t_M depends on the value of b, and is symmetrical about and minimum for $b = b_{opt}$. This means that t_M will have the same amount of efficiency loss whether b is an overestimation or an under estimation of b_{opt} by a certain amount. Therefore, it is our interest to examine how the efficiency of t_M with respect to t'_{HT} would change when b has a value other than b_{opt}.

Keeping in view $V(t_M) < V(t'_{HT})$, let us suppose that $b = pb_{opt}$ where $0 < p < 2$. Then substituting this value of b in (19) and using value of b_{opt} from (20), after simplification, we get:

$$V(t_M) = V(t'_{HT}) + \frac{p^2 d_2^2}{d_1} - \frac{2pd_2^2}{d_1} = V(t_M)_{opt} + (p-1)^2 \frac{d_2^2}{d_1}. \tag{25}$$

The second term in the *r.h.s.* of (25) is the amount of increase over the optimal variance when b deviates from b_{opt} by $(p-1)b_{opt}$, the sign of which indicates the direction of deviation. Obviously, when $p = 0$ or 2, $V(t_M) = V(t'_{HT})$ and when $p = 1$, $V(t_M) = V(t_M)_{opt}$.

Here sensitivity of the efficiency of t_M for varying values of b or alternatively the values of p is measured by the relative efficiency (RE) of t_M with respect to t'_{HT} being given by:

$$RE = \frac{V(t'_{HT})}{V(t_M)} = \frac{V(t'_{HT})}{V(t_M)_{opt} + (p-1)^2 d_2^2 / d_1}. \tag{26}$$

Nayak [10] evaluated numerical values of this RE of t_M using live data of a number of populations and investigated that it increases with the increase in the values of p.

4. Variance Comparison Under the Model

Consider the super population model (18) with an additional assumption that x_i's are *i.i.d.* gamma variates with a common parameter h $(h > 0)$ equal to the known value $\overline{X}$. [*cf.*, Durbin [7], Tin [21], Rao and Webster [14], Rao and Rao [15] among others]. With an objective of evaluating efficiency of the suggested estimation method under the model, we consider the following sampling strategies:

S_{HT} : t_{HT} under the usual HT scheme of sampling using x

S_1 : t_{HT} under the MS scheme of sampling using x

S_2 : t'_{HT} under the MS scheme of sampling using z

S_3 : t_M under the MS scheme of sampling using z

Using the concepts developed in Rao and Webster [14] and Arnab [1], we obtain the following model-based variance expressions of different strategies (details are omitted to save space):

$$\in V(S_{HT}) = \sigma^2 \frac{N}{n} H\left[(N-n)h-(n-1)(g-1)\right] \tag{27}$$

$$\in V(S_1) = \frac{N-1}{(n-1)^2}\sigma^2 \frac{H}{h}(g+h-1)\left[(Nn-2N+n)h-(N-n)g\right] \tag{28}$$

$$\in V(S_2) = \left(\frac{\sigma^2}{gh}\right)\frac{N}{n} H\left[(N-n)h-(n-1)(g-1)+\Delta_1+\Delta_2\right] \tag{29}$$

$$\in V(S_3) = \in V(S_2)+(bc)^2\left[\frac{N^2h(1-f_1-c^2)-N(N-n)-N^2c}{n(h-1)}\right] \tag{30}$$

where $\in$denotes expectation operator *w.r.t.* the super population, $H = \dfrac{\Gamma(g+h-2)}{\Gamma(h-1)}$,

$\Delta_1 = \dfrac{(ch)(N-n)g(N-1)(hg-2)}{(g+h-2)}$ and $\Delta_2 = \dfrac{(ch)^2(g-1)\{2(N-1)-(n-1)g\}}{2(g+h-2)}$.

Comparisons of the above expressions provide us with the following results:

(i) $\in V(S_2) < \in V(S_1)$, when $g > \dfrac{1}{2}, hg < 2$ and $c < \left(\dfrac{N-n}{N-1}\right)\left(\dfrac{g}{1-g}\right)\dfrac{1}{h}$.

(ii) $\in V(S_3) < \in V(S_2)$, when $c > \dfrac{h+1}{3}$ and $h > 1$.

(iii) $\in V(S_1) < \in V(S_{HT})$, when $g < 1$ and $h < \dfrac{(N-1)(N-n)}{(n^2-N)(n-1)}, n^2 > N$.

Finally, after combining these three results, we have $\in V(S_3) < \in V(S_2) < \in V(S_1) < \in V(S_{HT})$, when:

$$\frac{1}{2} < g < 1, 1 < h < \frac{2}{g} \text{ and } \frac{h+1}{3} < c < \left(\frac{N-n}{N-1}\right)\left(\frac{g}{1-g}\right)\frac{1}{h}. \tag{31}$$

Here it should be made clear that the achieved conditions are only sufficient conditions. It is however difficult to find out the necessary conditions. Also, one can see that the restrictions imposed on c in (31) are satisfied when h is marginally above unity and g is very near to unity from the left. It is therefore observed from (31) that there is enough scope of improving upon the HT estimator through our method.

5. Illustration of the Suggested Methodology

In order to illustrate how the suggested methodology operates, we consider data of three small populations as given in table 1.

Table 1. Three Small Populations.

Pop. No.	Source	y	x	Values of (y, x)
1	Srivenkataramana and Srinath [18]	artificial	artificial	(9,1), (13,2), (15,3), (15,4), (15,5)
2	Ray *et al.* [16]	artificial	artificial	(7,1), (8,3), (10,4), (11,6), (11,7), (13,9)
3	Sukhatme *et.al.* [20], p. 201	mean no. of live stock per village	mean agric. area per village	(25.4,63.7), (50.1,155.3), (76.0,245.7), (99.2,344.4), (150.8,491.6), (244.4,767.5), (425.1,1604.0)

Considering $n = 3$, it is seen that the MS scheme cannot be applied in a straightforward manner to these populations with revised selection probability obtained through x. Using (7), let us calculate the values of the transformed size variable z, for $d = 0.25c_1, 0.50c_1$ and $0.75c_1$ (or $c = 1.25c_1$, $1.50c_1$ and $1.75c_1$), where $c_1 = 3.0, 3.0$ and 2.9536 respectively for populations 1, 2 and 3. Table 2 shows $z-$values for different values of c.

Table 2. Values of z for Different Values of c.

Pop. No.	c	Values of z	Z
1	1.25 c_1	4.0833, 4.4167, 4.7500, 5.0833, 5.4167	23.7500
	1.50 c_1	4.8333, 5.1667, 5.5000, 5.8333, 6.1667	27.5000
	1.75 c_1	5.8333, 5.9167, 6.2500, 6.5833, 6.9167	31.2500
2	1.25 c_1	3.95, 4.35, 4.55, 4.95, 5.15, 5.55	28.50
	1.50 c_1	4.70, 5.10, 5.30, 5.70, 5.90, 6.30	33.00
	1.75 c_1	5.45, 5.85, 6.05, 6.45, 6.65, 7.05	37.50
3	1.25 c_1	3.8134, 3.9880, 4.1603, 4.3485, 4.6291, 5.1550, 6.7495	32.8438
	1.50 c_1	4.5518, 4.7264, 4.8987, 5.0869, 5.3675, 5.8934, 7.4879	38.0126
	1.75 c_1	5.2902, 5.4648, 5.6371, 5.8253, 6.1059, 6.6318, 8.2263	43.1814

We note that the $z-$values in respect of three populations satisfy the condition (8). Thus, the generated sets of $z-$values can be safely utilized for implementation of the MS scheme under revised probability of selection to compose t'_{HT}. Finally, one can compose the modified estimator t_M with b obtainable from (23).

6. Numerical Study of the Efficiency of the Proposed Strategy

A desirable goal is to study the performance of the proposed methodology. But, we see that a direct theoretical evaluation in this regard is not possible. In section 4, although some results have been derived concerning efficiency of the proposed strategy S_3 over some other strategies under a super population model, it is not very clear how far one can fulfill the sufficient conditions (31) in order to draw a conclusion. However, as a counter part to the theoretical comparison, here we carry out a numerical study by considering 12 natural populations from Rao and Bayless [13] and Bayless and Rao [2] as described in table 3. The following performance measures of the comparable strategies S_{HT}, S_1, S_2 and S_3 were taken into consideration:

(i) Percentage relative efficiency (PRE) with respect to the strategy S_0 which involves mean per unit estimator $N\sum_{i\in s} y_i / n$ under simple random sampling without replacement (SRSWOR).

(ii) Percentage of non-negative variance estimators (PNNVE). This performance measure gives us an idea about the non-negativeness property of the variance estimator *i.e.*, number of times one can get negative estimates of the variance of a strategy.

Table 3. Description of the Populations.

Pop. No.	Source	N	y	x
1	Horvitz and Thomson [8]	20	no. of house holds	eye estimated no. of house holds
2	Des Raj [11] (modification of 1)	20	no. of house holds	eye estimated no. of house holds
3	Rao [12]	14	corn acreage in 1960	corn acreage in1958
4	Cochran [6] p.152 (1-20)	20	no. of people in 1930	no. of people in 1920
5	Cochran [6] p.152 (21-40)	20	no. of people in 1930	no. of people in 1920
6	Sukhatme and Sukhatme [19] p.185	20	wheat acreage in 1937	wheat acreage in1936
7	Sukhatme *et al.* [20], p.297	20	wheat acreage	no. of villages
8	Yates [22] p.163	20	volume of timber	eye estimated volume of timber
9	Yates [22] p.159	20	no. of absentees	total no. of persons
10	Cochran [6] p.152 (1-16)	16	no. of people in 1930	no. of people in 1920
11	Cochran [6] p.152 (17-32)	16	no. of people in 1930	no. of people in 1920
12	Cochran [6] p.152 (33-49)	17	no. of people in 1930	no. of people in 1920

Numerical values of PRE and PNNVE of the comparable strategies are computed for $n = 2,3,4$ and 5. Our computations are based on all $C(N,n)$ possible samples when $C(N,n) \le 2000$ and on 2000 independent samples when $C(N,n) > 2000$. To calculate PRE of S_{HT}, we consider $\pi_i = np_i, \forall i$.

Table 4. PRE of Different Strategies.

Pop. No.	n	S_{HT}	S_1	S_2			S_3		
				$c = 1.25c_1$	$c = 1.50c_1$	$c = 1.75c_1$	$c = 1.25c_1$	$c = 1.50c_1$	$c = 1.75c_1$
1	2	529.62	253.04	493.05	478.52	463.58	521.54	518.87	516.28
	3	547.69	179.50	270.61	245.62	222.62	482.86	476.09	470.33
	4	549.24	151.06	197.25	181.16	169.52	462.55	455.60	450.30
	5	183.16	136.58	165.11	153.95	146.02	450.37	444.14	439.30
2	2	183.10	164.81	190.26	190.23	189.95	190.49	190.25	190.01
	3	184.70	141.56	167.92	162.25	157.27	186.77	186.08	185.48
	4	185.74	129.24	147.59	141.87	137.31	184.67	183.93	183.33
	5	179.43	121.84	135.03	130.21	126.52	183.35	182.66	182.11
3	2	110.17	78.52	55.34	58.89	61.94	96.95	97.50	97.97
	3	70.96	90.83	82.78	85.48	87.58	100.10	100.38	100.56
	4	123.59	95.39	91.69	93.31	94.48	100.77	100.87	100.91
	5	21.27	97.52	95.64	96.58	97.23	100.96	100.98	100.99
4	2	3629.82	656.50	525.79	433.83	373.44	4700.14	4803.71	4881.08
	3	3107.42	346.08	287.76	248.57	222.72	5022.34	5081.89	5124.29
	4	4269.73	253.08	217.08	193.12	177.23	5139.54	5179.06	5206.40
	5	284.54	208.82	183.46	166.60	155.38	5197.97	5226.62	5246.48
5	2	586.18	488.65	622.52	602.89	574.21	647.86	663.66	667.00
	3	604.19	304.29	396.15	346.12	308.75	727.34	740.60	750.84
	4	589.96	230.70	278.77	245.01	221.47	758.78	768.95	776.53
	5	231.36	193.58	223.08	199.32	183.05	775.40	783.46	789.37
6	2	2689.41	379.71	324.01	280.29	250.56	4437.37	4378.24	4323.68
	3	2473.28	229.29	202.28	182.49	169.05	4198.90	4127.99	4069.98
	4	2479.81	180.76	163.76	151.52	143.20	4053.47	3988.15	3936.75
	5	497.23	156.96	144.96	136.39	130.54	3959.77	3901.20	3856.50
7	2	152.58	155.54	164.15	161.94	159.51	166.95	166.77	166.50
	3	161.29	138.51	145.81	141.22	137.36	164.71	163.89	163.17
	4	160.82	128.37	133.39	129.67	125.85	162.67	161.78	161.05
	5	170.11	121.95	125.57	121.98	119.27	161.26	160.41	159.74
8	2	238.74	222.96	224.16	212.04	201.45	252.67	251.15	294.77
	3	238.07	174.75	171.87	161.46	153.56	246.13	244.35	242.90
	4	241.47	150.81	147.68	140.14	134.63	242.33	240.70	239.43
	5	118.01	137.24	134.49	128.85	124.77	239.98	238.47	237.36
9	2	223.06	190.59	222.42	218.38	213.83	235.10	237.08	238.69
	3	244.78	156.11	179.27	170.07	162.56	244.60	245.51	246.08
	4	241.93	138.87	153.55	145.87	140.04	246.54	246.80	246.87
	5	239.42	128.87	138.95	123.89	128.43	246.95	246.90	246.77
10	2	117.43	69.65	67.20	69.70	71.85	77.82	80.45	82.60
	3	119.56	80.53	79.78	81.96	83.72	88.11	90.22	91.84
	4	124.33	28.31	85.71	87.47	88.84	92.61	94.28	95.54
	5	108.15	94.39	110.55	113.36	115.08	124.19	136.41	138.99
11	2	1856.42	823.11	676.82	560.71	481.43	2476.82	2520.87	2553.97
	3	1963.67	439.54	364.17	310.30	274.32	2614.33	2640.33	2659.02
	4	2025.34	329.65	273.48	232.59	205.38	2744.77	2772.31	2791.02
	5	2082.51	263.72	204.19	185.46	164.22	2798.92	2855.05	2891.13
12	2	1142.79	557.50	834.56	714.90	619.38	1355.82	1402.94	1441.35
	3	1329.41	305.00	377.74	318.72	279.35	1556.98	1590.28	1614.94
	4	1536.51	225.06	256.26	223.13	201.30	1630.50	1653.61	1670.07
	5	106.44	187.04	204.36	182.55	168.16	1666.85	1683.82	1695.66

An examination of the results on the PRE of the strategies presented in table 4 reveals that the proposed strategy S_3 is decidedly more efficient than others. Although, S_2 appears to be better than S_1 in many cases, the overall performance of S_{HT} seems to be better than that of S_1 or S_2. Hence, our findings give an indication that our improvement over HT estimator under MS scheme of sampling is only marginal when transformation is considered on size variable only. But, this improvement is considerable when transformations on both variables are taken into account.

PNNVE – values of the strategies S_1, S_2 and S_3 are displayed in table 5 for $n = 2$ and 3. Although, in respect of this measure the strategies show inconsistent results, there is some indication of better performance of S_3 than both S_1 and S_2, and slightly better performance of S_2 than S_1. Results for $n = 4$ and 5 are not given here because the strategies for these cases behave very erratically and there is no clear indication that which one of them would have a decidedly better overall performance compared to others. This is probably because of the complexity in expressions for joint inclusion probabilities as well as variance estimator with enlargement of sample size.

Table 5. PNNVE of Different Strategies.

Pop. No.	n	S_1	S_2			S_3		
			$c = 1.25c_1$	$c = 1.50c_1$	$c = 1.75c_1$	$c = 1.25c_1$	$c = 1.50c_1$	$c = 1.75c_1$
1	2	49.23	53.45	51.23	45.15	55.63	54.81	48.92
	3	65.85	58.77	57.24	50.36	52.13	49.14	47.48
2	2	51.43	45.47	47.35	52.41	50.29	55.13	59.22
	3	54.10	55.91	56.17	59.45	54.55	58.29	61.27
3	2	38.69	45.84	49.73	49.95	47.35	50.54	51.41
	3	47.95	41.68	42.55	47.81	43.83	44.61	48.25
4	2	31.23	50.42	49.32	48.35	53.97	49.83	47.87
	3	50.51	54.51	51.89	50.50	54.65	53.61	52.75
5	2	25.74	45.11	47.23	51.33	44.74	48.65	53.47
	3	43.39	41.18	49.31	50.81	40.31	50.59	58.33
6	2	33.31	38.23	37.45	39.25	36.33	36.61	39.90
	3	44.49	33.48	39.39	45.24	40.51	44.75	45.83
7	2	69.51	55.51	56.24	60.35	58.18	60.34	68.53
	3	55.53	50.63	54.11	58.46	54.19	55.35	56.58
8	2	42.85	48.74	46.28	43.71	47.48	45.31	40.47
	3	51.75	55.45	54.34	55.76	53.85	59.36	61.51
9	2	34.73	41.11	47.46	54.50	45.55	47.44	54.38
	3	39.61	35.15	37.51	40.67	36.63	38.12	39.22
10	2	57.44	51.38	58.55	61.55	58.74	58.51	63.81
	3	61.11	60.39	61.90	62.18	59.65	63.55	65.79
11	2	43.27	41.91	41.85	44.21	40.34	44.14	46.23
	3	49.74	43.71	49.87	51.25	47.49	49.27	55.85
12	2	51.81	52.30	51.73	55.43	48.90	56.44	54.53
	3	43.67	48.25	47.23	46.82	46.70	48.31	49.48

7. Conclusions

On the basis of the analytical and empirical results derived in this work, we may conclude that our suggested estimation procedure may be better than the Horvitz-Thompson method of estimation under MS scheme of sampling in many situations. No general conclusion can be drawn from the numerical study as the conclusion is based on the results of 12 populations only. However, it shows that our method cannot be inferior to its competing methods.

Acknowledgement

The authors are grateful to Dr. J. Sahoo for his suggestions and guidance during the preparation of the paper and to the referees whose constructive comments led to an improvement in the paper.

References

[1]. Arnab, R. (1979). On the relative efficiencies of some sampling strategies under a super-population model. *Jour. Indian Soc. Agric. Stat.*, 31, 89-96.
[2]. Bayless, D. L. and Rao, J. N. K. (1970). An empirical study of the stabilities of estimators and variance estimators in unequal probability sampling ($n = 3 \text{ or } 4$). *Jour. Amer. Stat. Assoc*., 65, 1645-1667.
[3]. Brewer, K. R. W. and Hanif, M. (1983). *Sampling with Unequal Probabilities*. Springer–Verlag.
[4]. Chao, M.T. (1982). A general purpose unequal probability sampling plan. *Biometrika,* 69, 653-656.
[5]. Chaudhuri, A. and Vos, J. E. W. (1988). *Unified Theory and Strategies of Survey Sampling*. North Holland.
[6]. Cochran, W. G. (1977). *Sampling Techniques* (3rd Edition). Wiley Eastern Limited.
[7]. Durbin, J. (1959). A note on the application of Quenouille's method of bias reduction in the estimation of ratios. *Biometrika*, 46, 477-480.
[8]. Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Jour. Amer. Stat. Assoc*., 47, 663-685.
[9]. Midzuno, H. (1952). On the sampling system with probability proportionate to sum of sizes. *Ann. Inst. Stat. Math*., 3, 99-107.
[10]. Nayak, R. K. (2009). *Some Estimation Strategies in Finite Population Survey Sampling Using Auxiliary Information*. Unpublished Ph. D. Thesis, Utkal University, Bhubaneswar, India.
[11]. Raj, Des. (1965). Variance estimation in randomized systematic sampling with probability proportional to size. *Jour. Amer. Stat. Assoc*., 60, 278-284.
[12]. Rao, J. N. K. (1963). On the three procedures of unequal probability sampling without replacement. *Jour. Amer. Stat. Assoc*., 58, 202-215.
[13]. Rao, J. N. K. and Bayless, D. L. (1969). An empirical study of the stabilities of estimators and estimators in unequal probability sampling of two units per stratum. *Jour. Amer. Stat. Assoc*., 64, 540-549.
[14]. Rao, J. N. K. and Webstar, J. T. (1966). On two methods of bias reduction in the estimation of ratios. *Biometrika*, 53, 571-577.
[15]. Rao, P. S. R. S. and Rao, J. N. K. (1971). Small sample results for ratio estimators. *Biometrika*, 58, 625-630.
[16]. Ray, S.K., Sahai, Ajit and Sahai, Ashok (1979). A note on ratio and product type estimators. *Ann. Inst. Stat. Math*., 31, 141-144.
[17]. Sen, A. R. (1953). Present status of probability sampling and its use in estimation of farm characteristics (abstract). *Econometrica*, 20, 103.

[18]. Srivenkataramana, T. and Srinath, K. P. (1982). Change of scale in different estimator in sampling. *Jour. Indian Soc. Agric. Statist*. 34, 51-60
[19]. Sukhatme, P. V. and Sukhatme, B. V. (1977). *Sampling Theory of Surveys with Applications*. Asia Publishing House, New Delhi.
[20]. Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S. and Asok, C. (1984). *Sampling Theory of Survey with Applications.* Iowa State University Press, Ames, Iowa.
[21]. Tin, M. (1965). Comparison of some ratio estimators. *Jour. Amer. Stat. Assoc*., 60, 294-307.
[22]. Yates, F. (1960). *Sampling Methods for Censuses and Surveys*. Charls & Griffin, London.
[23]. Yates, F. and Grundy, P. M. (1953). Selection without replacement from within strata with probability proportional to size. *Jour. Roy. Stat. Soc*., B15, 253-261.