J. Ind. Soc. Agril. Statist.
58(2), 2004 : 163-176

A Family of Unbiased Estimators in Two Stage Sampling

Housila P. Singh and Rajesh Tailor
Vikram University, Ujjain–456006, M.P.
(Received : March, 2002)

SUMMARY

This paper proposes a family of unbiased estimators for population mean in two-stage sampling. It has been shown that the unbiased estimator cited in Sukhatme and Sukhatme (1970) is a particular case of the proposed family. Asymptotically optimum unbiased estimator (AOUE) in the class is identified with its variance formula. An estimator based on estimated optimum value is also obtained. It has been shown to the first degree of approximation that the variance of the estimator based on "estimated optimum value" is the same as that of AOUE. An empirical study is carried out to demonstrate the performance of the constructed estimator over other estimators.

Key words : Finite population mean, Family of unbiased estimators, Simple random sampling without replacement, Variance.

1. Introduction

Assume that the population is composed of $\sum_{i=1}^{N} M_i = N\overline{M}$ elements grouped into N first–stage units (fsu's) of $M_i, (i = 1, 2, \ldots, N)$ second–stage units (ssu's) in the ith first–stage unit. Let n denote the number of first–stage units in the sample and m_i $(i = 1, 2, \ldots, n)$ the number of second–stage units to be selected from the ith first stage unit, if it is in the sample. It is assumed that the units at each stage are selected with simple random sampling without replacement (SRSWOR) scheme. Let y_{ij} be the values of the j^{th} second-stage unit in the i^{th} first stage unit (j = 1, 2, ..., M_i; i = 1, 2, ... N). Then $\overline{Y}_i = \frac{1}{M_i}\sum_{j=1}^{M_i} y_{ij}$, the mean per second-stage unit in the i^{th} first-stage unit in the population $(i = 1, 2, \ldots, N)$, $\overline{Y} = \frac{1}{M_0}\sum_{i=1}^{N}\sum_{j=1}^{M_i} y_{ij} = \frac{1}{N}\sum_{i=1}^{N} u_i \overline{Y}_i =$ the mean per second–stage unit in the population, $u_i = \frac{M_i}{\overline{M}}$, $M_0 = \sum_{i=1}^{N} M_i$, and

$\overline{y}_i = \frac{1}{m_i}\sum_{j=1}^{m_i} y_{ij}$ = the mean per second–stage unit of the i^{th} first stage unit in the sample.

To estimate the population mean $\overline{Y}$, the well known estimators are

$$\overline{y}_{s2} = \frac{1}{n}\sum_{i=1}^{n}\overline{y}_i \tag{1.1}$$

$$\overline{y}'_{s2} = \frac{1}{n}\sum_{i=1}^{n}u_i\overline{y}_i \tag{1.2}$$

$$\overline{y}''_{s2} = \frac{1}{n\overline{M}_n}\sum_{i=1}^{n}M_i\overline{y}_i = \frac{1}{n\overline{u}_n}\sum_{i=1}^{n}u_i\overline{y}_i = \frac{\overline{y}'_{s2}}{\overline{u}_n} \tag{1.3}$$

and

$$\overline{y}'''_{s2} = \overline{y}_{s2} + \frac{(N-1)}{N\overline{M}}\cdot\frac{1}{(n-1)}\sum_{i=1}^{n}(M_i - \overline{M}_n)(\overline{y}_i - \overline{y}_{s2}) \tag{1.4}$$

or equivalently

$$\overline{y}'''_{s2} = \overline{y}_{s2} + \frac{(N-1)}{N}\cdot\frac{n}{(n-1)}(\overline{y}'_{s2} - \overline{u}_n\overline{y}_{s2}) \tag{1.5}$$

where $\overline{M}_n = \sum_{i=1}^{n} M_i / n$ and $\overline{u}_n = \frac{1}{n}\sum_{i=1}^{n}u_i$

It is to be mentioned that the estimators $\overline{y}'_{s2}$ and $\overline{y}'''_{s2}$ are unbiased and the estimators $\overline{y}_{s2}$ and $\overline{y}''_{s2}$ are biased.

The variances of $\overline{y}_{s2}$ and $\overline{y}'_{s2}$ are respectively given by

$$V(\overline{y}_{s2}) = \frac{(1-f)}{n}S_{by}^2 + \frac{1}{nN}\sum_{i=1}^{N}\frac{(1-f_i)}{m_i}S_{iy}^2 \tag{1.6}$$

$$V(\overline{y}'_{s2}) = \frac{(1-f)}{n}S_{by}'^2 + \frac{1}{nN}\sum_{i=1}^{N}u_i^2\frac{(1-f_i)}{m_i}S_{iy}^2 \tag{1.7}$$

where $f = \frac{n}{N}$ and $f_i = \frac{m_i}{M_i}$

$$S_{by}^2 = \frac{1}{(N-1)} \sum_{i=1}^{N} (\overline{Y}_i - \overline{\overline{Y}})^2 \tag{1.8}$$

$$S_{by}'^2 = \frac{1}{(N-1)} \sum_{i=1}^{N} (u_i \overline{Y}_i - \overline{Y})^2 \tag{1.9}$$

$$S_{iy}^2 = \frac{1}{(M_i - 1)} \sum_{j=1}^{M_i} (Y_{ij} - \overline{Y}_i)^2 \tag{1.10}$$

and $$\overline{\overline{Y}} = \frac{1}{N} \sum_{i=1}^{N} \overline{Y}_i \tag{1.11}$$

As the estimator $\overline{y}_{s2}$ is a biased estimator of $\overline{Y}$, therefore, we write its mean square error (MSE) for the comparison purpose as

$$MSE(\overline{y}_{s2}) = \frac{(1-f)}{n} S_{by}^2 + \frac{1}{nN} \sum_{i=1}^{N} \frac{(1-f_i)}{m_i} S_{iy}^2 + (\overline{\overline{Y}} - \overline{Y})^2 \tag{1.12}$$

To the first degree of approximation, the variance of $\overline{y}''_{s2}$ and $\overline{y}'''_{s2}$ are respectively given by

$$V(\overline{v}''_{s2}) = \frac{(1-f)}{n} S_{by}''^2 + \frac{1}{nN} \sum_{i=1}^{N} u_i \frac{(1-f_i)}{m_i} S_{iy}^2 \tag{1.13}$$

and

$$V(\overline{y}'''_{s2}) = \frac{(1-f)}{n} S_{by}'''^2 + \frac{1}{nN} \sum_{i=1}^{N} u_i^2 \frac{(1-f_i)}{m_i} S_{iy}^2 \tag{1.14}$$

where

$$S_{by}''^2 = \frac{1}{(N-1)} \sum_{i=1}^{N} u_i^2 (\overline{Y}_i - \overline{Y})^2 \tag{1.15}$$

$$S_{by}'''^2 = \frac{1}{(N-1)} \sum_{i=1}^{N} \left\{ (u_i \overline{Y}_i - \overline{Y}) - \overline{\overline{Y}}(u_i - 1) \right\}^2 \tag{1.16}$$

2. The Suggested Family of Unbiased Estimators

Taking the linear combination of $\overline{y}'_{s2}, \overline{y}_{s2}$ and $\overline{u}_n \overline{y}_{s2}$, we propose the following class of estimators for $\overline{Y}$

$$\hat{\bar{Y}} = \theta_1 \bar{y}'_{s2} + \theta_2 \bar{y}_{s2} + \theta_3 \bar{u}_n \bar{y}_{s2} \tag{2.1}$$

where θ_i's $(i = 1, 2, 3)$ are suitably chosen scalars such that their sum is unity i.e.

$$\theta_1 + \theta_2 + \theta_3 = 1 \tag{2.2}$$

The estimator $\hat{\bar{Y}}$ would be unbiased if

$$B(\hat{\bar{Y}}) = \theta_2 \, B(\bar{y}_{s2}) + \theta_3 B(\bar{u}_n \bar{y}_{s2}) = 0$$

or if
$$\theta_3 = -\theta_2 \frac{B(\bar{y}_{s2})}{B(\bar{u}_n \bar{y}_{s2})} \tag{2.3}$$

The exact expressions for biases of $\bar{y}_{s2}$ and $\bar{u}_n \bar{y}_{s2}$ are respectively given by

$$B(\bar{y}_{s2}) = -\frac{(N-1)}{N\bar{\bar{M}}} \cdot \frac{1}{(N-1)} \sum_{i=1}^{N} (M_i - \bar{M})(\bar{Y}_i - \bar{\bar{Y}}) \tag{2.4}$$

and

$$B(\bar{u}_n \, \bar{y}_{s2}) = -\frac{(n-1)}{n\bar{M}} \cdot \frac{1}{(N-1)} \sum_{i=1}^{N} (M_i - \bar{M})(\bar{Y}_i - \bar{\bar{Y}}) \tag{2.5}$$

If follows from (2.4) and (2.5) that

$$\frac{B(\bar{y}_{s2})}{B(\bar{u}_n \bar{y}_{s2})} = \frac{(N-1)n}{(n-1)N} = \eta \qquad \text{(say)} \tag{2.6}$$

Using (2.6) in (2.3) we obtain

$$\theta_3 = -\theta_2 \eta \tag{2.7}$$

Putting $\theta_3 = -\theta_2 \eta$ in (2.2) we get

$$\theta_1 = \{1 - \theta_2 (1 - \eta)\} \tag{2.8}$$

Thus putting $\theta_1 = \{1 - \theta_2(1-\eta)\}$ and $\theta_3 = -\theta_2 \eta$ with $\theta_2 = \theta$ (a constant) we get a general class of unbiased estimators for $\bar{Y}$ as

$$\hat{\bar{Y}}_u = [\{1 - \theta(1-\eta)\} \bar{y}'_{s2} + \theta \bar{y}_{s2} - \theta \eta \bar{u}_n \bar{y}_{s2}]$$

$$= \left[\left[1 - \theta\left\{1 - \frac{(N-1)n}{(n-1)N}\right\}\right] \bar{y}'_{s2} + \theta \bar{y}_{s2} - \theta \frac{(N-1)n}{(n-1)N} \bar{u}_n \bar{y}_{s2}\right]$$

$$= \left[(1-\theta)\bar{y}'_{s2} + \theta\left\{\bar{y}_{s2} + \frac{(N-1)n}{(n-1)N}(\bar{y}'_{s2} - \bar{u}_n \bar{y}_{s2})\right\}\right] \tag{2.9}$$

Remark 2.1

(i) For $\theta = 0$ in $\hat{\hat{Y}}_u$ gives the unbiased estimator $\bar{y}'_{s2}$ while for $\theta = 1$ it reduces to the unbiased estimator

$$\bar{y}''_{s2} = \bar{y}_{s2} + \frac{(N-1)n}{(n-1)N}(\bar{y}'_{s2} - \bar{u}_n \bar{y}_{s2})$$

which is cited in Sukhatme and Sukhatme (1970).

(ii) For $\theta = (1-\eta)^{-1}$, $\hat{\hat{Y}}_u$ boils down to the unbiased estimator

$$\hat{\hat{Y}}_1 = \frac{n(N-1)}{(N-n)}\bar{y}_{s2}\bar{u}_n - \frac{(n-1)N}{(N-n)}\bar{y}_{s2} \tag{2.10}$$

while for $\theta = \eta^{-1}$, $\hat{\hat{Y}}_u$ comes out to be

$$\hat{\hat{Y}}_2 = \left[\frac{N(n-1)}{(N-1)n}\bar{y}_{s2} + \frac{\{N(n+1)-2n\}}{(N-1)n}\bar{y}'_{s2} - \bar{u}_n \bar{y}_{s2}\right] \tag{2.11}$$

It is to be mentioned that the unbiased estimator $\hat{\hat{Y}}_1$ and $\hat{\hat{Y}}_2$ are analogous to the estimators suggested by Ruiz Espejo and Santos Penas (1989).

(iii) For $\theta = -1$ it reduces to the estimator

$$\hat{\hat{Y}}_3 = \left[\frac{\{n(N+1)-2N\}}{(n-1)N}\bar{y}'_{s2} - \bar{y}_{s2} + \frac{(N-1)n}{(n-1)N}\bar{u}_n \bar{y}_{s2}\right] \tag{2.12}$$

Many other unbiased estimators can be generated from the proposed family $\hat{\hat{Y}}_u$ just by substituting the suitable values of θ.

3. Optimum Estimator in the Proposed Family $\hat{\hat{Y}}_u$

The variance of $\hat{\hat{Y}}_u$ is given by

$$\begin{aligned} V(\hat{\hat{Y}}_u) &= V\left[(1-\theta)\bar{y}'_{s2} + \theta\{\bar{y}_{s2} + \eta(\bar{y}'_{s2} - \bar{u}_n \bar{y}_{s2})\}\right] \\ &= V\left[(1-\theta)\bar{y}'_{s2} + \theta\bar{y}''_{s2}\right] \\ &= (1-\theta)^2 V(\bar{y}'_{s2}) + \theta^2 V(\bar{y}''_{s2}) + 2\theta(1-\theta)\operatorname{cov}(\bar{y}'_{s2}, \bar{y}''_{s2}) \\ &= V(\bar{y}'_{s2}) + \theta^2\left[V(\bar{y}'_{s2}) + V(\bar{y}''_{s2}) - 2\operatorname{cov}(\bar{y}'_{s2}, \bar{y}''_{s2})\right] \\ &\quad - 2\theta\left[V(\bar{y}'_{s2}) - \operatorname{cov}(\bar{y}_{s2}, \bar{y}''_{s2})\right] \end{aligned} \tag{3.1}$$

which is minimized for

$$\theta = \frac{[V(\bar{y}'_{s2}) - \text{cov}(\bar{y}'_{s2}, \bar{y}''_{s2})]}{[V(\bar{y}'_{s2}) + V(\bar{y}''_{s2}) - 2\,\text{cov}(\bar{y}'_{s2}, \bar{y}''_{s2})]}$$

$$= \theta_{opt} \qquad \text{(say)} \tag{3.2}$$

θ_{opt} can be re-expressed as

$$\theta_{opt} = \frac{\text{cov}(\bar{y}'_{s2}, \hat{\bar{Y}}^*)}{V(\hat{\bar{Y}}^*)} \tag{3.3}$$

where $\hat{\bar{Y}}^* = (\bar{y}'_{s2}, \bar{y}''_{s2})$

Substitution of (3.3) in (3.1) yields the minimum variance of $\hat{\bar{Y}}_u$ as

$$\min.V(\hat{\bar{Y}}_u) = V(\bar{y}'_{s2})(1 - \rho_1^2) \tag{3.4}$$

where $\rho_1 = \dfrac{\text{cov}(\bar{y}'_{s2}, \hat{\bar{Y}}^*)}{\sqrt{V(\bar{y}'_{s2})V(\hat{\bar{Y}}^*)}}$ is the correlation coefficient between $\bar{y}'_{s2}$ and $\hat{\bar{Y}}^*$.

Remark 3.1 : From (3.4) it is immediate that

$$\min.V(\hat{\bar{Y}}_u) \le (\bar{y}'_{s2})$$

or

$$V(\hat{\bar{Y}}_u^{opt}) \le V(\bar{y}'_{s2}) \tag{3.5}$$

We have from (2.4) that

$$V(\bar{y}''_{s2}) - V(\hat{\bar{Y}}_u^{opt})\ (\text{or min}.\,V(\hat{\bar{Y}})) = \frac{[V(\bar{y}''_{s2}) - \text{cov}(\bar{y}'_{s2}, \bar{y}''_{s2})]^2}{V(\hat{\bar{Y}}^*)}$$

$$> \theta \text{ provided } V(\bar{y}''_{s2}) \ne \text{Cov}(\bar{y}'_{s2}, \bar{y}''_{s2})$$

which gives

$$V(\hat{\bar{Y}}_u^{opt}) \le V(\bar{y}''_{s2}) \tag{3.6}$$

It follows from (3.5) and (3.6) that the optimum unbiased estimator (OUE) $\hat{\bar{Y}}_u^{opt}$ is more efficient than both the unbiased estimators $\bar{y}'_{s2}$ and $\bar{y}''_{s2}$.

Now to have tangible idea about the variance of $\hat{\bar{Y}}_u$, we write the variance and covariance expressions, to the first degree of approximation as

$$V(\bar{y}''_{s2}) = \frac{(1-f)}{n}\left[S'^2_{by} + \bar{\bar{Y}}^2 S'^2_{bu} - 2\bar{\bar{Y}} S'_{byu}\right] + \frac{1}{nN}\sum_{i=1}^{N} u_i^2 \frac{(1-f_i)}{m_i} S_{iy}^2 \quad (3.7)$$

$$\mathrm{cov}(\bar{y}'_{s2}, \bar{y}''_{s2}) = \frac{(1-f)}{n}\left[S'^2_{by} - \bar{\bar{Y}} S'_{byu}\right] + \frac{1}{nN}\sum_{i=1}^{N} u_i^2 \frac{(1-f_i)}{m_i} S_{iy}^2 \quad (3.8)$$

where

$$S'^2_{bu} = \frac{\sum_{i=1}^{N} (u_i - 1)^2}{(N-1)}$$

$$S'_{byu} = \frac{\sum_{i=1}^{N} (u_i \bar{Y}_i - \bar{\bar{Y}})(u_i - 1)}{(N-1)}$$

Substituting (1.7), (3.7) and (3.8) in (3.1) we get the variance of $\hat{\bar{Y}}_u$ to the first degree of approximation as

$$V(\hat{\bar{Y}}_u) = \frac{(1-f)}{n}\left[S'^2_{by} + \theta^2 \bar{\bar{Y}}^2 S'^2_{bu} - 2\theta \bar{\bar{Y}} S'_{byu}\right] + \frac{1}{nN}\sum_{i=1}^{N} u_i^2 \frac{(1-f_i)}{m_i} S_{iy}^2 \quad (3.9)$$

or

$$V(\hat{\bar{Y}}_u) = \frac{(1-f)}{n}\left[S'^2_{by} + \theta \bar{\bar{Y}}^2 S'^2_{bu}(\theta - 2K_{yu})\right] + \frac{1}{nN}\sum_{i=1}^{N} u_i^2 \frac{(1-f_i)}{m_i} S_{iy}^2 \quad (3.10)$$

where

$$K_{yu} = \frac{S'_{byu}}{\bar{\bar{Y}} S'^2_{bu}} = \rho'_b \frac{C'_{by}}{C'_{bu}} \quad (3.11)$$

and

$$C'_{by} = \frac{S'_{by}}{\bar{\bar{Y}}},\ \rho'_b = \frac{S'_{byu}}{S'_{by} S'_{bu}},\ C'_{bu} = \frac{S'_{bu}}{\bar{\bar{U}}},\ \bar{U} = 1 \quad (3.12)$$

The variance of $\hat{\bar{Y}}_u$ at (3.10) is minimized for

$$\theta = K_{yu} = \theta_{opt} \quad (3.13)$$

Substitution of (3.13) in (3.10) yield the minimum variance of $\hat{\bar{Y}}_u$ as

$$\min.V(\hat{\bar{Y}}_u) = \frac{(1-f)}{n} S'^2_{by}(1-\rho'^2_b) + \frac{1}{nN}\sum_{i=1}^{N} u_i^2 \frac{(1-f_i)}{m_i} S^2_{iy} \quad (3.14)$$

which is same as the approximate variance of regression estimator $\hat{\bar{Y}}_{1yu} = \bar{y}'_{s2} + \hat{\beta}'_{byu}(\bar{U} - \bar{u}_n)$ (with $\bar{U} = 1$), where $\hat{\beta}_{byu}$ is the sample estimate of population regression coefficient $\beta_{byu} = \frac{S'_{byu}}{S'^2_{bu}}$ of $u_i\bar{Y}_i$ on u_i.

Putting (3.11) in (2.9) we get the "asymptotically optimum unbiased" (AOU) estimator in the family of estimators $\hat{\bar{Y}}_u$ in (2.9) as

$$\hat{\bar{Y}}_{u(opt)} = \left[\left(1 - K_{yu}\left\{1 - \frac{n(N-1)}{(n-1)N}\right\}\right)\bar{y}'_{s2} + K_{yu}\bar{y}_{s2} - \frac{n(N-1)}{(n-1)N}K_{yu}\bar{u}_n\bar{y}_{s2}\right] \quad (3.15)$$

whose variance is same as given in (3.14).

Remark 3.2 : It is to be noted that the AOU estimator $\hat{\bar{Y}}_{u(opt)}$ can be used in practice only when the exact value of K_{yu} is known in advance. However, in practice, one can use the value of K_{yu} from some earlier survey or pilot study in constructing the AOU estimator $\hat{\bar{Y}}_{u(opt)}$ in (3.15), for instance see Murthy (1967), Reddy and Rao (1977), Reddy (1974, 1978), Sahai and Sahai (1985), Srivenkataramana and Tracy (1980), Tracy *et al.* (1996) and Singh and Singh (1993).

Remark 3.3 : When ss'u equals to the fsu's (i.e. $m_i = M_i$), the results of the present study reduce to the results cited in Singh (1994) and Singh (1999). Thus the present investigation generalizes the work of Singh (1994) and Singh and Singh (1999).

4. Estimator Based on Estimated Optimum Value

It is to be mentioned that the optimum value $\theta_{opt} = K_{yu}$ at (1.13) depends upon the variances and co-variances of estimators in two-stage sampling and involves population parameters $\bar{\bar{Y}}, S'_{byu}, S'^2_{bu}$ on which it is hard to find prior information in some practical situations. In such circumstances, it is worth advisable to obtain the consistent estimate of the optimum value θ_{opt} of θ from the sample data at hand, replacing the population parameters $\bar{\bar{Y}}, S'_{byu}$ and S'^2_{bu} by their unbiased estimators $\bar{y}_{s2}, s'_{byu}$ and s'^2_{bu} respectively, where $\bar{y}_{s2} = \frac{1}{n}\sum_{i=1}^{n}\bar{y}_i$ is same as defined in (1.1), $s'_{by} = \frac{\sum_{i=1}^{n}(u_i\bar{y}_i - \bar{y}'_{s2})(u_i - 1)}{(n-1)}$, $s'^2_{bu} = \frac{\sum_{i=1}^{n}(u_i - 1)^2}{(n-1)}$, and $\bar{y}'_{s2} = \frac{1}{n}\sum_{i=1}^{n}u_i\bar{y}_i$ is same as defined in (1.2). Thus a consistent estimate of θ_{opt} is given by

$$\hat{\theta}_{opt} = \hat{K}_{yu} = \frac{s'_{byu}}{\bar{y}_{s2}s'^2_{bu}} \tag{4.1}$$

Replacing K_{yu} by $\hat{K}_{yu}$· in (3.15) we get a consistent estimator of population mean $\bar{Y}$ based on estimated optimum value as

$$\hat{\bar{Y}}_{u(o\hat{p}t)} = \left[\left\{1 - \hat{K}_{yu}\left(1 - \frac{n(N-1)}{N(n-1)}\right)\right\}\bar{y}'_{s2} + \hat{K}_{yu}\bar{y}_{s2} - \frac{n(N-1)}{N(n-1)}\hat{K}_{yu}\bar{u}_n\bar{y}_{s2}\right]$$

or

$$\hat{\bar{Y}}_{u(o\hat{p}t)} = \left[(1 - \hat{K}_{yu})\bar{y}'_{s2} + \hat{K}_{yu}\left\{\bar{y}_{s2} + \frac{n(N-1)}{N(n-1)}(\bar{y}_{s2} - \bar{u}_n\bar{y}_{s2})\right\}\right]$$

or

$$\hat{\bar{Y}}_{u(o\hat{p}t)} = \left[(1 - \hat{K}_{yu})\bar{y}'_{s2} + \hat{K}_{yu}\bar{y}'''_{s2}\right] \tag{4.2}$$

where $\bar{y}'''_{s2}$ is an unbiased estimator of population mean $\bar{Y}$ defined at (1.5).

It can be shown that

$$E(\hat{\hat{\bar{Y}}}_{u(o\hat{p}t)}) = \bar{Y} + o(n^{-1}) \tag{4.3}$$

which shows that the estimator based on estimated optimum value is not unbiased. To the first degree of approximation, the variance of $\hat{\hat{\bar{Y}}}_{u(o\hat{p}t)}$ is given by

$$V\left(\hat{\hat{\bar{Y}}}_{u(o\hat{p}t)}\right) = \frac{(1-f)}{n}S'^2_{by}(1 - \rho'^2_b) + \frac{1}{nN}\sum_{i=1}^{N}u_i^2\frac{(1-f_i)}{m_i}S^2_{iy} \tag{4.4}$$

which is same as the "asymptotic optimum unbiased" estimator $\hat{\bar{Y}}_{u(opt)}$ i.e.

$$V\left(\hat{\hat{\bar{Y}}}_{u(o\hat{p}t)}\right) = V\left(\hat{\bar{Y}}_{u(opt)}\right) = \min. V(\hat{\bar{Y}}_u) \tag{4.5}$$

Thus we note from (4.2) that if unbiasedness is not the primary concern, one can successfully use the estimator $\hat{\hat{\bar{Y}}}_{u(o\hat{p}t)}$ in (4.2) "based on estimated optimum value" in practice with the same efficiency as $\hat{\bar{Y}}_{u(opt)}$.

5. *Efficiency Comparisons*

From (1.7), (1.14) and (3.14) it can easily be proved that $\hat{\bar{Y}}_{u(opt)}$ (or $\hat{\hat{\bar{Y}}}_{u(o\hat{p}t)}$) is better than the unbiased estimators $\bar{y}'_{s2}$ and $\bar{y}''_{s2}$.

We have from (1.7) and (3.10) that

$$V(\bar{y}'_{s2}) - V(\hat{\bar{Y}}_u) = \frac{(1-f)}{n} \theta S'^2_{bu} \bar{\bar{Y}}^2 (2K_{yu} - \theta)$$

which is positive if $\hat{\bar{Y}}_u$ is more efficient than $\bar{Y}''_{s2}$ if

$$\left.\begin{array}{ll} \text{either} & 0 < \theta < 2K_{yu} \\ \text{or} & 2K_{ku} < \theta < 0 \end{array}\right\} \tag{5.1}$$

Further, we note from (1.14) and (3.10) that the proposed unbiased estimator

$$V(\bar{y}''_{s2}) - V(\hat{\bar{Y}}_u) = \frac{(1-f)}{n} \bar{\bar{Y}}^2 S'^2_{bu} (1-\theta)[\theta - (2K_{yu} - 1)] \tag{5.2}$$

is greater than zero i.e. $\hat{\bar{Y}}_u$ is more efficient than $\bar{y}''_{s2}$ if

$$\left.\begin{array}{ll} \text{either} & (2K_{yu} - 1) < \theta < 1 \\ \text{or} & 1 < \theta < (2K_{yu} - 1) \end{array}\right\} \tag{5.3}$$

It is to be mentioned that $\bar{y}'_{s2}$ is better than $\bar{y}''_{s2}$ if

$$K_{yu} > \frac{1}{2} \tag{5.4}$$

Thus from (5.1), (5.3) and (5.4), the following theorem can be easily proved.

Theorem 5.1 : Any member of the family $\hat{\bar{Y}}_u$ would be more efficient than $\bar{y}'_{s2}$ and $\bar{y}'''_{s2}$ up to first order of approximation, if θ is so chosen as to satisfy the following inequalities

(i) $0 < \theta < 2K_{yu}$ when $0 < K_{yu} < \frac{1}{2}$

(ii) $(2K_{yu} - 1) < \theta < 1$ when $\frac{1}{2} < K_{yu} < 1$

(iii) $1 < \theta < (K_{yu} - 1)$ when $K_{yu} > 1$

6. Empirical Study

Population : Source : Cochran (1977)

$N = 3, n = 1, m_i = 2, (i = 1, 2, 3)$

The data of artificially constructed population are presented in Table 6.1.

Table 6.1. Artificial population with units of unequal sizes

Unit	y_{ij}	M_i	m_i	$\bar{Y}_i$	S^2_{iy}
1	0,1	2	2	0.5	0.500
2	1,2,2,3	4	2	2.0	0.667
3	3,3,4,4,5,5	6	2	4.0	0.800

The required parameters are

$\bar{\bar{Y}} = 2.167$, $\bar{Y} = 2.75$, $\bar{M} = 4$

$S^2_{by} = 3.9219$, $S'^2_{by} = 8.6875$, $S''^2_{by} = 3.9219$

$S'''^2_{by} = 3.6314$, $S'^2_{bu} = 0.2500$, $S'_{byu} = 1.4375$

Table 6.2 gives the percent relative efficiency (PRE) of $\hat{\bar{Y}}_u$ with respect to various estimators $\bar{y}_{s2}, \bar{y}'_{s2}, \bar{y}''_{s2}$ and $\bar{y}'''_{s2}$ of population mean $\bar{Y}$. For computing the percent relative efficiencies of $\hat{\bar{Y}}_u$ with respect to $\bar{y}_{s2}, \bar{y}'_{s2}, \bar{y}''_{s2}$ and $\bar{y}'''_{s2}$ we have used the following formulae respectively

$$\text{PRE}(\hat{\bar{Y}}_u, \bar{y}_{s2}) = \frac{\text{MSE}(\bar{y}_{s2})}{V(\hat{\bar{Y}}_u)} \times 100$$

$$= \frac{\left[\frac{(1-f)}{n} S^2_{by} + \frac{1}{nN} \sum_{i=1}^{N} \frac{(1-f_i)}{m_i} S^2_{iy} + (\bar{\bar{Y}} - \bar{Y})^2\right]}{\left[\frac{(1-f)}{n} A + B\right]} \times 100$$

$$\text{PRE}\,(\hat{\bar{Y}}_u, \bar{y}'_{s2}) = \frac{V(\bar{y}'_{s2})}{V(\hat{\bar{Y}}_u)} \times 100$$

$$= \frac{\left[\frac{(1-f)}{n} S'^2_{by} + B\right]}{\left[\frac{(1-f)}{n} A + B\right]} \times 100$$

$$\text{PRE}(\hat{\bar{Y}}_u, \bar{y}''_{s2}) = \frac{V(\bar{y}''_{s2})}{V(\hat{\bar{Y}}_u)} \times 100$$

$$= \frac{\left[\frac{1-f}{n} S''^2_{by} + B\right]}{\left[\frac{(1-f)}{n} A + B\right]} \times 100$$

and

$$\text{PRE}(\hat{\bar{Y}}_u, \bar{y}'''_{s2}) = \frac{V(\bar{y}'''_{s2})}{V(\hat{\bar{Y}}_u)} \times 100$$

$$= \frac{\left[\frac{(1-f)}{n} S'''^2_{by} + B\right]}{\left[\frac{(1-f)}{n} A + B\right]} \times 100$$

where

$$A = \left[S'^2_{by} + \theta^2 \bar{\bar{Y}}^2 S'^2_{bu} - 2\theta \bar{\bar{Y}} S'_{byu}\right]$$

and

$$B = \frac{1}{nN} \sum_{i=1}^{N} u_i^2 \frac{(1-f_i)}{m_i} S^2_{iy}$$

Table 6.2 exhibits that when θ attains its optimum value (i.e. $\theta_{opt} = 2.65338$), the maximum gain in efficiency by using the proposed family of unbiased estimators $\hat{\bar{Y}}_u$ over $\bar{y}_{s2}, \bar{y}'_{s2}, \bar{y}''_{s2}$ and $\bar{y}'''_{s2}$, is observed. It is also observed from Table 6.2 that even if θ deviates from its optimum value ($\theta_{opt} = 2.65338$) the gain efficiency is substantial over $\bar{y}_{s2}, \bar{y}'_{s2}, \bar{y}''_{s2}$ and $\bar{y}'''_{s2}$.

Thus we conclude from the ranges of θ that there is enough scope of choosing θ in obtaining more efficient estimators from the proposed family of estimators $\hat{\bar{Y}}_u$.

Table 6.2. Showing the percent relative efficiencies (PRE's) of $\hat{\bar{Y}}_u$ with respect to various estimators of $\bar{Y}$

θ	PRE($\bar{Y}_u$, $\hat{\bar{y}}_{s2}$)	PRE($\bar{Y}_u$, $\hat{\bar{y}}'_{s2}$)	PRE($\bar{Y}_u$, $\hat{\bar{y}}''_{s2}$)	PRE($\bar{Y}_u$, $\hat{\bar{y}}'''_{s2}$)
–0.06101	40.29	99.99	45.53	42.46
–0.06001	40.32	100.05	45.56	42.49
0.00099	42.03	104.30	47.50	44.29
0.25000	50.22	124.62	56.75	52.92
0.50000	60.96	151.29	68.89	64.24
0.75000	75.31	186.89	85.11	79.36
0.92599	88.42	219.43	99.93	93.12
0.92699	88.51	219.64	100.02	93.27
0.99999	94.89	235.49	107.24	99.99
1.00000	94.89	235.49	107.24	100.00
1.00099	94.99	235.72	107.34	100.10
1.05000	99.64	247.26	112.60	104.99
1.05300	99.93	248.00	112.93	105.31
1.10000	104.71	259.86	118.33	110.35
1.25000	122.20	303.36	138.10	128.78
1.50000	160.94	399.39	181.87	169.60
1.75000	216.03	536.10	244.13	227.66
2.00000	219.57	723.58	329.50	307.27
2.25000	382.33	948.81	432.07	402.91
θ_{opt} = 2.50000	457.33	1134.93	516.82	481.95
2.65338	473.02	1173.85	534.55	498.47
2.75000	466.66	1158.09	527.37	491.78
3.00000	402.52	998.91	454.88	424.18
3.25000	311.43	772.85	351.94	328.19
3.50000	231.33	574.07	261.42	243.78
3.75000	171.82	426.40	194.17	181.07
4.00000	129.83	322.20	146.72	136.82
4.25000	100.30	248.92	113.35	105.70
4.25300	100.01	248.18	113.02	105.39
4.25400	99.91	247.94	112.91	105.29
4.25500	99.81	247.69	112.79	105.18
4.30599	94.96	235.66	107.32	100.07
4.30699	94.87	235.44	107.21	99.98
4.37899	88.57	219.80	100.09	93.34
4.37999	88.57	219.59	99.99	93.25
4.50000	79.22	196.60	89.53	83.49
5.00000	52.40	130.04	59.22	55.22
5.36669	40.32	100.04	45.57	42.49
5.36799	40.29	99.97	45.53	42.45
Range of θ	(1.0530, 4.2540)	(–0.0610, 5.3667)	(0.9260, 4.3799)	(0.9999, 4.3060)

ACKNOWLEDGEMENT

Authors are thankful to the referee for his valuable suggestions in improving the paper.

REFERENCES

Cochran, W.G. (1977). *Sampling Techniques.* Third Edition, Wiley Eastern Limited, New Delhi.

Murthy, M.N. (1967). *Sampling Theory and Methods.* Statistical Publishing Society, Calcutta.

Reddy, V.N. and Rao T.J. (1977). Modified PPS method of estimation. *Sankhya,* **C39,** 185–197.

Reddy, V.N. (1974). On a transformed ratio method of estimation. *Sankhya,* **C36,** 59–70.

Reddy, V.N. (1978). A study of the prior knowledge on certain population parameters. *Sankhya,* **C40,** 29–37.

Ruiz Espezo, M. and Santos Penas, J. (1989). Unbiased mean ratio estimators. *Statistica, anno,* **49(4),** 617–622.

Sahai, A. and Sahai, A. (1985). On efficient use of auxiliary information. *J. Statist. Plann. Inf.,* **12,** 203–212.

Singh, R. (1994). Study of some unbiased estimators for finite population mean in sample surveys. M. Phil dissertation submitted to Vikram University, Ujjain, M.P.

Singh, R. and Singh, H.P. (1999). A class of unbiased estimators in cluster sampling. *J. Ind. Soc. Agril. Statist.,* **52(3),** 290–298.

Singh, S. and Singh, R. (1993). A new method : Almost separation of bias precipitates in sample surveys. *J. Ind. Statist. Assoc.,* **31,** 99–105.

Srivenkatramana, T. and Tracy, D.S. (1980). An alternative to ratio method in sample surveys. *Ann. Inst. Statist. Math.,* **32,** 111–120.

Sukhatme P.V. and Sukhatme, B.V. (1970). *Sampling Theory of Surveys with Applications.* Second Revised Edition, Iowa State University Press, Ames, U.S.A. and Indian Society of Agricultural Statistics, New Delhi.

Tracy, D.S., Singh, H.P. and Singh, R. (1996). An alternative to the ratio–cum-product estimator in sample surveys. *J. Statist. Plann. Inf.,* **53,** 375–387.