Jour. Ind. Soc. Ag. Statistics
54(1), 2001 : 29–42

On a Biased Estimator in Repeat Surveys

U.C. Sud, A.K. Srivastava and D.P. Sharma
Indian Agricultural Statistics Research Institute, New Delhi
(Received : May, 2000)

SUMMARY

A general theory of estimation for repeat surveys in case of the biased estimator is developed. Estimates for the current occasion and change are developed as particular cases in respect of two occasions. The theory so developed assumes the knowledge of coefficient of variation of the population. Suitable estimators based on estimated coefficient of variation have been proposed.

Key words : Biased estimator, Repeat surveys.

1. *Introduction*

Jessen [2], Tikkiwal [9], Patterson [3] and Raj [4] developed the theory of unbiased estimation of mean of characteristics in successive sampling. Searls [6] considered the use of information pertaining to the coefficient of variation in developing an estimator for population mean with smaller squared error. This approach is extended to develop general theory of estimation for repeat surveys in case of the biased estimator.

2. *Some General Results*

Consider a finite population $\Omega = (U_1, \dots U_N)$ of 'N' identifiable units. Let y be the variable under study taking value y_i on unit U_i, $i = 1, \dots, N$. Let the parameter of interest be $\overline{Y} = \frac{1}{N} \sum_{i}^{N} y_i$. The variance of a single observation is defined by

$$\sigma^2 = \frac{1}{N} \sum_{i}^{N} (y_i - \overline{Y})^2 = \frac{N-1}{N} S^2$$

where $S^2 = \frac{1}{N-1} \sum_{i}^{N} (y_i - \overline{Y})^2$

A sample of size 'n' units is drawn from the population by SRSWOR. For the sake of simplicity in exposition we assume that the fpc s^s are ignorable. Consider linear estimators of the type

$$\hat{T} = \sum_{i}^{n} l_i y_i \text{ such that } \sum_{i}^{n} l_i = k$$

The bias of $\hat{T}$ is given by

$$(k-1)\bar{Y}$$

and the corresponding mean squared error by

$$MSE(\hat{T}) = \sigma^2 \sum_{i}^{n} l_i^2 - \frac{\sigma^2}{(N-1)} \sum_{i \neq j}^{n} l_i l_j + (k-1)^2 \bar{Y}^2 \tag{2.1}$$

We determine optimum values of l_i's by minimizing (2.1) subject to the condition that $\sum_{i}^{n} l_i = k$, in other words we minimize

$$\sigma^2 \sum_{i}^{n} l_i^2 - \frac{\sigma^2}{(N-1)} \sum_{i \neq j}^{n} l_i l_j + (k-1)^2 \bar{Y}^2 - 2\lambda \left(\sum_{i}^{n} l_i - k \right)$$

where λ is the Lagrangian multiplier. We thus, obtain following equalities

$$\sigma^2 l_i - \frac{\sigma^2}{(N-1)} \sum_{j(\neq i)}^{n} l_j = \text{Cov}\left(y_i, \sum_{i}^{n} l_i y_i \right) = \lambda \tag{2.2}$$

$$(1-k)\bar{Y}^2 = \lambda \tag{2.3}$$

$$\sum_{i}^{n} l_i = k \tag{2.4}$$

It may be noted that the minimized mean squared error is

$$\sigma^2 l_{iopt} - \frac{\sigma^2}{(N-1)} \sum_{j(\neq i)}^{n} l_{iopt} = \lambda = (1-k_{opt})\bar{Y}^2$$

The optimum values of l_i's and k are given by

$$l_{iopt} = \frac{1}{C^2 + n} \text{ and } k_{opt} = \frac{n}{(C^2 + n)}$$

where C^2 is the square of coefficient of variation of the population of y's defined by

$$C^2 = \frac{\sigma^2}{\overline{Y}^2}$$

Substituting the optimum value of 'k' in (2.3) we obtain the minimum mean squared error of $\hat{T}$, reported by Searls (1964), as

$$\frac{\sigma^2}{C^2 + n}$$

It can be seen that (2.2) holds good for $\forall$ i. Thus

$$\text{Cov}\left(\sum_i^n \xi_i y_i,\ \sum_i^n l_i y_i \right) = \lambda = \text{ Minimized mean squared error of } \hat{T}$$

where $\sum_i^n \xi_i y_i$ is any unbiased estimator of $\overline{Y}$ i.e. $\sum_i^n \xi_i = 1$

Thus, we have the following result :

The minimum mean squared error of $\hat{T}$ is the covariance of any unbiased estimator of $\overline{Y}$ and $\hat{T}$.

This minimum mean squared error is equal to $(1 - k_{opt})\, \overline{Y}^2$. Further, the minimum mean squared error linearly biased estimator is uncorrelated with every zero function of 'y' values.

Next, we consider a set up wherein there are number of variates divided into different sets. The i-th variate in the j-th set is denoted by y_{ij} $(1 \le i \le n_j;\ 1 \le j \le h)$. The population mean in the m-th set is denoted by $\overline{Y}_m$ and its estimate by $\hat{T}_m$ $(1 \le m \le h)$. Consider estimator of the type

$$\hat{T}_m = \sum_i^{n_j} \sum_j^{h} w_{ij} y_{ij} \quad \text{such that}$$

$$\sum_i^{n_j} w_{ij} = k_j \quad \text{for } j = m$$

$$= 0 \quad \text{otherwise} \tag{2.5}$$

The w_{ij}'s are to be determined by minimizing mean square error of $\hat{T}_m$ subject to the condition (2.5).

The bias of $\hat{T}_m$ can be seen equal to $(1-k_m)\,\overline{Y}_m$.

The expression to be minimized is thus

$$\sigma^2 \sum \sum w_{ij}^2 + \sum_{i} \sum_{j \neq j'} w_{ij} w_{ij'} \operatorname{Cov}(y_{ij}, y_{ij'}) + \sum_{i \neq i'} \sum_{j} w_{ij} w_{i'j} \operatorname{Cov}(y_{ij}, y_{i'j})$$

$$+ \sum_{i \neq i'} \sum_{j \neq j'} w_{ij} w_{i'j} \operatorname{Cov}(y_{ij}, y_{i'j}) + (k_m - 1)^2 \overline{Y}_m^2 - 2 \sum_{j} \lambda_j \left(\sum_{i} w_{ij} - k_j \right)$$

where λ_j's are the Lagrangian multipliers. The minimization leads to following equations

$$\operatorname{Cov}(y_{ij}, \hat{T}_m) = \lambda_j \;\forall\; i, j \tag{2.6}$$

$$(1 - k_m)\,\overline{Y}_m^2 = \lambda_m \tag{2.7}$$

$$\sum_{i}^{n_j} w_{ij} = k_j \tag{2.8}$$

In view of (2.6), (2.7) and (2.8) we have the following result :

The minimum mean squared error of $\hat{T}_m$ is the covariance of any unbiased estimator of $\overline{Y}_m$ and $\hat{T}_m$. This minimum mean square error is equal to $(1-k_m)\,\overline{Y}_m^2$. Further, $\hat{T}_m$ is uncorrelated with every zero function of $y_{ij}\ \forall\ i, j$.

The foregoing result will be used to develop minimum mean squared linearly biased estimators for sampling on successive occasions with partial replacements of units.

3. The Minimum Mean Squared Linearly Biased Estimator for the h-th Occasion

Let there be a finite population of 'N' units sampled over a number of occasions (say, h). A sub-sample of $m = n\lambda$ units observed on the (h-1)-th occasion is retained for use on the h-th occasion and an independent sample of $u = n - m = n\mu$ units is selected on the h-th occasion. For the sake of simplicity in exposition we assume that the population variances remain constant and that the sample sizes and the proportions retained are the same on each occasion. Further, correlations between observations on the same units, one, two, three, ... occasions apart are $\rho, \rho^2, \rho^3, \ldots$. We also assume that the fpc's are ignorable. The population mean on the j-th occasion will be denoted by M_j ($j = 1, 2, \ldots, h$). We define

$\bar{x}'_{h-1}$ = mean of the observations on occasion (h-1) associated with $n\lambda$ units common with occasion 'h'

$\bar{y}'_h$ = mean of the $n\lambda$ units on occasion 'h' common with the (h − 1) occasion

$\bar{x}''_{h-1}$ = mean of $n\mu$ units not common to h-th occasion

$\bar{y}''_h$ = mean of $n\mu$ units not common to (h − 1) occasion

We consider minimum mean squared linear estimator of M_h as

$$\hat{M}_h = A\hat{M}_{h-1} + k_{h-1} B\bar{x}''_{h-1} - k_{h-1}(A+B)\bar{x}'_{h-1} + \phi_{1h}\bar{y}''_h + \phi_{2h}\bar{y}_h'$$

where $\hat{M}_{h-1}$ is the minimum mean squared linear estimator of M_{h-1} and $\phi_{1h} + \phi_{2h} = k_h$.

To determine A and B we apply the condition that $\hat{M}_h$ is uncorrelated with every zero function. In particular

$$\text{Cov}(\bar{x}''_{h-1}, \hat{M}_h) = \text{Cov}(\bar{x}'_{h-1}, \hat{M}_h)$$

$$\text{Cov}(\bar{x}''_{h-2}, \hat{M}_h) = \text{Cov}(\bar{x}'_{h-2}, \hat{M}_h)$$

Thus, we get $B = 0$ and $A = \dfrac{\phi_{2h}\,\rho}{k_{h-1}}$

The minimum mean squared linear estimator is

$$\hat{M}_h = \frac{\phi_{2h}\,\rho}{k_{h-1}}\hat{M}_{h-1} - \phi_{2h}\,\rho\bar{x}'_{h-1} + \phi_{1h}\bar{y}_h'' + \phi_{2h}\bar{y}'_h \tag{3.1}$$

The quantities ϕ_{1h} and ϕ_{2h} will be determined from the conditions

$$\text{Cov}(\bar{y}''_h, \hat{M}_h) = \text{Cov}(\bar{y}'_h, \hat{M}_h)$$

$$\phi_{1h} + \phi_{2h} = k_h \tag{3.2}$$

Further, $\text{MSE}_{\min}(\hat{M}_h) = \text{Cov}(\hat{M}_h, \bar{y}_h'') = \frac{\phi_{1h}}{n\mu}\sigma^2 = (1-k_h)\bar{Y}_h^2$ (3.3)

Using (3.2) and (3.3), we obtain

$$\varphi_{2h} = 1 - \phi_{1h}\left(\frac{C_h^2 + n\mu}{n\mu}\right)$$

and

$$\phi_{1h} = \frac{\rho^2 \frac{\text{MSE}_{\min}(\hat{M}_{h-1})}{k_{h-1}} + \frac{\sigma^2}{n\mu}(1-\rho^2)}{\frac{\sigma^2}{n} + \frac{\sigma^2}{m}\left(\frac{C_h^2 + n\mu}{n\mu}\right)\left[(1-\rho^2) + \frac{\rho^2}{k_{h-1}}\text{MSE}_{\min}(\hat{M}_{h-1})\right]} \tag{3.4}$$

where $C_j^2 = \frac{\sigma^2}{M_j^2}$ = square of the coefficient of variation of the variable observed on the j-th occasion $\forall j = 1, 2, \ldots, h$.

We define

$$G_j = \frac{\text{MSE}_{\min}(\hat{M}_j)}{\text{MSE}_{\min}(\hat{M}_1)} = \frac{\phi_{1j}}{n\mu}(C_1^2 + n)$$

It may be noted that $G_1 = 1$.

From (3.4) it can be easily seen that

$$\frac{1}{G_h} = \frac{C_h^2 + n(1-\lambda_h)}{C_1^2 + n} + \frac{n\lambda_h k_{h-1}}{\rho^2 n\,\lambda_h\, G_{h-1} + k_{h-1}(C_1^2 + n)(1-\rho^2)} \tag{3.5}$$

Maximizing (3.5) with respect to λ_h, we obtain

$$\lambda_{hopt} = \frac{k_{(h-1)opt}(C_1^2 + n)\sqrt{1-\rho^2}}{n(1+\sqrt{1-\rho^2})\,G_{(h-1)opt}} \tag{3.6}$$

$$\frac{1}{G_{hopt}} = \frac{(C_h^2 + n)G_{(h-1)opt}(1+\sqrt{1-\rho^2}) + k_{(h-1)opt}(1-\sqrt{1-\rho^2})(C_1^2 + n)}{(C_1^2 + n)\,G_{(h-1)opt}(1+\sqrt{1-\rho^2})} \tag{3.7}$$

Also, the optimum value of k_h is given by

$$k_{hopt} = \frac{n(1+\sqrt{1-\rho^2})G_{(h-1)opt} + k_{(h-1)opt}(C_1^2 + n)(1-\sqrt{1-\rho^2})}{(C_h^2 + n)(1+\sqrt{1-\rho^2})\,G_{(h-1)opt} + k_{(h-1)opt}(C_1^2 + n)(1-\sqrt{1-\rho^2})}$$

It can be shown that

$$\frac{k_{(h-1)opt}}{G_{(h-1)opt}} = \frac{n}{(C_1^2+n)} + \frac{k_{(h-2)opt}}{G_{(h-2)opt}}B$$

where $B = \dfrac{(1-\sqrt{1-\rho^2})}{(1+\sqrt{1-\rho^2})}$

Repeated use of the recurrence relation gives

$$\frac{k_{(h-1)opt}}{G_{(h-1)opt}} = \frac{n}{(C_1^2+n)}\left[\frac{1-B^{h-1}}{1-B}\right] \tag{3.8}$$

Substituting (3.8) in (3.6) we get

$$\lambda_{hopt} = \frac{1}{2}[1-B^{h-1}] \tag{3.9}$$

Similarly, it can be shown that

$$\frac{1}{G_{hopt}} = \frac{2(C_h^2+n)(1+\sqrt{1-\rho^2})\sqrt{1-\rho^2}+n\rho^2(1-B^{h-1})}{2\sqrt{1-\rho^2}(1+\sqrt{1-\rho^2})(C_1^2+n)} \tag{3.10}$$

$$MSE(\hat{M}_h) = \frac{2\sqrt{1-\rho^2}(1+\sqrt{1-\rho^2})}{2(C_h^2+n)\sqrt{1-\rho^2}(1+\sqrt{1-\rho^2})+n\rho^2(1-B^{h-1})} \tag{3.11}$$

From (3.8) and (3.11) we can see that

$$\lambda_\infty = \frac{1}{2}$$

$$MSE(\hat{M}_\infty) = \frac{2\sqrt{1-\rho^2}(1+\sqrt{1-\rho^2})}{2C_\infty^2\sqrt{1-\rho^2}(1+\sqrt{1-\rho^2})+n(1+\sqrt{1-\rho^2})^2} \tag{3.12}$$

In particular, if only two occasion data are available

$$\hat{M}_2 = \left[C_2^2\frac{(1-\mu\rho^2)}{n}+1-\rho^2\mu^2\right]^{-1}[\lambda\bar{y}'+\mu(1-\rho^2\mu)\bar{y}''+\lambda\mu(\bar{x}''-\bar{x}')] \tag{3.13}$$

$$\lambda_{2opt} = \frac{\sqrt{1-\rho^2}}{1+\sqrt{1-\rho^2}},\quad \mu_{2opt} = \frac{1}{1+\sqrt{1-\rho^2}},$$

$$k_{2opt} = \frac{2n}{2n+C_2^2(1+\sqrt{1-\rho^2})}$$

and $$\text{MSE}(\hat{M}_2)_{\text{opt}} = \frac{\sigma^2(1+\sqrt{1-\rho^2})}{2n + C_2^2(1+\sqrt{1-\rho^2})} \tag{3.14}$$

Let $\hat{M}'_2$ be the minimum variance unbiased estimator on the second occasion, then

$$\hat{M}'_2 = [1-\rho^2\mu^2]^{-1}\,[\lambda\bar{y}' + \mu(1-\rho^2\mu)\,\bar{y}'' + \lambda\mu\rho\,(\bar{x}''-\bar{x}')]$$

$$V(\hat{M}'_2)_{\text{opt}} = \frac{\sigma^2\,(1+\sqrt{1-\rho^2})}{2n}$$

The percent gain in precision of $\hat{M}_2$ over $\hat{M}'_2$ is given by

$$\text{Percent gain in precision} = C_2^2\,\frac{[1+\sqrt{1-\rho^2}\,]}{2n} \times 100 \tag{3.15}$$

It can be seen that the $\text{MSE}(\hat{M}_2)_{\text{opt}}$ is always less than $V(\hat{M}'_2)_{\text{opt}}$. Also, the gain decreases as sample size increases.

4. Empirical Study

Data, listed in Sukhatme ([7], area under wheat in 1936 and 1937), Samford ([5], total acreage of tillage crops and grass in 1947 and the actual area recorded under oats in 1957), and Cochran ([1], sizes of large United States cities in 1920 and 1930), were utilized for empirical study. In addition to this, data pertaining to production of apple from a survey conducted by IASRI on cost of production of apple in hilly areas of UP for the years 1972-73, 1973-74 and secondary data on wheat production for a zone comprising of two blocks of Bullandshahar district of UP for the two years 1984-85 and 1985-86, in a survey also conducted by IASRI, were utilised. For the purpose of illustration we treat this data as population and call these populations as I, II, III, IV and V respectively. For the empirical study 20%, 15% and 10% samples were drawn from these populations respectively. The results of the study are presented in Table 1.

Table 1. % Gain in precision of $\hat{M}_2$ over $\hat{M}'_2$

Populations	C_2^2	ρ^2	% Gain in precision		
			Percentage of Population Sampled		
			20%	15%	10%
I	0.554	0.865	5.57	7.57	11.14
II	0.495	0.702	5.46	7.35	10.92
III	0.909	0.964	5.52	7.36	11.04
IV	4.429	0.080	22.36	29.81	44.73
V	0.521	0.579	8.95	11.93	17.90

It can be seen from Table 1 that the percentage gain in precision of $\hat{M}_2$ over $\hat{M}'_2$ increases with decrease in the sample size. Also, gain in precision is substantial when C_2^2 is high and ρ^2 is low.

The minimum mean squared linearly biased estimator on the first occasion is given by

$$\hat{M}_1 = \left[C_1^2 \frac{(1-\mu\rho^2)}{n} + 1 - \rho^2\mu^2\right]^{-1} \left[\lambda\bar{x}' + \mu(1-\rho^2\mu)\,\bar{x} + \lambda\mu\,(\bar{y}'' - \bar{y}')\right]$$

The estimate of change is given by

$$\hat{M}_2 - \hat{M}_1$$

and the corresponding minimum mean squared error as

$$\text{MSE}\,(\hat{M}_2 - \hat{M}_1)_{\min} =$$

$$\sigma^2 \frac{(1-\rho)(1+\mu)}{n}\left[\frac{1}{\frac{C_2^2}{n}(1-\mu\rho^2) + 1 - \mu^2\rho^2} + \frac{1}{\frac{C_1^2}{n}(1-\mu\rho^2) + 1 - \mu^2\rho^2}\right]$$

5. *Alternative Estimators*

If C_2^2 is unknown, we propose estimating C_2^2 by

$$c_2^2 = \frac{s_y^2}{\bar{y}_u^2}, \quad s_u^2 = \frac{1}{U-1}\sum_{i=1}^{U} (y_i - \bar{y}_u)^2$$

Substituting c_2^2 for C_2^2 in (3.13) and after some readjustments we obtain

$$\hat{M}_{2A} = [c_2^2(n - u\rho^2) + (n^2 - u^2\rho^2)]^{-1}\ [\lambda\,\{\bar{y}' + \beta\,(\bar{x}_n - \bar{x}')\} + u\,(n - u\rho^2)\,\bar{y}'']$$

$$= \left[\frac{c_2^2 T}{n} + 1\right]^{-1}\left[\frac{\lambda}{(1-\mu^2\rho^2)}\{\bar{y}' + \beta\,(\bar{x}_n - \bar{x}')\} + \mu T\,\bar{y}''\right] \qquad (4.1)$$

where $T = [1 - \mu\,\rho^2]\,[1 - \mu^2\rho^2]^{-1}$

We now investigate the large sample properties of $\hat{M}_2$

Write

$$\bar{y}_u = \bar{Y} + \varepsilon_{1u}\,; \quad \bar{y}_m = \bar{Y} + \varepsilon_{1m}\,; \quad \bar{x}_n = \bar{X} + \varepsilon_2\,; \quad \bar{x}_m = \bar{X} + \varepsilon_{2m}$$

$$s_u^2 = \sigma^2 + v$$

$$\hat{M}_{2A} = \frac{\bar{Y}}{1-\mu^2\rho^2}\left\{\lambda\left(1+\frac{2\varepsilon_{1u}}{Y}+\frac{\varepsilon_{1u}^2}{Y^2}\right)\left(1+\frac{\varepsilon_{1m}}{Y}+\frac{\beta}{Y}(\varepsilon_2-\varepsilon_{2m})\right)\right.$$

$$\left.+\mu(1-\mu\rho^2)\left(1+\frac{3\varepsilon_{1u}}{\bar{Y}}+\frac{3\varepsilon_{1u}^2}{\bar{Y}^2}+\frac{3\varepsilon_{1u}^3}{\bar{Y}^3}\right)\right\}$$

$$\left\{1+\frac{C_2^2}{n}T+\frac{v}{n\bar{Y}^2}T+\frac{2\varepsilon_{1u}}{\bar{Y}}+\frac{\varepsilon_{1u}^2}{\bar{Y}^2}\right\}^{-1} E$$

expanding the right hand side and retaining terms to order $0_p(n^{-2})$, the bias and mean squared error of $\hat{M}_{2A}$ can be shown equal to

$$B(\hat{M}_{2A}) = -\frac{C_2^2}{n}T\left[\left\{1-\frac{C_2^2}{n}T+\frac{1}{n\mu}\left(C_2^2-\sqrt{\frac{\beta_1}{C_2^2}}\right)\right\}\right.$$

$$\left.+\frac{C_2^2\lambda}{n\mu(1-\mu^2\rho^{2n})}\left\{2-\sqrt{\frac{\beta_1}{C_2^2}}\right\}\right]\bar{Y} \qquad (4.2)$$

$$M(\hat{M}_{2A}) = \frac{\sigma^2}{n}T\left[1+\frac{w_{12}}{n}T\right]$$

where $w_{12} = C_2^2\left[3-2\sqrt{\frac{\beta_1}{C_2^2}}\right]$, $\sqrt{\beta_1} = \frac{\mu_{3y}}{\sigma_y^2}$ is the coefficient of skewness, μ_{3y} being the third central moment.

The mean squared error (4.3) will be smaller than the sampling variance of $\hat{M}'_2$ provided

$$\sqrt{\frac{\beta_1}{C_2^2}} > \frac{3}{2} \qquad (4.4)$$

or in other words w_{12} is negative. The value of w_{12} was computed for populations I, II and III,.. described in Section 3, and was found to be negative in respect of populations II and III. However, for normal population, $\beta_1 = 0$ so that the estimator M_{2A} will not only be biased but it will have larger mean squared error. This will be true for all populations where in the distribution is symmetrical.

One of the feasible optimum values of μ and λ in the sense of minimum mean squared error of $\hat{T}_{1A}$ can be seen equal to

$$\mu_{opt} = [1+\sqrt{1-\rho^2}\,]^{-1},\ \lambda_{opt} = [\sqrt{1-\rho^2}\,]\,[1+\sqrt{1-\rho^2}\,]^{-1}$$

and the resulting minimum mean squared error as

$$\frac{\sigma^2}{2n}(1+\sqrt{1-\rho^2}\,)\left[1+\frac{w_{12}}{2n}(1+\sqrt{1-\rho^2}\,)\right] \tag{4.5}$$

while the other feasible optimum values of μ and λ are given by (if w_{12} is negative)

$$\mu_{opt} = -\frac{w_{12}}{n}-\frac{1}{\rho}\sqrt{\frac{w_{12}^2}{n^2}\rho^2+1+2\frac{w_{12}}{n}}$$

$$\lambda_{opt} = 1+\frac{w_{12}}{n}+\frac{1}{\rho}\sqrt{\frac{w_{12}^2}{n^2}\rho^2+1+2\frac{w_{12}}{n}}$$

subject to the conditions that $-\frac{w_{12}}{n} > 1$ and $\frac{w_{12}^2}{n^2}\rho^2+1 > -2\frac{w_{12}}{n}$, and the resultant minimum mean squared error as

$$M(\hat{M}_{2A})_{min} = -\frac{\sigma^2}{4w_{12}}$$

The value of $-\frac{w_{12}}{n}$ was computed for populations I, II and III. The condition $-\frac{w_{12}}{n} > 1$ was not satisfied for all these populations. Therefore, percent gain in precision of $\hat{M}_{2A}$ over $\hat{M}'_2$ was obtained only by utilizing expression (4.5). This percent gain in precision can be seen equal to

$$-\frac{w_{12}(1+\sqrt{1-\rho^2}\,)}{2n+w_{12}(1+\sqrt{1-\rho^2}\,)}\times 100 \tag{4.6}$$

Next, we consider an alternative estimator of C_2^2 wherein we put unbiased estimators of σ_y^2 and $\overline{Y}^2$ is C_2^2. Thus, we propose estimating C_2^2 by

$$\hat{C}_{2A}^2 = s_u^2\left[\overline{y}_u^2-\frac{s_u^2}{u}\right]^{-1}$$

Substituting $\hat{C}^2_{2A}$ for C_2^2 in (5), we get

$$\hat{M}_{2A1} = \frac{1}{(1-\mu^2\rho^2)}\,[\lambda\{\bar{y}_m + \beta(\bar{x}_n - \bar{x}_m)\}$$

$$+ \mu(1-\mu\rho^2)\,\bar{y}_u]\left(\bar{y}_u^2 - \frac{s_u^2}{u}\right)\left(\frac{s_u^2}{n}T + \bar{y}_u^2 - \frac{s_u^2}{u}\right)^{-1}$$

However, instead we consider a slightly general class of estimators given by

$$\hat{M}_{2Ag} = \frac{1}{(1-\mu^2\rho^2)}\,[\lambda\{\bar{y}_m + \beta(\bar{x}_n - \bar{x}_m)\}$$

$$+ \mu(1-\mu\rho^2)\,\bar{y}_u]\,\frac{\bar{y}_u^2 + L\dfrac{s_u^2}{u}}{\left[s_u^2\left(\dfrac{T}{n} - \dfrac{1}{u}\right) + \bar{y}_u^2\right]} \tag{4.7}$$

where L is a constant to be chosen suitably.

The bias and mean squared error of $\hat{M}_{2Ag}$ to order $0_p(n^{-2})$ can be shown equal to

$$B(\hat{M}_{2Ag}) = \bar{Y}\frac{C_2^2}{n\mu}\left[L - T_1 + \frac{C_2^2}{n\mu}T_1^2 - \frac{LC_2^2}{n\mu}T_1 + \left[\sqrt{\frac{\beta_1}{C_2^2}} - 1\right]C_2^2\frac{(T_1 - L)}{n}T\right.$$

$$\left. + \frac{C_2^2(T_1 - L)}{n\mu}\,\frac{\lambda}{(1-\mu^2\rho^2)}\left[2\sqrt{\frac{\beta_1}{C_2^2}} - 3\right]\right] \tag{4.8}$$

$$M(\hat{M}_{2Ag}) = \frac{\sigma^2}{n}\left[T + \frac{L^2C_2^4}{n\mu^2} + \frac{2(L-T_1)}{n\mu}\sqrt{\beta_1 C_2^2 T} - \frac{2LC_2^2}{n\mu}T\right.$$

$$\left. - \frac{2LC_2^2T_1}{n\mu^2} + \frac{3C_2^2T_1T}{n\mu} - \frac{C_2^2T_1}{n\mu^2}\right] \tag{4.9}$$

where $T_1 = T\mu - 1$

It can be easily seen that the proposed estimator $\hat{M}_{2Ag}$ has smaller mean squared error than variance of the unbiased estimator $\hat{M}'_2$ if either of the conditions hold

$$T_1 < L < T_1 + 2\mu T\left(1 - \sqrt{\frac{\beta_1}{C_2^2}}\right) \text{ and } 1 > \sqrt{\frac{\beta_1}{C_2^2}}$$

or

$$T_1 + 2\mu T\left(1 - \sqrt{\frac{\beta_1}{C_2^2}}\right) < L < T_1 \text{ and } 1 < \sqrt{\frac{\beta_1}{C_2^2}} \quad (4.10)$$

Minimizing (4.9) with respect to L, we find optimum value of L given by

$$L_{opt} = 1 + 2T_1 - \mu T\sqrt{\frac{\beta_1}{C_2^2}} \quad (4.11)$$

and the resultant mean squared error as

$$M(\hat{M}_{2Ag})_{min} = \frac{\sigma^2}{n} T\left[1 - \frac{C_2^2 T}{n}\left(1 - \sqrt{\frac{\beta_1}{C_2^2}}\right)^2\right] \quad (4.12)$$

One of the feasible optimum values of μ and λ in the sense of minimum mean squared error of $M(\hat{M}_{2Ag})$ can be seen equal to

$$\mu_{opt} = \frac{1}{[1 + \sqrt{1 - \rho^2}]}, \quad \lambda_{opt} = \frac{\sqrt{1 - \rho^2}}{[1 + \sqrt{1 - \rho^2}]}$$

and the corresponding minimum mean squared error as

$$\frac{\sigma^2}{2n}(1 + \sqrt{1 - \rho^2})\left[1 - \frac{Q}{2n}(1 + \sqrt{1 - \rho^2})\right] \quad (4.13)$$

While the other feasible optimum values of μ and λ in the sense of minimum mean squared error of $M(\hat{M}_{2Ag})$ can be seen equal to

$$\mu_{opt} = \frac{Q}{n} - \frac{1}{\rho}\sqrt{\frac{Q^2}{n^2}\rho^2 - 2\frac{Q}{n} + 1}, \quad \lambda_{opt} = 1 - \frac{Q}{n} + \frac{1}{\rho}\sqrt{\frac{Q^2}{n^2}\rho^2 - 2\frac{Q}{n} + 1}$$

subject to the conditions that $\frac{Q}{n} > 1$ and $\frac{Q^2}{n^2}\rho^2 + 1 > 2\frac{Q}{n}$

and the resultant mean squared error as

$$M_{opt}(\hat{M}_{2Ag})_{min} = \frac{\sigma^2}{4Q}, \quad Q = C_2^2\left(1 - \sqrt{\frac{\beta_1}{C_2^2}}\right)^2$$

The value of $\frac{Q}{n}$ was found to be less than 1 for all the three populations, I, II and III. The percent gain in precision of $\hat{M}_{2Ag}$ over $\hat{M}'_2$ is, therefore, obtained utilizing the expression (4.13). This percent gain in precision is given by

$$\frac{Q(1+\sqrt{1-\rho^2})}{2n - Q(1+\sqrt{1-\rho^2})} \times 100 \tag{4.14}$$

ACKNOWLEDGEMENT

The authors are very thankful to referees for their valuable comments/suggestions.

REFERENCES

[1] Cochran, W.G. (1977). *Sampling Techniques* (Third edition). John Wiley and Sons Inc., New York, London.

[2] Jessen, R.J. (1942). Statistical investigation of a sample survey for obtaining farm facts. *Iowa Ag. Expt. Stat. Res. Bul.*, **304**.

[3] Patterson, H.D. (1950). Sampling on successive occasions with partial replacements of units. *J. Roy. Stat. Soc.*, **B 12,** 241-255.

[4] Raj, D. (1965). On sampling over two occasions with probability proportional to size. *Ann. Math. Stat.*, **36,** 327-330.

[5] Samford, M.R. (1962). *An Introduction to Sampling Theory*. Oliver and Boyce Ltd., Edinburg and London.

[6] Searls, D.T. (1964). The utilization of known coefficient of variation in the estimation procedure. *J. Amer. Stat. Assoc.*, **59,** 1225-1226.

[7] Sukhatme, P.V. (1954). *Sampling Theory of Surveys with Applications. J. Ind. Soc. Agril. Stat.*, New Delhi.

[8] Srivastava, V.K. (1974). On the use of coefficient of variation in estimating mean. *J. Ind. Soc. Agril. Stat.*, **26 (2),** 33-36.

[9] Tikkiwal, B.D. (1951). Theory of successive sampling. *Unpublished Diploma Thesis*. I.C.A.R., New Delhi.

[10] Upadhyaya, L.N. and Srivastava, S.R. (1976). A note on the use of coefficient of variation in estimating mean. *J. Ind. Soc. Agril. Stat.*, **28 (2),** 97-99.