COMMUN. STATIST.-THEORY METH., 20(8), 2589-2602 (1991)

ESTIMATION OF MEAN USING DOUBLE SAMPLING FOR STRATIFICATION AND MULTIVARIATE AUXILIARY INFORMATION

T. P. Tripathi

Stat-Math. Division, Indian Statistical Institute, Calcutta - 700 035, India

Shashi Bahl

Dept. of Mathematics, M. D. University, Rohtak - 124 001, India

Key Words and Phrases : Combined and Separate Estimators; Relative Performance, Stratification.

ABSTRACT

Several estimators for estimating the mean of a principal variable are proposed based on double sampling for stratification (DSS) and multivariate auxiliary information. The general properties of the proposed estimators are studied, search for optimum estimators is made and the proposed estimators are compared with the corresponding estimators based on unstratified double sampling (USDS).

1. INTRODUCTION

When the sampling frame within strata is known, stratified sampling is used; but there are many situations of practical importance where the strata weights are known and the frame within strata is not available. In these situations the technique of post-stratification may be employed. However in other situations strata weights may not be known exactly as they become outdated with the passage of time. Further the information on stratification variable may not be readily available but could be made available by diverting a part of the survey budget. Under these circumstances the method of double sampling for stratification (DSS) can be used.

In the proposed DSS Scheme we select a preliminary large sample $S_{(1)}$ of size n' rather inexpensively from a population of N units with simple random sampling without replacement (SRSWOR) and observe the auxiliary variables $x_1, x_2, \ldots, x_p$. Let (x_{ij}), $i = 1,2,\ldots,p$; $j = 1,2,\ldots,n'$ denote the x-observations and $\bar{x}_i' = \sum_{j=1}^{n'} x_{ij}/n'$, the sample means. The sample $S_{(1)}$ is then stratified into L strata on the basis of information for one or more x_i's obtained through $S_{(1)}$. Let n_h' denote the number of units in $S_{(1)}$ falling into h-th stratum ($h = 1,2,\ldots,L$, $\sum_h n_h' = n'$) yielding the representation

$$\bar{x}_i' = \sum_h w_h' \bar{x}_{ih}' \quad \text{where} \quad \bar{x}_{ih}' = \sum_{j=1}^{n_h'} x_{ijh}/n_h' \quad \text{and} \quad w_h' = \frac{n_h'}{n'} .$$

Subsamples of sizes $n_h = v_h n_h'$ $0 < v_h < 1$; $h=1,2,\ldots,L$, v_h being predetermined for each h, are then selected independently, using SRSWOR within each stratum and y, the variable of main interest is observed. Let

$n = \sum_{h=1}^{L} n_h$, and (y_{jh}), $j = 1,2,\ldots,n_h$; $h = 1,2,\ldots,L$ denote y-observations and $\bar{y}_h = \sum_{j=1}^{n_h} y_{jh}/n_h$

$$\bar{x}_{ids} = \sum_h w'_h \, \bar{x}_{ih} \quad \text{where} \quad \bar{x}_{ih} = \sum_{j=1}^{n_h} x_{ijh}/n_h$$

Clearly w'_h is an unbiased estimator of strata weights $W_h = \frac{N_h}{N}$.

Similarly the sample means $\bar{x}'_i$ and $\bar{x}_{ids}$ based on first sample and subsample respectively are unbiased estimators of population mean $\bar{X}_i = \sum_h W_h \bar{X}_{ih}$ of auxiliary variable x_i. For estimating the population mean $\bar{Y}$, the customary unbiased estimator based on DSS and its variance are given by

$$\bar{y}_{ds} = \sum_h w'_h \, \bar{y}_h \tag{1.1}$$

and $$V(\bar{y}_{ds}) = \frac{1-f}{n'} S_o^2 + \frac{1}{n'} \sum_h \left(\frac{1}{v_h} - 1\right) W_h S_{ho}^2 \tag{1.2}$$

[Rao (1973), Cochran (1977)]

where $f = \frac{n'}{N}$; $S_o^2 = \sum_{j=1}^{N} (y_j - \bar{Y})^2/(N-1)$

$$S_{ho}^2 = \sum_{j=1}^{N_h} (y_{jh} - \bar{Y}_h)^2/(N_h - 1) \; ; \; \bar{Y}_h = \sum_{j=1}^{N_h} y_{jh}/N_h$$

Some estimators based on DSS and information on a single auxiliary variable have been proposed by Ige and Tripathi (1987) for improving the precision of estimation compared to $\bar{y}_{ds}$. In this paper we discuss several methods of estimation, based on multivariate auxiliary information, as an effort for further improvement of precision of estimation.

2. MULTIVARIATE COMBINED AND SEPARATE ESTIMATORS BASED ON DSS

Utilizing the information collected on x-variates through the preliminary sample $S_{(1)}$, we define multivariate combined difference, ratio and ratio-cum-product estimators in DSS by

$$e = \sum_{i=1}^{p} a_i \alpha_i$$

$$e = e_{DMC} \quad \text{if} \quad \alpha_i = \bar{y}_{ds} - \lambda_i(\bar{x}_{ids} - \bar{x}_i') \quad i=1,2,\ldots,p \tag{2.1}$$

$$e = e_{RMC} \quad \text{if} \quad \alpha_i = \frac{\bar{y}_{ds}}{\bar{x}_{ids}} \bar{x}_i' \quad i=1,2,\ldots,p \tag{2.2}$$

$$e = e_{RPMC} \quad \text{if} \quad \alpha_i = \frac{\bar{y}_{ds}}{\bar{x}_{ids}} \bar{x}_i' \quad \text{for } i=1,\ldots,q \tag{2.3}$$

$$= \frac{\bar{y}_{ds}\bar{x}_{ids}}{\bar{x}_i'} \quad \text{for } i=q+1,\ldots,p$$

where $a = (a_1, a_2, \ldots, a_p)'$ with $\sum_{i=1}^{p} a_i = 1$ is a weigh-function and λ_i's are suitably chosen constants.

Using the same amount of information, we can define multivariate separate difference, ratio and ratio-cum-product estimators in DSS by

$$e^* = \sum_{h=1}^{L} w_h' d_h \quad \text{with} \quad d_h = \sum_{i=1}^{p} a_{ih} d_{ih}$$

$$e^* = e_{DMS} \quad \text{if} \quad d_{ih} = \bar{y}_h - \lambda_{ih}(\bar{x}_{ih} - \bar{x}_{ih}') \tag{2.4}$$

$$i = 1,2,\ldots,p$$

$$e^* = e_{RMS} \quad \text{if} \quad d_{ih} = \frac{\bar{y}_h}{\bar{x}_{ih}} \bar{x}_{ih}' \quad i = 1,2,\ldots,p \tag{2.5}$$

$$e^* = e_{RPMS} \quad \text{if} \quad d_{ih} = \frac{\bar{y}_h}{\bar{x}_{ih}} \bar{x}_{ih}' \quad i=1,2,\ldots,q \tag{2.6}$$

$$= \frac{\bar{y}_h \bar{x}_{ih}}{\bar{x}'_{ih}} \quad i=q+1,\ldots,p$$

The variables $x_1, x_2, \ldots, x_q$ in e_{RPMC} and e_{RPMS} are those ones of x which are positively correlated with y. Obviously e_{DMC} and e_{DMS} are unbiased for $\bar{Y}$ and exact expressions for their variances are given by

$$V(e_{DMC}) = \frac{1-f}{n'} S_o^2 + \frac{1}{n'} \sum_{h=1}^{L} W_h\left(\frac{1}{v_h} - 1\right) a' B_h a \qquad (2.7)$$

$$V(e_{DMS}) = \frac{1-f}{n'} S_o^2 + \frac{1}{n'} \sum_{h=1}^{L} W_h\left(\frac{1}{v_h} - 1\right) a_h' D_h a_h \qquad (2.8)$$

with $B_h = (b_{hik})$; $D_h = (d_{hik})$ $i,k=1,\ldots,p$

$b_{hik} = S_{ho}^2 - \lambda_i S_{hoi} - \lambda_k S_{hok} + \lambda_i \lambda_k S_{hik}$

$d_{hik} = S_{ho}^2 - \lambda_{ih} S_{hoi} - \lambda_{kh} S_{hok} + \lambda_{ih} \lambda_{kh} S_{hik}$

where $S_{hik} = \frac{1}{N_h - 1} \sum_{j=1}^{N_h} (x_{ihj} - \bar{X}_{ih})(x_{khj} - \bar{X}_{kh})$ $i,k=0,1,\ldots,p$

the subscripts $0,1,2,\ldots,p$ refering to the variables $y, x_1, \ldots, x_p$ respectively.

For large samples, the approximate expressions for the biases and MSES of the estimators e_{RMC}, e_{RMS}, e_{RPMC}, e_{RPMS} are given by

$$B(e_{RMC}) = \frac{1}{n'} \sum_h W_h\left(\frac{1}{v_h} - 1\right) \sum_{i=1}^{p} \frac{a_i}{\bar{X}_i} (R_i S_{hi}^2 - S_{hoi})$$

$$B(e_{RMS}) = \frac{1}{n'} \sum_h W_h\left(\frac{1}{v_h} - 1\right) \sum_{i=1}^{p} \frac{a_{ih}}{\bar{X}_{ih}} (R_{ih} S_{hi}^2 - S_{hoi})$$

$$B(e_{RPMC}) = \frac{1}{n'} \sum_h W_h\left(\frac{1}{v_n} - 1\right)\left[\sum_{i=1}^{q} \frac{a_i}{\bar{X}_i} (R_i S_{hi}^2 - S_{hoi}) + \sum_{i=q+1}^{p} \frac{a_i}{\bar{X}_i} S_{hoi}\right]$$

$$B(e_{RPMS}) = \frac{1}{n'} \sum_h W_h(\frac{1}{v_n}-1)[\sum_{i=1}^{q} \frac{a_{ih}}{\bar{X}_{ih}} (R_{ih}S_{hi}^2-S_{hoi}) + \sum_{i=q+1}^{p} \frac{a_{ih}}{\bar{X}_{ih}} S_{hoi}]$$

$$M(e_{RMC}) = V(e_{DMC}) \quad \text{with} \quad \lambda_i = R_i = \frac{\bar{Y}}{\bar{X}_i} \tag{2.9}$$

$$M(e_{RMS}) = V(e_{DMS}) \quad \text{with} \quad \lambda_{ih} = R_{ih} = \frac{\bar{Y}_h}{\bar{X}_{ih}} \tag{2.10}$$

$$M(e_{RPMC}) = V(e_{DMC}) \quad \text{with} \quad \lambda_i = R_i,\ \lambda_k = R_k \quad \text{for each } i,k=1,2,..,q$$

$$\lambda_i = -R_i,\ \lambda_k = -R_k \quad \text{for each } i,k=q+1,..,p$$

$$\lambda_i = R_i\ ,\ \lambda_k = -R_k \quad \text{for } i = 1,2,...,q;\ k = q+1,...,p. \tag{2.11}$$

$$M(e_{RPMS}) = V(e_{DMS}) \quad \text{with} \quad \lambda_{ih} = R_{ih};\ \lambda_{kh} = R_{kh} \quad \text{for each } i,k=1,2,..,q$$

$$\lambda_{ih} = -R_{ih};\lambda_{kh} = -R_{kh} \quad \text{for each } i,k=q+1,..,p$$

$$\lambda_{ih} = R_{ih};\ \lambda_{kh} = -R_{kh} \quad \text{for } i = 1,...,q,\ k = q+1,..,p. \tag{2.12}$$

Using the results of Rao (1973), non-negative unbiased estimators of $V(e_{DMC})$ and $V(e_{DMS})$ are given by

$$v(e_{DMC}) = \frac{1-f}{n'} s_{ods}^2 + \frac{1}{n'}\sum_h(\frac{1}{v_h}-1)w_h'(\sum_{ik}\sum a_i a_k(s_{ho}^2-\lambda_i s_{hoi} -\lambda_k s_{hok}+\lambda_i \lambda_k s_{hik})$$

$$v(e_{DMS}) = \frac{1-f}{n'} s^2_{ods} + \frac{1}{n'}\sum_h \left(\frac{1}{v_h}-1\right) w'_h \sum\sum a_{ih} a_{kh} (s^2_{ho} - \lambda_{ih} s_{hoi} - \lambda_{kh} s_{hok} + \lambda_{ih}\lambda_{kh} s_{hik})$$

where

$$s^2_{ods} = \sum_h \frac{w'_h}{n_h} \left[(n_h-1) + \frac{(n'_h-1)}{(n'-1)}\right] s^2_{ho} + \frac{n'}{(n'-1)} \sum_h w'_h (\bar{y}_h - \bar{y}_{ds})^2$$

and

$$s_{hik} = \frac{1}{n_h-1} \sum_{j=1}^{n_h} (x_{ihj} - \bar{x}_{ih})(x_{khj} - \bar{x}_{kh}),\ i,k=0,1,\ldots,p$$

with $s^2_{ho} = s_{hoo}$.

Further, non-negative but biased estimators for the MSES of e_{RMC}, e_{RMS}, e_{RPMC}, e_{RPMS} are given by

$m(e_{RMC}) = v(e_{DMC})$ with $\lambda_i = r_i = \bar{y}_{ds}/\bar{x}_{ids}$

$m(e_{RMS}) = v(e_{DMS})$ with $\lambda_{ih} = r_{ih} = \bar{y}_h/\bar{x}_{ih}$

$m(e_{RPMC}) = v(e_{DMC})$ with $\lambda_i = r_i;\ \lambda_k = r_k$
for each $i,k=1,2,\ldots,q$
$\lambda_i = -r_i;\ \lambda_k = -r_k$
for each $i,k=q+1,\ldots,p$
$\lambda_i = r_i;\ \lambda_k = -r_k$
for $i = 1,2,\ldots,q$
$k = q+1,\ldots,p.$

$m(e_{RPMS}) = v(e_{DMS})$ with $\lambda_{ih} = r_{ih};\ \lambda_{kh} = r_{kh}$
for each $i,k=1,2,\ldots,q$
$\lambda_{ih} = -r_{ih};\ \lambda_{kh} = -r_{kh}$
for each $i,k=q+1,\ldots,p$
$\lambda_{ih} = r_{ih};\ \lambda_{kh} = -r_{kh}$
for $i = 1,2,\ldots,q,$
$k = q+1,\ldots,p.$

3. OPTIMUM ESTIMATORS

Let $\beta_{oi} = \sum_h C^*_{ih}\beta_{oih}/\sum_h C^*_{ih}$ with $C^*_{ih}=(\frac{1}{v_h}-1)W_h S^2_{hi}$ be the weighted average of the strata population regression coefficients $\beta_{oih} = S_{hoi}/S^2_{hi}$ of y on x_i and

$$\rho_{ik} = \frac{\sum_h (\frac{1}{v_h}-1)W_h\rho_{hik}S_{hi}S_{hk}}{[\sum_h(\frac{1}{v_h}-1)W_h S^2_{hi}\cdot \sum_h (\frac{1}{v_h}-1)W_h S^2_{hk}]^{1/2}}$$

where $\rho_{hik} = S_{hik}/S_{hi}S_{hk}$ is the correlation coefficient between x_i and x_k in stratum h.
For p = 1, when information on only x_i is used, following Ige and Tripathi (1987) the optimum value of λ_i in (2.7) is given by

$$\lambda_{oi} = \beta_{oi}$$

When the choices $\lambda_i = \beta_{oi}$ are made for each i, the resulting variance is given by

$$[V(e_{DMC})]_{\lambda_i=\beta_{oi}} = \frac{1-f}{n'} S^2_o + \frac{1}{n'}(a'Ba)\sum_h W_h(\frac{1}{v_h}-1)S^2_{ho} \quad (3.1)$$

where $B = (b_{ik}) \quad i,k = 1,\ldots,p$

$$b_{ik} = 1 - \rho^2_{oi} - \rho^2_{ok} + \rho_{ik}\rho_{oi}\rho_{ok}$$

Further, when optimum weight vector

$$a_o = \frac{B^{-1} g}{g'B^{-1}g}, \quad g = (1,1,\ldots,1)'$$

is used, we obtain

$$[V(e_{DMC})]_{\substack{\lambda_i=\beta_{oi} \\ a = a_o}} = \frac{1-f}{n'} S^2_o + \frac{1}{n'}(g'B^{-1}g)^{-1}\sum_h W_h(\frac{1}{v_h}-1)S^2_{ho}$$

In practice, when exact value of $\lambda_{oi}= \beta_{oi}$ is not available, it may be estimated through

$$\beta^*_{oi}= \sum_h w'_h(\frac{1}{v_h}-1)s_{hoi}/\sum_h w'_h(\frac{1}{v_h}-1)s^2_{hi}$$

Using the estimated optimum values, we may define a combined multivariate estimator for $\bar{Y}$ in DSS by

$$e_{rgMC}^{(1)} = \bar{y}_{ds} - \sum_{i=1}^{p} a_i \beta_{oi}^{*} (\bar{x}_{ids} - \bar{x}_i')$$

For large samples $M(e_{rgMC}^{(1)})$ would again be given by (3.1).

One may in fact obtain simultaneous optimum values of $T_i = a_i \lambda_i$ $(i = 1,2,\ldots,p)$ as follows.

Let $S^* = (S_{ik}^*)$; $Q = (Q_1, Q_2, \ldots, Q_p)'$

where $S_{ik}^* = \sum_h W_h (\frac{1}{v_h} - 1) S_{hik}$; $Q_i = S_{oi}^*$ $i,k = 0,1,2,\ldots,p$

Then $$V(e_{DMC}) = \frac{1-f}{n'} S_o^2 + \frac{1}{n'} (S_o^{*2} - 2T'Q + T'S^*T)$$

which gives $$T_{opt} = T_o = S^{*-1} Q \tag{3.2}$$

and $$V_o(e_{DMC}) = \frac{1-f}{n'} S_o^2 + \frac{1}{n'} S_o^{*2}(1 - R^2)$$

where $R^2 = \frac{Q'S^{*-1}Q}{S_o^{*2}}$, R being the multiple correlation coefficient between $\bar{y}_{ds}$ and $(d_1, d_2, \ldots, d_p)$, with $d_i = \bar{x}_{ids} - \bar{x}_i'$.

The optimum value of T may be estimated by

$$T^* = s^{*-1} Q^*; \quad s^* = (s_{ik}^*), \quad Q^* = (Q_1^*, \ldots, Q_p^*)'$$

where $$s_{ik}^* = \sum_h w_h' (\frac{1}{v_h} - 1) s_{hik}$$

and $$Q_i^* = \sum_h w_h' (\frac{1}{v_h} - 1) s_{hoi}$$

Using these estimated values, we may define a combined multiple regression estimator for $\bar{Y}$ as

$$e_{rgMC}^{(2)} = \bar{y}_{ds} - \sum_{i=1}^{p} T_i^* (\bar{x}_{ids} - \bar{x}_i')$$

whose variance for large samples is given by (3.2).

For separate estimator, when optimum λ_{ih} is used separately for each i,

$$V(e_{DMS})_{\lambda_{ih}=\beta_{oih}} = \frac{1-f}{n'} S_o^2 + \frac{1}{n'} \sum_h W_h\left(\frac{1}{v_h}-1\right) S_{ho}^2 a_h' B_h a_h \qquad (3.3)$$

where $B_h = (b_{hik})$

$b_{hik} = 1 - \rho_{hoi}^2 - \rho_{hok}^2 + \rho_{hik}\rho_{hoi}\rho_{hok}$

Further if optimum weight vector $a_{oh} = \dfrac{B_h^{-1} g}{g' B_h^{-1} g}$ is used,

we obtain

$$V(e_{DMS})_{\substack{\lambda_{ih}=\beta_{oih} \\ a_{ih}=a_{oih}}} = \frac{1-f}{n'} S_o^2 + \frac{1}{n'} \sum_h W_h\left(\frac{1}{v_h}-1\right) S_{ho}^2 (g' B_h^{-1} g)^{-1}$$

In practice when the optimum choice $\lambda_{oih} = \beta_{oih}$ may not be made, it may be estimated through $\beta_{oih}^* = s_{hoi}/s_{hi}^2$ and a separate multivariate regression-type estimator for $\bar{Y}$ may be defined as

$$e_{rgMS}^{(1)} = \sum_h w_h' [\bar{y}_h - \sum_i a_{ih} \beta_{oih}^* (\bar{x}_{ih} - \bar{x}_{ih}')]$$

It may be noted that $M(e_{rgMS}^{(1)})$ may be approximated, for large n_h' in each strata, through the expression in (3.3).

For obtaining simultaneous optimum values of $T_{ih} = a_{ih}\lambda_{ih}$ $(i=1,2,\ldots,p)$ let

$T_h = (T_{1h}, T_{2h}, \ldots, T_{ph})'$; $S_h = (S_{hik})$, $Q_h = (Q_{h1}, Q_{h2}, \ldots, Q_{hp})'$ where $Q_{hi} = S_{hoi}$. We may express

$$V(e_{DMS}) = \frac{1-f}{n'} S_o^2 + \frac{1}{n'} \sum_h W_h\left(\frac{1}{v_h}-1\right)(S_{ho}^2 - 2T_h' Q_h + T_h' S_h T_h)$$

which gives $T_{hopt} = T_{oh} = S_h^{-1} Q_h$

and $\quad V_o(e_{DMS}) = \dfrac{1-f}{n'} S_o^2 + \dfrac{1}{n'} \sum_h W_h\left(\dfrac{1}{v_h}-1\right) V_h$

with $\quad V_h = S_{ho}^2 (1 - R_{ho(1,2,\ldots,p)}^2)$

where $R_{ho(1,2,\ldots,p)}$ is the multiple correlation coefficient between y and x 's in the h-th stratum. The estimated value of T_{oh} is given by $T^*_h = s_h^{-1} Q^*_h$ where $s_h = (s_{hik})$, $Q^*_h = (Q^*_{1h}, \ldots, Q^*_{ph})'$, $Q^*_{ih} = s_{hoi}$.

Using the above estimated optimum value, a separate multiple regression estimator for $\bar{Y}$ may be defined as

$$e^{(2)}_{rgMS} = \sum_h w'_h(\bar{y}_h - \sum_{i=1}^{p} T^*_{ih} (\bar{x}_{ih} - \bar{x}'_{ih}))$$

whose variance, for large samples, is given by (3.4).

4. RELATIVE PERFORMANCE OF THE PROPOSED ESTIMATORS

From (3.2) we observe that if the weighted partial regression coefficients T_{oi} are used as $T_i = a_i \lambda_i$, the variance of the corresponding estimator would be always smaller than that of the customary estimator $\bar{y}_{ds}$.

In practice, however, exact optimum T_o may not be known. Let $T = \alpha T_o = \alpha S^{*-1} Q$, then for any T we find from (1.2), (2.7) and (3.2) after some algebraic simplification that

$$V(\bar{y}_{ds}) - V(e_{DMC}) = \frac{1}{n'} \alpha(2-\alpha) T'_o S^* T_o \qquad (4.1)$$

We note that e_{DMC} would be better than $\bar{y}_{ds}$ as far as $0 < \alpha < 2$. In practice good guessed values T^*_o of T_o may be available through census data, past sample survey data or pilot survey and be used in e_{DMC} which would be better than $\bar{y}_{ds}$ if $T^*_o = \alpha T_o$, $0 < \alpha < 2$. Similarly from (1.2), (2.8) and (3.4) we find that e_{DMS} would be better than $\bar{y}_{ds}$ if

$$T_h = \alpha_h T_{oh}, \; 0 < \alpha_h < 2 \qquad \text{for each } h = 1,2,\ldots,L.$$

From (1.2) and (2.9) we find that a sufficient condition for e_{RMC} to be better than $\bar{y}_{ds}$ is given by

$$\rho_{hoi} \frac{C_{ho}}{C_{hi}} \frac{R_{ih}}{R_i} > \frac{1}{2} \qquad \text{for all } i = 1,2,\ldots,p$$
$$h = 1,2,\ldots,L .$$

If the strata ratios $R_{ih} = R_i$, then the condition reduces to

$$\rho_{hoi} \frac{C_{ho}}{C_{hi}} > \frac{1}{2} \tag{4.2}$$

which is the usual condition for customary separate ratio estimator to be better than mean per unit. Similarly from (1.2) and (2.10) it follows that e_{RMS} would be better than $\bar{y}_{ds}$ if (4.2) holds. It may be noted that the separate ratio, ratio-cum-product and regression type estimators discussed in Section 3 are suitable only for large values of n_h in each stratum.

5. COMPARISON WITH CORRESPONDING UNSTRATIFIED DOUBLE SAMPLING (USDS) ESTIMATORS

The multivariate difference (Raj (1965)), multivariate ratio (Khan and Tripathi (1967)) and multivariate-ratio-cum-product (Rao and Mudholkar (1967)) estimators for the population mean $\bar{Y}$ in USDS are defined by

$$\begin{aligned} \bar{y}'_{DM} &= \sum_{i=1}^{p} a_i\alpha_i \quad \text{where } \alpha_i = \bar{y} - \lambda_i(\bar{x}_i - \bar{x}'_i) \quad i=1,2,\ldots,p \\ \bar{y}'_{RM} &= \sum_{i=1}^{p} a_i\alpha_i \quad \text{where } \alpha_i = \frac{\bar{y}}{\bar{x}_i}\bar{x}'_i \quad i=1,2,\ldots,p \end{aligned} \tag{5.1}$$

$$\begin{aligned} \bar{y}'_{RPM} = \sum_{i=1}^{p} a_i\alpha_i \quad \text{where } \alpha_i &= \frac{\bar{y}}{\bar{x}_i}\bar{x}'_i \quad i=1,2,\ldots,q \\ &= \frac{\bar{y}}{\bar{x}'_i}\bar{x}_i \quad i=q+1,\ldots,p \end{aligned}$$

and $\sum_{i=1}^{p} a_i = 1$. Further

$$V(\bar{y}'_{DM}) = \left(\frac{1}{n'} - \frac{1}{N}\right)S_o^2 + \left(\frac{1}{n} - \frac{1}{n'}\right)\sum_{i}\sum_{k} a_i a_k b_{ik} \tag{5.2}$$

$$b_{ik} = S_o^2 - \lambda_i S_{oi} - \lambda_k S_{ok} + \lambda_i\lambda_k S_{ik}$$

$$M(\bar{y}'_{RM}) = V(\bar{y}'_{DM}) \quad \text{with } \lambda_i = R_i = \frac{\bar{Y}}{\bar{X}_i} \quad i,k=1,2,\ldots,p \qquad (5.3)$$

$$M(\bar{y}'_{RPM}) = V(\bar{y}'_{DM}) \quad \text{with } \lambda_i=R_i;\ \lambda_k=R_k \quad i,k=1,2,\ldots,q \qquad (5.4)$$

$$\lambda_i=-R_i;\lambda_k=-R_k \quad i,k=q+1,\ldots,p$$

$$\lambda_i=R_i;\ \lambda_k=-R_k \quad i=1,2,\ldots,q$$

$$k=q+1,\ldots,p$$

where n is the size of the second phase sample selected randomly. It may be noted that expression in (5.2) is valid for all sample sizes while the expressions in (5.3) and (5.4) are approximate and valid for large samples.

For comparison, we assume in case of DSS estimators that sample allocation to the strata is proportional ($n_h \propto n'_h$, $h = 1,2,\ldots,L$) that is

$$v_h = \frac{n_h}{n'_h} = \frac{n}{n'}$$

We obtain that

$$V(\bar{y}'_{DM}) - V(e_{DMC}) = \left(\frac{1}{n} - \frac{1}{n'}\right) \sum_h W_h a' D_h^{(1)} a$$

$$M(\bar{y}'_{RM}) - M(e_{RMC}) = \left(\frac{1}{n} - \frac{1}{n'}\right) \sum_h W_h a' D_h^{(2)} a$$

$$M(\bar{y}'_{RPM}) - M(e_{RPMC}) = \left(\frac{1}{n} - \frac{1}{n'}\right) \sum_h W_h a' D_h^{(3)} a$$

where $D_h^{(m)} = (d_{hik}^{(m)}) \qquad m = 1,2,3.$

$$d_{hik}^{(1)} = [(\bar{Y}_h-\bar{Y})-\lambda_i(\bar{X}_{ih}-\bar{X}_i)]\ [(\bar{Y}_h-\bar{Y})-\lambda_k(\bar{X}_{kh}-\bar{X}_k)];$$

$$i,k=1,2,\ldots,p$$

$$d_{hik}^{(2)} = [\bar{Y}_h-R_i\bar{X}_{ih}]\ [\bar{Y}_h-R_k\bar{X}_{kh}]; \qquad i,k=1,2,\ldots,p$$

$$d_{hik}^{(3)} = [\bar{Y}_h-R_i\bar{X}_{ih}]\ [\bar{Y}_h-R_k\bar{X}_{kh}] \quad \text{for } i,k=1,2,\ldots,q$$

$$= [\bar{Y}_h-R_i\bar{X}_{ih}]\ [\bar{Y}_h+R_k\bar{X}_{kh}] \quad \text{for } i=1,2,\ldots,q$$

$$k=q+1,\ldots,p$$

$$= [\bar{Y}_h + R_i \bar{X}_{ih}] \, [\bar{Y}_h + R_k \bar{X}_{kh}] \quad \text{for} \quad i,k = q+1,\ldots,p$$

It is noted that $D_h^{(1)}$, $D_h^{(2)}$, $D_h^{(3)}$ are all positive definite matrices. Thus under proportional allocation of the second sample, the multivariate combined difference, ratio and ratio-cum-product estimators in DSS are always better than the corresponding estimators in USDS.

ACKNOWLEDGEMENTS

The authors are thankful to the referee for valuable suggestions leading to a better presentation of the paper. Further, the second author expresses her gratitude to the authorities of F. C. College, Hisar, Haryana, for granting study leave and to Prof. R.K. Tuteja for providing facilities to work at Deptt. of Maths., M.D. University.

BIBLIOGRAPHY

Cochran, W.G. (1977). Sampling Techniques; 3rd Edition, New York, Wiley.

Ige, Abel F. and Tripathi, T.P. (1987). On double sampling for stratification and use of auxiliary information; J. Ind. Soc. Agr. Stat., 39, 191-201.

Khan, S. and Tripathi, T.P. (1967). The use of multivariate auxiliary information in double sampling; J. Ind. Stat. Assoc., 5, 42-48.

Raj, D. (1965 (a)). On method of using multi-auxiliary information in sample surveys; J. Amer. Statist. Assoc., 60, 270-277.

Rao, J.N.K. (1973). On double sampling for stratification and analytical surveys; Biometrika, 60, 125-133.

Rao, P.S.R.S. and Madholkar, G.S. (1967). Generalized multivariate estimators for the mean of a finite population; J. Amer. Statist. Assoc., 62, 1009-1012.

Received December 1990; Revised April 1991.

Recommended Anonymously.