ON DOUBLE SAMPLING FOR MULTIVARIATE RATIO AND DIFFERENCE METHODS OF ESTIMATION*

By

T. P. Tripathi

*Indian Institute of Technology***, New Delhi*

(Received in December 1971;
Revised in July, 1974)

1. Introduction

Let $U = \{u_1, .. , u_N\}$ denote the population consisting of N units and let $y, x_1,..., x_p$ and z be the variates defined on U taking the values $y_j, x_{1j},..., x_{pj}$ and z_j respectively on u_j ($j=1, ..., N$), that is, $y_j=y(u_j)$, $x_{ij}=x_i(u_j)$, $i=1,..., p$ and $z_j=z(u_j)$.

Let

$$Y=\sum_{j=1}^{N} y_j=N\bar{Y},\ X_i=\sum_{j=1}^{N} x_{ij}=N\bar{X}_i,\ \underline{X}=(X_1 ..., X_p) \text{ and } Z=\sum_{j=1}^{N} z_j=N\bar{Z}.$$

We want to estimate Y using information on z and $\underline{x}=(x_1,..., x_p)$.

It is well known that in sampling designs the use of auxiliary information increases the precision of an estimator considerably. The auxiliary information may be used (*i*) at the pre-selection stage *e.g.* in stratifying the population (*ii*) at the selection stage *e.g.* in selecting the units with unequal probabilities and (*iii*) at the estimation stage *e.g.* in forming the estimators such as ratio, regression, difference or product estimators.

We may utilize the auxiliary information in mixed ways also; for example, by selecting the units with probabilities proportional to a suitable measure of size (*pps*), based on a variate z, and by forming

---

*This paper is based on a part of the author's unpublished Ph.D. Thesis submitted to Punjabi University, Patials, India, 1970.

The work was financially supported by the Punjabi University, Patiala.

**Now at Research and Training School, Indian Statistical Institute, Calcutta.

ratio, regression, difference or product estimator using the information on yet other auxiliary variate (s) $\underline{x}$ where $\underline{x}$ may be a vector or a scaler.

Tripathi (1969) presented a regression type estimator for Y in *pps wr* sampling when information on z is readily available and the population total X of a character x is known. Also he (1970, 1973) developed a unified technique of double sampling for regression method of estimation and extended the above results to the situations :

(*a*) information on z is available but X is not known and

(*b*) neither the information on z is readily available nor X is known. Further the author (1968) has proposed multivariate ratio and difference estimators when information on z is readily available and the population totals $(X_1, \ldots . X_p)$ are known. In this paper, we develop a general double sampling scheme for multivariate ratio and difference methods of estimation and present the results for the situations similar to (*a*) and (*b*) cited above.

Let $\{\alpha_i\}$, $i=1, 1 \ldots, p$, be p-functions of sample observations which could be used as estimators for a population parameter θ. Let $w=(w_1, \ldots, wp)$ be a weight vector such that

$$\sum_{i=1}^{p} w_i = 1.$$

Proof of the following result is straightforward and will not be given here.

LEMMA 1 : The bias B (d) and the mean square error (MSE) $M(d)$ of the estimator

$$d = \sum_{i=1}^{p} w_i \alpha_i,$$

for θ, are given by

$$B(d) = \sum_{i=1}^{p} w_i \, b_i = wb'$$

and

$$M(d)=\sum_{i}^{p}\sum_{k}^{p} w_i w_k a_{ik}=wAw'$$

respectively, where $b_i=E(\alpha_i-\theta)$

$a_{ik}=E(\alpha_i-\theta)(\alpha_k-\theta)$ and $A=(a_{ik})$ $i, k=1,\ldots,p$. A is a symmetric positive definite matrix. The condition of positive definiteness of A is necessary to avoid non-degeneracy of the p α_i-estimators [Tikkiwal, 1960, Lemma 2.1, *p*. 132].

The results in lemmas 2 and 3 below are true in general and will not be proved here. Olkin (1958) proved them in equal probability selection while dealing with a multivariate ratio estimator for y.

Lemma 2 : The vector of weights $w=(w_1,\ldots,w_p)$ which minimizes $M(d)$, subject to the restriction $\sum_1^p w_i=1$, is given by $w_{opt}=eA^{-1}/eA^{-1}e', e=(1,\ldots,1) : 1xp$, and then the resulting bias and *MSE* are given by

$$B_0(d)=(eA^{-1}/eA^{-1}e')b'$$

and $$M_0(d)=1/eA^{-1}e'=1/\Sigma_i\Sigma_k a^{ik}\ 1, k=1,\ldots,p$$

respectively, where a^{ik} is the (I, k)th element of A^{-1}.

Lemma 3 : Let $M_0(d/p)$ denote the optimum *MSE* of d obtained by using $p\alpha$'s. Then

$$M_0(d/q)\leqslant M_0(d/p) \text{ for } q>p.$$

Unless otherwise stated, we shall consistently employ the following notations:

$P_j=z_j/Z$: probability of selecting the jth unit of the population, at each draw $(j=1,\ldots,N)$.

$$C(x_i,x_k)=\sum_{j=1}^{N}P_j(x_{ij}/P_j-X_i)(x_{kj}/P_j-X_k);\ R_i=Y/X_i$$

$$V(x_i)=C(x_i,x_i)=\sigma^2(x_i);\ V(y)=C(y,y)=\sigma^2(y)$$

$\delta_{ik}=C(x_i,x_k)/\sigma(x_i)\sigma(x_k)$ $i,k=0,1,\ldots,p$ where x_0 denotes y.

$C_{ik}=C(x_i,x_k)/X_i X_k$; $C^2_0=C_{oo}=V(y)/Y^2$

$C_i=\sigma(x_i)/X_i$, $i=0,1,\ldots,p$.

$\delta u=(u-Eu)/Eu$: delta notation, where u is a random variable.

E_1 (.), V_1 (.), C_1 () denote the unconditional and E_2 (.), V_2 (.), C_2(..) denote the conditional (given the first-phase sample) mean, variance and covariance respectively. Symbols

$$\hat{V}_1, \hat{V}_2 \text{ and } \hat{C}_2$$

will denote unbiased estimators of V_1, V_2 and C_2 respectively.

2. A General Double Sampling Scheme for Multivariate Difference and Ratio Methods of Estimation

To estimate Y, the general multivariate difference and ratio estimators for any probability sampling design (*psd*) may be defined by

$$\hat{Y}_d=\sum_{i=1}^{p} w_i\, \alpha_i,\ \alpha_i=\hat{Y}-\lambda_i\,(\hat{X}_i-X_i) \tag{2.1}$$

and

$$\hat{Y}_R=\sum_{i=1}^{p} w_i\alpha_i,\ \alpha_i=(\hat{Y}/\hat{X}_i)X_i \tag{2.2}$$

respectively, where

$$\hat{Y} \text{ and } \hat{X}_1$$

are unbiased estimators of Y and X_i respectively based on any *psd* and λ_i's are suitably chosen constants.

The estimators due to Takkiwal (1967) and Raj (1965*a*) are of the form (2.1) and that due to Ulkin (1958) is of the form (2.2).

In $\hat{Y}_d$ and $\hat{Y}_R$ defined above X_i's ($i=1,...,p$) are assumed to be known. When X_i's are not known but could be estimated by a large sample, rather inexpensively, we use the double sampling technique.

We select a preliminary large sample of n units at moderate cost according to a specified sampling design and then a subsequent small sample of m units according to a different (or same) sampling design. The second sample may either be a subsample of the first or independent of the first. Let $X_{i(1)}$ be an unbiased estimator of X_i based on the observations of the first phase sample alone and $X_{i(2)}$

and $Y_{(2)}$ be the unbiased estimators of X_i and Y respectively based on the observations of the second phase sample mainly. That is

$$EX_{i(2)}=E\,X_{i(1)}=X_i \text{ and } EY_{(2)}=Y \tag{2.3}$$

For the case of subsamples we shall further require that

$$E_2X_{i(2)}=X_{i(1)} \tag{2.4}$$

We note that in the case of subsamples and independent samples both,

$$C_2\,(X_{i(2)}, X_{i(1)})=C_2(Y_{(2)}, X_{i(1)})=0 \tag{2.5}$$

It is well known that in two-phase sampling, for any estimators d and d^* we have that

$$\text{Cov}\,(d, d^*)=E_1\,C_2\,(d, d^*)+C_1\,(E_2d, E_2d^*) \tag{2.6}$$

Following results can be easily proved with the help of Lemma 1 and the statements (2.3) to (2.6).

Theorem 1 : The multivariate difference estimator

$$Y_d=\sum_{i=1}^{p} w_i.\alpha_i,\ \alpha_i=Y_{(2)}-\lambda_i\,(X_{i(2)}-X_{i(1)}) \tag{2.7}$$

is unbiased for Y and the general expression for its variance is given by

$$V\,(Y_d)=wBw',\ B=(b_{ik})\ i, k=1,\dots, p$$

$$\begin{aligned} b_{ik}=\text{Cov}\,(\alpha_i, \alpha_k)=V(Y_{(2)})&-\lambda_i\,\{\text{Cov}\,(Y_{(2)}, X_{i(2)})-C_1\,(E_2Y_{(2)}, X_{i(1)})\}\\ &-\lambda_k\,\{\text{Cov}\,(Y_{(2)}, X_{k(2)})-C_1\,(E_2Y_{(2)}, X_{k(1)})\}\\ &+\lambda_i\lambda_k\,\{\text{Cov}\,(X_{i(2)}, X_{k(2)})-C_1\,(E_2X_{i(2)}, X_{k(1)})\\ &-C_1(E_2\,X_{k(2)}, X_{i(1)})+C_1\,(X_{i(1)}, X_{k(1)})\} \end{aligned} \tag{2.8}$$

Corollary 1.1. If the second-phase sample is a subsample of the first, then

$$\text{Cov}(\alpha_i, \alpha_k)=b_{ik}=V_1\,(Y_{(1)})+E_1V_2\,(Y_{(2)})-\lambda_iE_1C_2\,(Y_{(2)}, X_{i(2)})-\lambda_kE_1C_2\,(Y_{(2)}, X_{k(2)})+\lambda_i\lambda_k\,E_1C_2(X_{i(2)}, X_{k(2)}) \tag{2.9}$$

where $\qquad Y_{(1)}=E_2Y_{(2)}.$

The expression in (2.8) will be used to derive the results for independent samples.

It is to be noted that in the case of independent samples if $E_2\,Y_{(2)}=Y$ and $E_2X_{i(2)}=X_i$, that is, if $Y_{(2)}$ and $X_{i(2)}$ do not use any information on the first phase sample, then

$$\text{Cov}\,(\alpha_i, \alpha_k)=b_{ik}$$

$$=V\,(Y_{(2)})-\lambda_i\,\text{Cov}\,(Y_{(2)}, X_{i(2)})-\lambda_k\,\text{Cov}\,(Y_{(2)}, X_{k(2)})+\lambda_i\lambda_k\,\text{Cov}\,(X_{i(2)}, X_{k(2)})+\lambda_i\lambda_k\,C_1\,(X_{i(1)}, X_{k(1)}) \tag{2.10}$$

In case of subsamples an unbiased estimator of $V(Y_d)$ is given by

$$v(Y_d) = w\hat{B}w', \hat{B} = (\hat{b}_{ik})$$

$$\hat{b}_{ik} = \hat{V}_1(Y_{(1)}) + \hat{V}_2(Y_{(2)}) - \lambda_i \hat{C}_2(Y_{(2)}, X_{i(2)}) - \lambda_k \hat{C}_2(Y_{(2)}, X_{k(2)})$$
$$+ \lambda_i \lambda_k \hat{C}_2(X_{i(2)}, X_{k(2)})$$

where $\hat{V}_1(.)$ and $\hat{V}_2(.)$ are unbiased estimators of $V_1(.)$ and $V_2(.)$ respectively and $\hat{C}_2$'s are the unbiased estimators of C_2's.

In general, in the case of independent samples an unbiased estimator of $V(Y_d)$ may be obtained by substituting unbiased estimators of the parameters in (2.8) and (2.10).

A biased estimator of $V_0(Y_d) = 1/(eB^{-1}e')$ would be given by

$$v_0(Y_d) = 1/(e\hat{B}^{-1}e') = 1/\sum_{ik}\sum \hat{b}_{ik}.$$

A multivariate ratio estimator for Y would be defined as

$$Y_R = \sum_{i=1}^{p} w_i \alpha_i, \quad \alpha_i = (Y_{(2)}/X_{i(2)})\, X_{i(1)} \tag{2.11}$$

The estimator due to Khan and Tripathi (1967) is of this form.

In our discussion for the ratio estimator we shall assume that

$$E\,[(\delta Y_{(2)})^a\, (\delta X_{i(2)})^b\, (\delta X_{k(2)})^c\, (\delta X_{i(1)})^d\, (\delta X_{k(1)})^e]$$
$$\leqslant O\,(m^{-(a+b+c+d+e)/2}) \tag{2.12}$$

Theorem 2 : The multivariate ratio estimator Y_R is in general biased. Under the assumption (2.12) it is a consistent estimator of Y for large populations and if $|\delta X_{i(2)}| < 1$, its bias is given by

$$B(Y_R) = wC' + O(m^{-2}) \tag{2.13}$$

where C' is a column vector of

$$C_i^* = (1/X_i)\,[R_i\{V(X_{i(2)}) - C_1(E_2 X_{i(2)}, X_{i(1)})\}$$
$$- \text{Cov}\,(Y_{(2)}, X_{i(2)}) + C_1(E_2\, Y_{(2)}, X_{i(1)})]$$

and its MSE is given by

$$M(Y_R) = wAw' + O(m^{-2}),\ A = (a_{ik})\ i, k = 1, \ldots, p$$

$$a_{ik}=V(Y_{(2)})-R_i\{\text{Cov}\,(Y_{(2)},X_{i(2)})-C_1\,(E_2Y_{(2)},X_{i(1)})\}$$
$$-R_k\{\text{Cov}\,(Y_{(2)},\,X_{k(2)})-C_1\,(E_2Y_{(2)},\,X_{k(1)})\}+R_iR_k$$
$$\{\text{Cov}\,(X_{i(2)},\,X_{k(2)})-C_1\,(E_2X_{k(2)},\,X_{i(1)})-C_1\,(E_2X_{i(2)},\,X_{k(1)})+C_i\,(X_{i(1)},\,X_{k(1)})\} \qquad (2.14)$$

Corollary 2.1. In the case of subsamples

$$C_i^*=(1/X_i)\,[R_iE_1V_2\,(X_{i(2)})-E_1C_2\,(Y_{(2)},\,X_{i(2)})] \qquad (2.15)$$
$$a_{ik}=V_1(Y_{(1)})+E_1V_2(Y_{(2)})-R_iE_1C_2\,(Y_{(2)},\,X_{i(2)})$$
$$-R_kE_1C_2\,(Y_{(2)};\,X_{k(2)})+R_iR_kE_1C_2\,(X_{i(2)},\,X_{k(2)}) \qquad (2.16)$$

For large m, so that the terms $O(m^{-2})$ may be neglected, the optimum weights, bias and MSE may be obtained by using Lemma 2.

Remark : From Theorems 1 and 2 we observe that $V(Y_d)=M_a(Y_R)$ provided $\lambda_i=R_i$ where $M_a(Y_R)$ is the large sample approximation of the MSE of Y_R. In case of simple random sampling when λ_i is a good guess of R_i based on the past experience, Raj (1965*a*) has illustrated by a numerical example that provided the departure of λ_i from R_i is moderate the variance of difference estimator varies a little from the large sample approximation of MSE of ratio estimator.

3. Double Sampling For Multivariate Ratio And Difference Estimators With *pps wr* Selection

In case the population totals $X_i(i=1,\dots,p)$ are unknown but the information on z is readily available we follow the selection procedure given by Raj (1965*b*).

We select the first phase sample $\{x_{ij},\dots,xp_j\}$, $j=1,\dots,n$, of n units with replacement and with probabilities $P_j\,(j=1,\dots,N)$, $\Sigma P_j=1$, based on the character z (*pps wr*) at moderate cost and then a subsequent subsample $\{y_j\}$, $j=1,\dots,m$, of m units with equal probabilities and without replacement. If the second phase sample is to be independent of the first it is selected with *pps wr* in which y and $(x_1,\dots,x_p)$ are observed. For convenience of reference we shall denote this sampling procedure by D_1.

$$\text{Let } X_{i(1)}=(1/n)\sum_{j=1}^{n}x_{ij}/p_j,\ \text{where } p_j=z_j/\sum_{j=1}^{N}z_j \qquad (3.1)$$

$$X_{i(2)}=(1/m)\sum_{j=1}^{m} x_{ij}/P_j \tag{3.2}$$

and

$$Y_{(2)}=(1/m)\sum_{j=1}^{m} y_j/p_j \tag{3.3}$$

The estimators $X_{i(1)}$, $X_{i(2)}$ and $Y_{(2)}$ obviously satisfy the conditions (2.3) and (2.4). The proofs of the following lemmas are easily obtainable and will not be given here.

Lemma 4 : In the sampling scheme D_1 if the second-phase sample is a subsample of the first, we have

(*i*) $V_1(Y_{(1)})=(1/n)\,V(y)$

where

$$Y_{(1)}=E_2 Y_{(2)}=(1/n)\sum_{j=1}^{n} y_j/P_j.$$

(*ii*) $$C_2(X_{i(2)}, X_{k(2)})=(1/m-1/n)\sum_{j=1}^{n}(x_{ij}/P_j-X_{i(1)})(x_{kj}/P_j-X_{k(1)})/(n-1)$$

(*iii*) $$V_2(X_{i(2)})=(1/m-1/n)\sum_{j=1}^{n}(x_{ij}/P_j-X_{i(1)})^2/(n-1)$$

(*iv*) $E_1C_2(X_{i(2)}, X_{k(2)})=(1/m-1/n)\,C(x_i, x_k)$

(*v*) $E_1V_2\,(X_{i(2)})=(1/m-1/n)\,V\,(X_i)$

(*vi*) $$\hat{V}_1(Y_{(1)})=(1/n)\sum_{j=1}^{m}(y_j/p_j-Y_{(2)})^2/(m-1)$$

(*vii*) $$\hat{C}_2(X_{i(2)}, X_{k(2)})=(1/m-1/n)\sum_{j=1}^{m}(x_{ij}/p_j-X_{i(2)})(x_{kj}/p_j-X_{k(2)})/(m-1)$$

(*viii*) $\hat{V}_2(X_{i(2)})=(1/m-1/n)\sum_{j=1}^{m}(x_{ij}/P_j-X_{i(2)})^2/(m-1)$

Lemma 5 : In the sampling design D_1 if the second phase sample is independent of the first, we have

(*i*) $\text{Cov}\,(X_{i(2)}, X_{k(2)})=(1/m)\,C(x_i, x_k)$

(*ii*) $V\,(X_{i(2)})=(1/m)\,V(x_i)$

(*iii*) $\text{Cov}\,(X_{i(2)}, X_{k(1)})=C_1\,(E_2X_{i(2)}, X_{k(1)})=0$

(*iv*) $\text{Cov}\,(X_{i(1)}, Y_{(2)})=C_1\,(X_{i(1)}, E_2\,Y_{(2)})=0$

(*v*) $\text{Cov}\,(X_{i(1)}, X_{k(1)})=(1/n)\,C(x_i, x_k)$

(*vi*) $\hat{C}(X_{i(2)}, X_{k(2)})=(1/m)\sum_{j=1}^{m}(x_{ij}/P_j-X_{i(2)})\,(X_{kj}/P_j-X_{k(2)})/(m-1)$

(*vii*) $\hat{V}\,(X_{i(2)})=(1/m)\sum_{j=1}^{n}(x_{ij}/P_j-X_{i(2)})^2/(m-1)$

(*viii*) $\hat{C}_1(X_{i(1)}, X_{k(1)})=(1/n)\sum_{j=1}^{n}(x_{ij}/P_j-X_{i(1)})(x_{kj}/P_j-X_{k(1)})/(n-1)$

Using the results of section 2 and the lemmas given above we obtain the following results in D_1.

Variance of the differrnce estimator would be given by

$$V(Y_d)=wBw',\quad B=(b_{ik}),\ i,k=1,\dots,p$$

where in the case of subsamples

$$b_{ik}=(1/m)\,V(y)-(1/m-1/n)\,[\lambda_i\,C(y, x_i)+\lambda_k\,C(y, x_k)-\lambda_i\,\lambda_k\,C(x_i, x_k)] \qquad (3.4)$$

with

$$\hat{b}_{ik}=\{m(m-1)\}^{-1}\sum_{j=1}^{m}(y_j/P_j-Y_{(2)})^2-(1/m-1/n)\Big[\lambda_i\sum_{j=1}^{m}(y_j/P_j-Y_{(2)})\,(x_{ij}/P_j-X_{i(2)})+\lambda_k\sum_{j=1}^{m}(y_j/P_j-Y_{(2)})(x_{kj}/P_j-X_{k(2)})$$

$$-\lambda_1 \lambda_k \sum_{j=1}^{m} (x_{ij}/P_j - X_{i(2)})(x_{kj}/P_j - X_{k(2)})\Big]/(m-1) \quad (3.5)$$

and in the case of independent samples

$$b_{ik} = (1/m)\,[V(y) - \lambda_i\, C(y, x_i) + \lambda_k\, C(y, x_k) + \lambda_i \lambda_k\, C(x_i, x_k)] + (1/n)\,\lambda_i \lambda_k\, C(x_i, x_k) \quad (3.6)$$

with

$$\hat{b}_{ik} = \{m(m-1)\}^{-1}\Big[\sum_{j=1}^{m} (y_j/P_j - Y_{(2)})^2 - \lambda_i \sum_{j=1}^{m} (y_j/P_j - Y_{(2)})$$

$$(x_{ij}/P_j - X_{i(2)}) - \lambda_k \sum_{j=1}^{m} (y_j/P_j - Y_{(2)})(x_{kj}/P_j - X_{k(2)})$$

$$+\lambda_i \lambda_k \sum_{j=1}^{m} (x_{ij}/P_j - X_{i(2)})(x_{kj}/P_j - X_{k(2)})\Big]$$

$$+\lambda_i \lambda_k \{n(n-1)\}^{-1} \sum_{j=1}^{n} (x_{ij}/P_j - X_{i(1)})(x_{kj}/P_j - X_{k(1)}) \quad (3.7)$$

where $\hat{b}_{ik}$ is an unbiased estimator of b_{ik}.

Bias of the ratio estimator Y_R in D_1 is given by

$$B(Y_R) = (1/m - 1/n)\, Ywd' + O(m^{-2})$$

if the second sample is a subsample and by

$$B(Y_R) = (1/m)\; Ywd' + O(m^{-2})$$

in the case of independent samples, where

$$d_i = C_i^2 - C_{oi}$$

A sample estimate of $B(Y_R)$ is given by

$$\hat{B}(Y_R) = (1/m - 1/n)\, w\hat{d}'$$

$$\hat{d}_i = (1/X_{i(2)})\Big[r_i \sum_{j=1}^{m} (x_{ij}/P_j - X_{i(2)})^2$$

$$-\sum_{j=1}^{m} (x_{ij}/P_j - X_{i(2)})\;(y_j/P_j - Y_{(2)})\Big]/(m-1)$$

in the case of subsamples and is given by

$$\hat{B}(Y_R) = (1/m)\, w\hat{d}'$$

in the case of independent samples, where $r_i = Y_{(2)}/X_{i(2)}$.

Mean square error and an estimate of it are given by

$$M(Y_R) = wAw' + O(m^{-2}) \tag{3.8}$$

and

$$m(Y_R) = w\hat{A}w',\ \hat{A} = (\hat{a}_{ik})$$

where for the case of subsamples and independent samples a_{ik} is given by (3.4) and (3.6) respectively with λ_1 replaced by R_i and $\hat{a}_{ik}$ is given by (3.5) and (3.7) respectively with λ_i replaced by r_i.

It is found that

$$E\ \hat{B}(Y_R) = B_a(Y_R) + O(m^{-2}) \text{ and } E\, m\, Y_R = M_a(Y_R) + O(m^{-2})$$

where $B_a(Y_R)$ and $M_a(Y_R)$ are bias and *mse* of Y_R to the terms of order m^{-1}.

4. Double Sampling For Inclusion Probabilities and Multivariate Difference and Ratio Estimators

If the information on z is not avaiable and population totals X_i's are also unknown [situation similar to (*b*) mentioned in section 1] we select a first phase sample $(z_j, x_{ij}, \ldots, x_{pj})$, $j=1, \ldots, n$ of n units with equal probabilities and without replacement and second phase sample $(y_j, x_{1j}, \ldots, x_{pj})$, $j=1, \ldots, m$ of m units with probabilities $p_j = z_j / \sum_{j=1}^{n} z_j$ $(j=1, \ldots, n)$ and with replacement. We denote this sampling scheme by D_2, which was first given by Raj (1964) and then by Singh and Singh (1965).

In this case we define

$$X_{i(1)} = N\bar{x}_i,\ X_{i(2)} = (N/nm)\sum_{j=1}^{m} X_{ij}/P_j$$

and

$$Y_{(2)}=(N/nm)\sum_{j=1}^{m} y_j/P_j \tag{4.1}$$

Following results are easily proved.

Lemma 6. In D_2 the second phase sample is a subsample of the first then

(i) $V_1(Y_{(1)})=N^2(1/n-1/N)\,S_y^2$ where $Y_{(1)}=E_2\,Y_{(2)}=N\bar{y}$

(ii) $C_2(X_{i(2)}, X_{k(2)})=(N^2/mn^2)\sum_{j=1}^{n} p_j\,(x_{ij}/p_j-X_{in})\,(x_{kj}/p_j-X_{kn})$.

(iii) $V_2(X_{i(2)})=(N^2/mn^2)\sum_{j=1}^{n} p_j(x_{ij}/p_j-X_{in})^2$

(iv) $E_1C_2(X_{i(2)}, X_{k(2)})=\dfrac{(n-1)N}{n(-1)m}\,C(x_i, x_k)$

(v) $\hat{V}_1(Y_{(1)})=N^2(1/n-1/N)\;s_y^2$

(vi) $\hat{C}_2(X_{i(2)}, X_{k(2)})=(N^2/mn^2)\;\hat{C}^*(x_i, x_k)$

(vii) $\hat{V}_2(X_{i(2)})=(N^2/mn^2)\;\hat{V}^*(x_i)$

where

$$Y_n=\sum_{j=1}^{n} y_j=n\bar{y},\quad X_{in}=\sum_{j=1}^{n} x_{ij}=n\bar{x}_i.\quad S_y^2$$

$$=\sum_{j=1}^{N}(y_j-\bar{Y})^2/(N-1)$$

$$m\,(n-1)\;S_y^2=\sum_{j=1}^{m} y_j^2/p_j-\sum_{j\neq e}^{m}(y_j/p_j)\,(y_e/p_e)/n(m-1)$$

$$\hat{C}^*(x_i, x_k)=\sum_{j=1}^{m}(x_{ij}/p_j-\hat{x}_i)\,(x_{kj}/p_j-\hat{x}_k)/(m-1)$$

$$\hat{x}_i = (1/m) \sum_{j=1}^{m} x_{ij}/p_j \ ; \qquad \hat{V}^* (x_i) = \hat{C}^*(x_i, x_i)$$

Lemma 7 : In D_2, in case of independent samples we have

(*i*) $$\text{Cov}\ (X_{i(2)}, X_{k(2)}) = (1/m) \left\{ \left(\frac{1}{n} - \frac{1}{N} \right) C_z^2 + 1 \right\}$$

$$C\ (x_i, x_k) + R_{2i} R_{2k}\ N^2 \left(\frac{1}{n} - \frac{1}{N} \right) S^2_z$$

(*ii*) $\text{Cov}\ (X_{i(2)}, X_{k(1)}) = C_1(E_2 X_{i(2)}, X_{k(1)}) = N^2(1/n - 1/N)\ R_{2i}\ S_{kz}$

(*iii*) $\text{Cov}\ (X_{i(1)}, X_{k(1)}) = N^2\ (1/n - 1/N)\ S_{ik}$

(*iv*) $$\hat{C}(X_{i(2)}, X_{k(2)}) = \frac{N^2}{nm\,(m-1)} \left[\sum_{j=1}^{m} (x_{ij}\ x_{kj}/z^2_j) - m\ \bar{r}_{2i}\ \bar{r}_{2k} \right]$$

$$\left[n\bar{z}^2 - (1 - n/N)\, s_z^2 \right] + N(N-n)\, \bar{r}_{2i}\ \bar{r}_{2k}\ s_z^2/n$$

(*v*) $\hat{C}(X_{i(2)}, X_{k(1)}) = \hat{C}_1(E_2\ X_{i(2)}, X_{k(1)}) = N^2(1/n - 1/N)\ \bar{r}_{2i}\ s_{kz}$

(*vi*) $\hat{C}(X_{i(1)}, X_{k(1)}) = N^2\ (1/n - 1/N)\ s_{ik}$

where

$$S_{iz} = \sum_{j=1}^{N} (x_{ij} - \bar{X}_i)\ [z_j - \bar{Z})/(N-1),\ R_{2i} = X_i/Z,\ C_z^2 = S_z^2/\bar{Z}^2$$

$$S_{ik} = \sum_{j=1}^{N} (x_{ij} - \bar{X}_i)\ (x_{kj} - \bar{X}_k)/(N-1)$$

$$\bar{r}_{2i} = (1/m) \sum_{j=1}^{m} x_{ij}/z_j,\ s_{ik} = \sum_{j=1}^{m} (x_{ij} - \bar{x}_i)\ (x_{kj} - \bar{x}_k)/(m-1)$$

$$S_{kz} = \sum_{j=1}^{m} (x_{kj} - \bar{x}_k)\ (z_j - \bar{z})/(n-1) \qquad (4.3)$$

Results of the Section 2 and above lemmas yield the following results in D_2.

$$V(Y_D) = \sum_{i=1}^{p} \sum_{k=1}^{p} w_i w_k\ b_{ik} \qquad (4.4)$$

where in the case of subsamples

$$b_{ik} = \frac{(n-1)N}{(N-1)nm} \left[V(y) - \lambda_i\ C(y, x_i) - \lambda_k\ C(y, x_k) + \lambda_i \lambda_k\ C(x_i, x_k) \right] + (N/n)\ (N-n) s^2_y \qquad (4.5)$$

with

$$\hat{b}_{ik}=(N^2/mn^2)\ [\ \hat{V}_*(y)-\lambda_i\hat{C}_*(y,\ x_i)-\lambda_k\ \hat{C}_*(y,\ x_k)+\lambda_i\lambda_k\ \hat{C}_*(x_i,\ x_k)\ ]$$
$$+\ (N/n)\ (N-n)\ s^2_y \tag{4.6}$$

and in the case of independent samples

$$b_{ik}=(1/m)\ [V(y)-\lambda_i\ C(y,\ x_i)+\lambda_i\lambda_k\ C(x_i,\ x_k)-\lambda_k\ C(y,\ x_k)\ ]$$
$$\left\{1+\left(\frac{1}{n}-\frac{1}{N}\right)\ C_z^2\ \right\}+\left[\left(\ R_1^{*2}-R_1^*\ R_{2i}\ \lambda_i-R_1^*R_{2k}\lambda_k\right.\right.$$
$$\left.+\lambda_i\lambda_k R_{2i}R_{2k}\right)\ S_z^2$$
$$+(R_1^*\ \lambda_i-\lambda_i\ \lambda_k\ R_{2k})\ S_{iz}+(R_1^*\ \lambda_k-\lambda_i\ \lambda_k\ R_{2i})\ S_{kz}\ +\lambda_i\lambda_k\ S_{ik}]$$
$$(1/n-1/N)\ N^2 \tag{4.7}$$

with $\hat{b}_{ik}$ obtained by substituting unbiased estimators of various terms in the expansion (2.8) from Lemma 7; $R_1^*=Y/Z$.

In the case of subsamples and independent samples

$$B(Y_R)=\frac{(n-1)N}{(N-1)nm}Ywd'\ +0(m^{-2}) \tag{4.8}$$

and $$B(Y_R)=(1/m)\left\{1+\left(\frac{1}{n}-\frac{1}{N}\right)\ C_z^2\right\}\ Ywb'+0\ (m^{-2}) \tag{4.9}$$

respectively, where $d_i=\ C_i^2\ -C_{oi}$.

$$M(Y_R)=wAw'+0(m^{-2}) \tag{4.10}$$

where in the case of subsample a_{ik} is given by b_{ik} in (4.5) with λ_i replaced by R_i while in the case of independent samples

$$a_{ik}=(1/m)\ [V(y)-R_iC\ (y,\ x_i)-R_k\ C(y,\ x_k)+R_iR_k\ C(x_i,\ x_k)].$$
$$\left\{1+(1/n-1/N)\ C_z^2\ \right\}+N^2\ (1/n-1/N)\ R_iR_k\ S_{ik}.$$

In the case of subsamples, for large m, a set of estimators for bias and MSE to the terms of order $(1/m)$ would be

$$\hat{B}(Y_R)=\hat{w}\hat{g}',\ \hat{g}_i=(N/nm)\ (1/\hat{x}_i)\ [(\hat{y}/\hat{x}_i)\ \hat{V}^*(x_i)-\hat{C}^*\ (y,\ x_i)]$$

$$m(Y_R)=w\hat{A}w'$$

where $\hat{y}=(1/m)\ \sum_{i=1}^{m}\ y_j/p_j$ and $\hat{a}_{ik}$ is $\hat{b}_{ik}$ in (4.6) with λ_i replaced by $(\hat{y}/\hat{x}_i)$.

Henceforth we shall consider $M\ (Y_R)$ to the terms of order $(1/m)$.

5. Optimum n, m And Resulting MSE

Let a cost function be defined as

$$C=a_o+nC_n+mC_m \tag{5.1}$$

where C is the total cost, a_o is over head cost and C_n and C_m are the cost per unit in the first phase sample and the second phase sample respectively.

Theorem 3 : The values of n and m which minimize $M(Y_R)$ for the case of subsamples in D_t for a fixed cost of the form (5.1), are given by

$$n_{opt}=(C-a_o)\ (wHw')^{1/2}/\ [(c_n)^{1/2}\ \{(wGw'c_m)^{1/2}+(wHw'c_n)^{1/2}\}\ ]$$

and

$$m_{opt}=(C-a_o)\ (wGw')^{1/2}/[\ (c_m)^{1/2}\ \{\ (wGw'c_m)^{1/2}+(wHw'c_n)^{1/2}\}\ ]$$

respectively and then the resulting MSE is given by

$$M_{(opt)}\ (Y_R)=[((wGw')\ c_m)^{1/2}+((wHw')\ c_n)^{1/2}]^2/C-a_0)$$

where $H=(h_{ik}),\ G=(g_{ik}),\ i,\ k=1,\ \dots\ ,\ p$

$$h_{ik}=R_i\ C\ (y,\ x_i)+R_k\ C\ (y,\ x_k)-R_iR_k\ C\ (x_i,\ x_k) \tag{5.2}$$

and $$g_{ik}=V(y)-R_i\ C(y,x_i)-R_k\ C(y,\ x_k)+R_iR_k\ C(x_i,\ x_k) \tag{5.3}$$

Proof : From (3.8) and (3.4), to the terms of order $(1/m)$ we have

$$a_{ik}=(1/n)\ h_{ik}+(1/m)\ g_{ik}$$

whence $$M\ (Y_R)=(1/n)\ wHw'+(1/m)\ wGw'.$$

Following Cochran [(1963) p. 331] the results are easily arrived at.

For the difference estimator, h_{ik} and g_{ik} in above theorem will be given by replacing R_i by λ_i.

In case of independent samples, for ratio estimator we replace h_{ik} in (5.2) by

$$h_{ik}=R_iR_k\ C(x_i,\ x_k) \tag{5.4}$$

and for the the difference estimator g_{ik} and h_{ik} are given by (5.3) and (5.4) respectively with R_i replaced by λ_i.

For large n we may replace the terms $(n-1)/n$ and $N/(N-1)$ by unity giving, from (4.10),

$$M\Big(\ Y_R\Big)=(1/n)N^2\ S_y^2\ +(1/m)\ wGw'-NS_y^2$$

in the case of subsamples in the sampling scheme D_2. In this case

$$n_{opt}=(C-a_0)NS_y\,[c_n^{1/2}\,\{(wGw')^{1/2}\,c_m^{1/2}+NS_y\,c_n^{1/2}\}]^{-1}$$

$$m_{opt}=(C-a_0)\,(wGw')^{1/2}\,[c_m^{1/2}\,\{(wGw')^{1/2}\,c_m^{1/2}+NS_y\,c_n^{1/2}\}]^{-1}$$

$$M_{(opt)}\,(Y_R)=[(wGw')^{1/2}\,c_m^{1/2}+NS_y\,c_n^{1/2}]^2/(C-a_0)-NS_y^2 \qquad (5.5)$$

Similar results are obtained for the difference estimator in the case of subsamples in D_2.

6. Comparison Of Y_R With Other Estimators

The following comparisons are restricted to the case of subsamples in the sampling scheme D_1.

From (3.8), to the terms $0(m^{-1})$,

we have

$$M\,(Y_R)=Y^2\sum_i^p\sum_k^p w_i\,w_k\,d_{ik}$$

$$d_{ik}=(1/m)\,C_0^2-(1/m-1/n)[\delta_{0i}\,C_0\,C_i+\delta_{0k}\,C_0\,C_k-\delta_{ik}\,C_i\,C_k] \qquad (6.1)$$

From Lemma 3 it is obvious that when the optimum weights are used,

$$M_0\,(Y_R/q)<M_0\,(Y_R/p),$$

where

$$M\,(Y_R/q)$$

represents $M\,(Y_R)$ when the information on the characters

$$x_i,\,\ldots,\,x_p,\,x_p+1,\,\ldots,\,x_q$$

is used. In particular if

$$C_i=c,\ \delta_{0i}=\delta_0\,;\ \delta_{ik}=\delta(i\neq k)\ i,k=1,\,\ldots,\,p \qquad (6.2)$$

then the optimum weights reduce to uniform weights $w_i=1/p$ and from (6.1)

$$M\,(Y_R)=Y^2\,[(1/m-1/n)\,c^2\,(1-\delta)/p+(1/m)\,C_0^2-(1/m-1/n)\,(2\delta_0\,C_0\,c-\delta\,c^2)] \qquad (6.3)$$

giving

$$M\,(Y_R\mid p)-M\,(Y_R\mid q)=Y^2\,(1/m-1/n)\,c^2\,(1-\delta)\,(q-p)/qp.$$

obviously, if

$$q>p \text{ then } M(Y_R \mid q) \leqslant M(Y_R \mid p);$$

equality being hold when $\delta=1$.

Theorem 4 : Let the cost function be of the form (5.1) and the conditions (6.2) be satisfied. The inclusion of $(q-p)$ more x's will result in higher precision, that is

$$M_{(opt)}(Y_R \mid q) < M_{(opt)}(Y_R \mid p),$$

if

$$(c_m)^{1/2} \geqslant [(V_{q1}\, c_{nq})^{1/2}-(V_{p1}\, c_{np})^{1/2}]\, [(V_{p2})^{1/2}-(V_{q2})^{1/2}]^{-1} \quad (6.4)$$

where c_{np} is C_n, p denoting that information on $(x_1, \ldots, x_p)$ is being used;

and

$$V_{p1}=[2\delta_0\, C_0\, c-\delta\, c^2-c^2(1-\delta)/p]\, Y^2 \quad (6.5)$$

$$V_{p2}=[C_0^2\; -2\delta_0\, C_0\, c+\delta\, c^2+(1-\delta)\, c^2/p]\, Y^2. \quad (6.6)$$

Proof :

From (6.3),

$$M(Y_R/p)=(1/n)\, V_{p1}+(1/m)\, V_{p2}.$$

Following Cochran [(1963), p. 331], using optimum m and n we get

$$M_{(opt)}\,(Y)_R \mid p)=[(V_{p2}\, C_{mp})^{1/2}+(V_{p1}\, C_{np})^{1/2}]^2/\,(C-a_0) \quad (6.7)$$

Noting that

$$C_{mp}=C_{mq}=c_m$$

since only y is measured on the subsample,

we have that

$$M_{(opt)}\,(Y_R \mid q) \leq M_{(opt)}\,(Y_R \mid p)$$

if

$$(V_{q2}\, c_m)^{1/2}+(V_{q1}\, C_{nq})^{1/2} \leq (V_{p2}\, c_m)^{1/2}+(V_{p1}\, C_{np})^{1/2}$$

which reduces to (6.4).

DOUBLE SAMPLING VERSUS SINGLE SAMPLING :

If all resources are devoted to a single sample with equal probabilities, a simple cost function would be

$$C=a_0+m_s\, C_m \quad (6.8)$$

where the symbols C and a_0 are defined in (5.1), C_m being the cost per unit spent in observing the required character y. The size of this single sample would be $m_s=(C-a_0)/C_m$. We may then form the simple estimator $Y_{eq}=N\bar{y}$; and since information on z is readily available, assuming the population total Z to be known we may form ratio estimator :

$$Y_{Req}=(\bar{y}/\bar{z})\,Z$$

and

regression estimator :

$$Y_{rgeq}=N\bar{y}-b\,(N\bar{z}-Z).$$

where $\bar{y}$ and $\bar{z}$ are the simple random means and b is the sample regression coefficient of y on z.

For sample size $m_s=(C-a_0)/C_m$

$$V(Y_{eg})=N^2S_y^2\,/m_s=C_m\,V_s/(C-a_0) \tag{6.9}$$

where

$$V_s=N_1\,S_y^2\,,\;S_y^2=\sum_{j=1}^{N}\,(y_j-\bar{Y})^2/(N-1).$$

Similarly assuming the cost per unit of observing z negligible we have that for large samples,

$$V\,(Y_{Req})=(1/m_s)\,V_{rat}=C_m\,V_{rat}/(C-a_0) \tag{6.10}$$

and

$$V\,(Y_{rgeq})=(1/m_s)\,V_{reg}=C_m\,V_{reg}/(C-a_0) \tag{6.11}$$

where

$$V_{rat}=N^2\Big[\,S_y^2+(Y/Z)^2\,S_z^2-2(Y/Z)\,\rho_{yz}\,S_y\,S_z\Big]$$

and

$$V_{reg}=N^2\left((1-\rho_{yz}^2\right)\,S_y^2$$

(Cochran : 1963).

Theorem 5 : Let the cost function be of the form (5.1). The procedure of double sampling for multivariate ratio method of estimation with pps selection gives a smaller variance than the simple random sampling and ratio and regression methods of estimation in single-phase sampling with equal probability selection

if

$$(C_m/C_n)>(V_1/V)\,[1-(V_2/V)^{1/2}]^{-2},\;V_2<V \tag{6.12}$$

where

$$V = V_s,\ V = V_{rat}$$

and

$$V = V_{reg} \text{ respectively};$$

and

$$V_1 = wHw' \text{ and } V_2 = wGw'$$

are defined in the Theorem 3.

Proof : From the Theorem 3,

$$M_{(opt)}(Y_R) = [(V_2\, C_m)^{1/2} + (V_1\, C_n)^{1/2}]^2/(C - a_0) \tag{6.13}$$

From (6.12)

$$(C_m/C_n)^{1/2} > (V_1/V)^{1/2}\,[1 + V_2/V)^{1/2}]\,(1 - V_2/V)^{-1} \tag{6.14}$$

obviously

$$(C_m/C_n)^{1/2} > -(V_1/V)^{1/2}\,[1 - (V_2/V)^{1/2}]/(1 - V_2/V) \tag{6.15}$$

is always true. From (6.14) and (6.15)

$$(C_m/C_n)^{1/2} > (V_1/V)^{1/2}\,[(V_2/V)^{1/2} \pm 1]/(1 - V_2/V)$$
$$= [2(V_1\, V_2/V^2)^{1/2} \pm \{4\,(V_1\, V_2/V^2) - 4(1 - V_2/V)$$
$$(-V_1/V)\}^{1/2}]/2\,(1 - V_2/V)$$

which imples (with $V_2 < V$) that

$$(C_m/C_n)\,(1 - V_2/V) - 2\,\{(V_1\, V_2/V^2)\,(C_m/C_n)\}^{1/2} - V_1/V > 0$$

or

$$C_m > (1/V)\,[(V_2 C_m)^{1/2} + (V_1\, C_n)^{1/2}]^2 \tag{6.16}$$

For

$$V = V_s,\ (6.16) \text{ implies}$$

that

$$M_{(opt)}(Y_R) < V\,(Y_{eq})$$

where

$$V\,(Y_{eq}) \text{ is defined by (6.9).}$$

Similarly for

$$V = V_{rat} \text{ and } V = V_{reg},$$

(6.16) implies that

$$M_{(opt)}(Y_R) < V\,(Y_{Req})$$

and

$$M_{(opt)}(Y_R) < V\,(y_{rgeq}) \text{ respectively.}$$

This proves the Theorem.

Further, since the information on z is readily available we may devote all resources to have a single sample selected with the probabilities P_j, $\sum_{j}^{N} p_j=1$, based on the character z and with replacement and may consider a simple estimator $Y'_{pps}=(1/m_s) \sum_{j=1}^{m_s} y_j/p_j$.

Using (6.8) the variance of this estimator would be

$$V(Y'_{pps})=(1/m_s)\, V(y)=c_m'\, V(y)/(C-a_0) \qquad (6.17)$$

where C_m' is per unit cost in this case. Comparing (6.13) with (6.17) we observe that

$$M_{(opt)}(Y_R) < V(Y'_{pps})$$

if

$$C'_m > [1/V(y)]\, [(V_2\, c_m)^{1/2}+(V_1\, c_n)^{1/2}]^2.$$

In the sampling scheme D_2 an unbiased and consistent estimator for Y given by Raj (1964) is

$$Y_D=(N/nm) \sum_{j=1}^{n} y_j/p_j,\ p_j=z_j/\sum_{j=1}^{n} z_j.$$

In the case of subsamples

$$V(Y_D)=(1/n)\, Y^2\left[\frac{N(n-1)}{(N-1)m}\, C_0^2+\frac{(N-n)}{N}\, C_y^2\right] \qquad (6.18)$$

From (4.10), in the case of subsamples

$$V(Y_R)=(1/n)\, Y^2 \sum_{i}^{p} \sum_{k}^{p} w_i\, w_k\, d_{ik}$$

$$d_{ik}=\frac{(n-1)N}{(N-1)m}\left[C_0^2-\delta_{0i}\, c_0\, c_i-\delta_{0k} c_0\, c_k+\delta_{ik}\, c_i\, c_k\right]$$

$$+\frac{N-n}{N}\, C_y^2 \qquad (6.19)$$

where

$$C_y^2=S_y^2/\bar{y}^2.$$

Under the uniform weighting conditions (6.2),

$$M(Y_R)=(Y^2/pn)\left[\frac{(n-1)N}{(N-1)m}\left\{C^2(1-\delta)+p\left(c_0^2-2\delta_0 cc_0+\delta c^2\right)\right\}+p\frac{N-n}{N}c_y^2\right]. \quad (6.20)$$

From (6.18) and (6.20) it follows that,

$$M(Y_R)\leq V(Y_D)$$

iff

$$p\,\delta_0\,c_0\,[c\,\{1+\delta\,(p-1)\}]^{-1}>1/2 \quad (6.21)$$

the condition (6.21) reduces to $\delta > 1/(p+1)$
if

$$\delta_0=\delta \text{ and } c_0=c.$$

Further from (6.20) we observe that

$$M(Y_R\mid p)-M(Y_R\mid q)=Y^2\frac{(n-1)N}{n(N-1)m}C^2(1-\delta)(q-p)/qp \quad (6.22)$$

By (5.5) it follows that $M_{(opt)}(Y_R\mid q)\leq M_{(opt)}(Y_R\mid p)$
if

$$[(C_{nq})^{1/2}-(C_{np})^{1/2}]C_m^{-1/2}\leq[(wG_p\,w')^{1/2}-(wG_q\,w')^{1/2}]/NS_y. \quad (6.23)$$

Summary

The problem considered is to estimate Y, the population total of a character y, in case the units are to be selected with probabilities proportional to a suitable measure of size and with replacement (pps wr) and the information on yet a p-dimensional vector $x=(x_1, \dots, x_p)$ of auxiliary characters is to be used to form multivariate difference-type and ratio-type estimators. A general double sampling scheme is developed for multivariate difference and ratio-type estimators and then the general results are used to derive mean and mean square errors of these estimators in two particular sampling schemes. Expressions for optimum sizes of first phase and second phase samples and the resulting optimum mean square errors are obtained and comparison of our estimators is made with other estimators.

Acknowledgements

The author is indebted to Dr. A.R. Kokan for fruitful discussions and constant encouragement while working at Muslim University

Aligarh. Thanks are also due to Dr. O.P. Srivastava, Professor and Head, Department of Statistics, Haryana Agricultural University, Hissar, for his valuable suggestions during the preparation of this paper. The author wishes to express his thanks to the referee whose suggestions led to the better presentation of the paper.

REFERENCES

1. Cochran, W.G. (1963) : Sampling Techniques. Second Edition, John Wiley and Sons Inc., New York.

2. Khan, S. and Tripathi, T.P., (1967) : The use of multivariate auxiliary information in double sampling. Jour. Ind. Stat. Assoc., Vol. 5, 43-48.

3. Olkin, I. (1958) : Multivariate ratio estimation for finite populations. Biometrika 45, 154-165.

4. Raj, D. (1964) : On double sampling for pps estimation. Ann. Math. Stat., 35, 900-902.

5. Raj, D. (1965*a*) : On a method of using multi-auxiliary information in sample surveys. J. Amer. Stat. Assoc., 60, 270-77.

6. Raj, D. (1965*b*) : On a sampling over two occasions with pps. Ann. Math. Stat., 36, 327-30.

7. Singh, D. and Singh, B.D. (1965) : Some contributions to two-phase sampling. Austral. J. Statist., 7, 45-47.

8. Tikkiwal, B.D. (1960) : On the theory of classical regression and double sampling estimation. JRSS ; Series B, 22, 131-38.

9. Tikkiwal, B.D. (1967) : Theory of multiphase sampling form a finite or an infinite population on successive occasions. Rev. Int. Stat. Inst., 35 : 3, 247-63. Based mainly on the theoretical part of the author's Ph. D. dissertation submitted to North Carolina State College, Releigh, N.C., USA in 1955.

10. Tripathi, T.P. (1968) : The use of multivariate auxiliary information in sample surveys when the units are selected with pps and with replacement. Presented at the 55th annual session of 'Indian Science Congress' held at Varanasi in early January 1968. Now sent for publication.

11. Tripathi, T.P. (1969) : A regression-type estimator in sampling with replacement. Austral. J. Statist., 11 (3), 140-148.

12. Tripathi, T.P. (1970) : Double sampling for regression method of estimation using pps wr selection. A portion of Chapter II of author's Ph. D. thesis, 'Some Contributions to the Sampling Theory Using Multivariate Information' submitted to Punjabi University, Patiala, India.

13. Tripathi, T.P. (1973) : Double sampling for inclusion probabilities and regression method of estimation. Jour. Indian Stat. Assoc., Vol. 10; 33-46.