6.1. Introduction

The estimates of population ratio of means (or totals) are most frequently needed in sample surveys. It is often of much interest to know the ratio of two characters. The ratio could be of total land leased out to total land possessed, total irrigated land to total cultivable land in a particular period or in different periods of time. In apple orchard surveys, for example, one may be interested to know the yield per acre, yield per fruit bearing tree or per tree etc.

The use of auxiliary information in formulating ratio estimators is well known in literature. Cochran (1940) and Lahiri (1951), among others, considered the estimation of ratio in univariate case, while Olkin (1958), Raj (1965a) utilized the multivariate auxiliary information in constructing a multivariate ratio estimator. Desraj (1954a) considered the variance and unbiased variance estimator of the ratio estimators in case of a multi stage design where the sample at the first stage is selected through PPS. Sukhatame (1954), Singh and
Singh (1965b), Tripathi (1970), Tikkiwal (1964) are among others to consider the problem in one way or the other.

In large scale surveys, while using two stage sampling design, the auxiliary information may not be readily available for every first stage unit in the population. The information on several auxiliary characters may be collected for the selected first stage units and then used for stratifying the second stage units and in constructing the suitable estimators leading to increased precision of the estimator.

In this chapter, such a situation has been considered.


To estimate the population ratio $R$, in the absence of information on auxiliary characters for every first stage unit (f.s.u.) and the use of information on several characters collected from first stage units and to construct the suitable estimators, we proceed as follows:

Let a sample $s(1)$ of size $n$ first stage units be drawn from a population of $N$ units using a suitable sampling procedure and the observations on auxiliary characters be made. Let from $j$th selected first stage units, a sample $s(2)_j$, ($j=1,...,n$) of second stage units be selected independently
Let $\hat{X}_i(1)$ be an unbiased estimator of $X_i$ based on $s(1)$ and $\hat{X}_i(2)$, $\hat{Y}(2)$ and $\hat{X}_o$ be the estimators of $X_i$, $Y$ and $X_o$ respectively based on $s(2) = \{s(2)', \ldots, s(2)^n\}$.

We define below an estimator $\mathbb{R}^*$ for estimating the population ratio as,

$$
\mathbb{R}^* = \frac{\sum_{i=1}^{p} w_i(1) \alpha_i(1)}{\sum_{i=1}^{p} w_i(2) \alpha_i(2)}
$$

where $\alpha_i(1) = \hat{Y}(2) - t_i(1)(\hat{X}_i(2) - \hat{X}_i(1))$

and $\alpha_i(2) = \hat{X}_o(2) - t_i(2)(\hat{X}_i(2) - \hat{X}_i(1))$

The weight vectors $w(1)' = (w_1, w_2, \ldots, w_p)$ and $w(2)' = (w_1', w_2', \ldots, w_p')$ are such that $w(1)' e$ and $w(2)' e$ are unity where $e' = (1, 1, \ldots, 1)_p x_1$ is a unit vector. $t_i(1)$ and $t_i(2)$'s are suitably chosen statistics.

By introducing the statistics $t_i(1)$ and $t_i(2)$ in the definition of $\mathbb{R}^*$, the class becomes quite wide and several interesting estimators may be identified as particular members of the class defined in (6.2.1).
To study the general properties of the class \( \mathcal{R}^* \) in (6.2.1), we assume that the statistics \( t_i^{(1)} \) and \( t_i^{(2)} \) for all \( i = 1, \ldots, p \) are such that

(i) \[ E t_i^{(1)} = \lambda_i^{(1)} + \varepsilon_i^{(1)} \]

(ii) \[ E t_i^{(2)} = \lambda_i^{(2)} + \varepsilon_i^{(2)} \]

where \( \varepsilon_i^{(1)} \) and \( \varepsilon_i^{(2)} \) are negligible for large samples.

Let us write

\[ a_i^{(1)} = a_i^{(1)} + (\hat{X}_{i(2)} - \hat{X}_{i(1)}) (\lambda_i^{(1)} - t_i^{(1)}) \]

\[ a_i^{(2)} = a_i^{(2)} + (\hat{X}_{i(2)} - \hat{X}_{i(1)}) (\lambda_i^{(2)} - t_i^{(2)}) \]

where

\[ a_i^{(1)} = \hat{Y}_{i(2)} - \lambda_i^{(1)} (\hat{X}_{i(2)} - \hat{X}_{i(1)}) \]

\[ a_i^{(2)} = \hat{X}_{o(2)} - \lambda_i^{(2)} (\hat{X}_{i(2)} - \hat{X}_{i(1)}) \]

Now to our order of approximation the bias and mean square error of the estimator in \( \mathcal{R}^* \) may be given by the
expressions in Chapter 5. In fact all results of Section 5.2 are applicable in this case also for large samples.

We now discuss two proposed sampling schemes $D_1$ and $D_2$ and their estimation procedures using two stage stratified sampling.

6.3. Proposed Sampling Scheme $D_1$.

To estimate the population ratio $R$ using stratified multistage PPS sampling and multivariate information, we consider sampling scheme $D_1$ as given below:

A sample $s(1)$ of $n$ first stage units (fsu) from a population $U$ of $N$ first stage units (fsus), the $j$th fsu containing $M_j$ second stage units, is selected using simple random sampling without replacement (SRSWOR). The characters $z, x_1, x_2, \ldots, x_p$ are observed over $s(1)$ yielding the observations $\{z_{jr}, x_{ijr}\}, j = 1, \ldots, n; r = 1, \ldots, M_j; i = 1, \ldots, p$. Now each of the $j$th fsu is stratified into $L_j$ strata, the number of the second stage units (ssu's) in $h$th stratum being, say $M_{jh}$, such that $\sum_{h=1}^{L_j} M_{jh} = M_j; j = 1, \ldots, n$.

Further, $m_{jh}$ ssu's are selected from $M_{jh}$ ssu's contained in $h$th stratum using PPSWR with probabilities of selection as
and principal characters $y$ and $x_o$ are observed yielding the observations \{ $y_{jhr}, x_{o jhr}, x_{i jhr}$ \}; $r = 1, \ldots, m_{jh}; \ h = 1, \ldots, L_j; \ j = 1, \ldots, n; \ i = 1, \ldots, p.$

Let $Y_{jh} = \sum_{r=1}^{m_{jh}} y_{jhr}, \ j = 1, \ldots, n; \ h = 1, \ldots, L_j$ be the total of $y$ in $h$th stratum of $j$th selected fsu. Let $\hat{Y}_{jh}$ be the estimator of $Y_{jh}$ based on ppz sample of size $m_{jh}$ from $h$th stratum of $j$th f.s.u. given as,

$$\hat{Y}_{jh} = \frac{m_{jh}}{m_{jh}} \sum_{r=1}^{m_{jh}} y_{jhr}; \ p_{jhr} = \frac{z_{jhr}}{\sum_{r=1}^{M_{jh}} z_{jhr}}$$

where $E(\hat{Y}_{jh} / h, s(1)) = Y_{jh} = \sum_{r=1}^{m_{jh}} y_{jhr}$. We also note that

$$\hat{Y}_j = \sum_{h=1}^{L_j} \hat{Y}_{jh}$$

is an unbiased estimator of $Y_j = \sum_{h=1}^{L_j} Y_{jh}$ with $E(\hat{Y}_j) = Y_j$.

Similarly, we have

$$\hat{x}_{i jh} = \frac{1}{m_{jh}} \sum_{r=1}^{m_{jh}} \frac{x_{i jhr}}{p_{jhr}}$$
where \( E(\hat{X}_{ijh}/s(1)) = X_{ijh} = \sum_{r=1}^{M_{jh}} x_{ijhr} \) and

\[
\hat{\beta}_{ij} = \sum_{h=1}^{L_j} \hat{X}_{ijh} = \sum_{h=1}^{L_j} \frac{1}{m_{jh}} \frac{1}{r=1} x_{ijhr}
\]

Also, let \( X_{ij} = \sum_{h=1}^{L_j} \frac{1}{r=1} x_{ijhr} \) be the total of the auxiliary character \( x_i; i = 1, \ldots, p \) for \( j \)th selected fsu.

An unbiased estimator of population total \( X_i \) based on \( s(1) \) is given by,

\[
\hat{X}_{i(1)} = \frac{N}{n} \sum_{j=1}^{n} X_{ij} = N \bar{X}_i
\] (6.3.1)

Based on the observations on second stage samples, we can also define the estimators of \( X_i, Y \) and \( X_o \) as,

\[
\hat{X}_{i(2)} = \frac{N}{n} \sum_{j=1}^{n} \sum_{h=1}^{L_j} \frac{1}{m_{jh}} \frac{1}{r=1} x_{ijhr}
\]

with \( E(\hat{X}_{i(2)}/j) = \hat{X}_{i(1)} = \frac{N}{n} \sum_{j=1}^{n} x_{ij} \)

\[
\hat{Y}_{(2)} = \frac{N}{n} \sum_{j=1}^{n} \sum_{h=1}^{L_j} \frac{1}{m_{jh}} \frac{1}{r=1} y_{ijhr}
\]

and \( \hat{X}_{o(2)} = \frac{N}{n} \sum_{j=1}^{n} \sum_{h=1}^{L_j} \frac{1}{m_{jh}} \frac{1}{r=1} x_{oijhr} \) (6.3.3)
We define below a generalized ratio estimator \( R^* \) for estimating population ratio \( R \) using two stage stratified PPS sampling and multivariate auxiliary information, as

\[
R^* = \frac{\sum_{i=1}^{p} w_i^{(1)} \alpha_i^{(1)}}{\sum_{i=1}^{p} w_i^{(2)} \alpha_i^{(2)}}
\]

(6.3.5)

where \( \alpha_i^{(1)} = \hat{Y}_i^{(2)} - \hat{Y}_i^{(1)}(\hat{X}_i^{(2)} - \hat{X}_i^{(1)}) \)

and \( \alpha_i^{(2)} = \hat{Y}_i^{(2)} - \hat{Y}_i^{(1)}(\hat{X}_i^{(2)} - \hat{X}_i^{(1)}) \)

The weight vectors \( w^{(1)}' = (w_1^{(1)}, \ldots, w_p^{(1)}) \) and \( w^{(2)}' = (w_1^{(2)}, \ldots, w_p^{(2)}) \) being such that \( w^{(1)}' e = 1 \) and \( w^{(2)}' e = 1 \) and \( \lambda^{(1)} \) and \( \lambda^{(2)} \) being the vectors of suitably chosen constants.

The bias and the mean square error of the estimator \( R^* \) may be obtained from (5.2.2a) and (5.2.2) by substituting the values of \( a_{ik}, b_{ik} \) and \( d_{ik} \) as following:

\[
a_{ik} = N^2 \left( \frac{1}{n} - \frac{1}{N} \right) S_{by}^2 + \sum_{j=1}^{N} \sum_{h=1}^{L} \frac{1}{m_{jh}} \sigma_{jh}^2 (Y_j - \lambda_i^{(1)} \sigma_{jh}(y, x_i) \sigma (x_j, x_i))
\]

\[
- \lambda_k^{(1)} \sigma_{jh}(y, x_k) + \lambda_i^{(1)} \lambda_k^{(1)} \sigma_{jh}(x_i, x_k)
\]

(6.3.6)
where the terms and notations used are defined in the lemma 6.3.1 and the results that follow it.

**Lemma 6.3.1.** For the above sampling scheme, given that

\[
\hat{X}_{i(2)} = \frac{N}{n} \sum_{j=1}^{n} \hat{X}_{ij}
\]

and

\[
\hat{X}_{ij} = \sum_{h=1}^{L_j} \hat{X}_{ijh} = \sum_{h=1}^{L_j} \sum_{m_{jh} = 1}^{M_{jh}} \sum_{r=1}^{p_{jhr}} x_{ijhr}
\]

we have

\[
C_2(\hat{X}_{i(2)}, \hat{X}_{k(2)}) = \frac{N^2}{n^2} \sum_{j=1}^{L_j} \sum_{m_{jh} = 1}^{M_{jh}} \sum_{r=1}^{p_{jhr}} (\hat{X}_{ijhr} - \hat{X}_{ij})(\hat{X}_{kjhr} - \hat{X}_{kj})
\]

... (6.3.9)

**Proof.** We know

\[
C_2(\hat{X}_{i(2)}, \hat{X}_{k(2)}) = \mathbb{E}_2(\hat{X}_{i(2)} - \hat{X}_{i(1)})(\hat{X}_{k(2)} - \hat{X}_{k(1)})
\]

\[
= \mathbb{E}_2 \left[ \frac{N}{n} \sum_{j=1}^{n} \frac{n}{n} \hat{X}_{ij} - \frac{N}{n} \sum_{j=1}^{n} \bar{X}_{ij} \right] \left[ \frac{N}{n} \sum_{j=1}^{n} \frac{n}{n} \hat{X}_{kj} - \frac{N}{n} \sum_{j=1}^{n} \bar{X}_{kj} \right]
\]
\[
\begin{align*}
E_2 \left( \frac{N}{n} \sum_{j=1}^{n} \left( \hat{X}_{ij} - X_{ij} \right) \right) & = \frac{N^2}{n^2} E_2 \left( \sum_{j=1}^{n} \left( \hat{X}_{ij} - X_{ij} \right) \hat{X}_{kj} - X_{kj} \right) + E_2 \left( \sum_{j \neq j} \hat{X}_{kj} - X_{kj} \right) \left( \hat{X}_{kj} - X_{kj} \right)
\end{align*}
\]

\[
= \frac{N^2}{n^2} \left[ E_2 \left( \sum_{j=1}^{n} \hat{X}_{ij} \hat{X}_{kj} + \sum_{j \neq j} \hat{X}_{kj} \hat{X}_{kj} \right) \right]
\]

\[
= \frac{N^2}{n^2} \sum_{j=1}^{L_j} E_2 \left( \sum_{h=1}^{L_j} \left( \hat{X}_{ij} \hat{X}_{jh} - X_{ij} X_{jh} \right) \right)
\]

\[
= \frac{N^2}{n^2} \sum_{j=1}^{L_j} E_2 \left( \sum_{h=1}^{L_j} \left( \hat{X}_{ij} \hat{X}_{jh} - X_{ij} X_{jh} \right) \right)
\]

or \( C_2(\hat{X}(2), \hat{X}(2)) = \frac{N^2}{n^2} \sum_{j=1}^{L_j} \sum_{h=1}^{M_j} \sigma_{jh}(x_i, x_k) \)

where \( \sigma_{jh}(x_i, x_k) = \sum_{r=1}^{M_j} \sigma_{jhr} \left( \frac{x_{ijr} - X_{ij}}{p_{jhr} - X_{kjh}} \right) \)

and, the following results follow immediately,

(1) \( V_1 E_2(\hat{Y}(2)) = N^2 M' N \sum_{j=1}^{M_j} \left( \frac{1}{M_j} \right) - \bar{Y} \)

where \( \bar{S}^2_{by} = \frac{1}{N-1} \sum_{j=1}^{N} \left( \frac{1}{M_j} \right) - \bar{Y} \)

with \( M' = \frac{1}{N} \sum_{j=1}^{N} M_j \) and \( \bar{Y} = \frac{Y}{NM'} \)

(2) \( V_1 E_2(\hat{X}(2)) = N^2 M' \sum_{j=1}^{N} \frac{1}{N} S^2_{by} \)
(3) $C_1(E_2^\hat{Y}(2), E_2^\hat{X}_o(2)) = N^2 \frac{M}{n^2} \frac{N-n}{n} S_{byx_o}$

where $S_{byx_o} = \frac{1}{N-1} \sum_{j=1}^{N} \left( \frac{M}{M_i} \bar{Y}_j - \bar{Y} \right) \left( \frac{M}{M_i} \bar{X}_{oj} - \bar{X}_o \right)$

(4) $E_1 V_2(\hat{Y}(2)) = \frac{N}{n} \sum_{j=1}^{L} \sum_{h=1}^{m} \frac{L_j \sigma^2_{j(h)(y)}}{m_jh}$

(5) $E_1 V_2(\hat{X}_o(2)) = \frac{N}{n} \sum_{j=1}^{L} \sum_{h=1}^{m} \frac{L_j \sigma^2_{j(h)(o)}}{m_jh}$

(6) $E_1 C_2(\hat{X}_i(2), \hat{X}_k(2)) = \frac{N}{n} \sum_{j=1}^{L} \sum_{h=1}^{m} \frac{L_j \sigma_{j(h)(x_i,x_k)}}{m_jh}$

(7) $E_1 C_2(\hat{Y}(2), \hat{X}_i(2)) = \frac{N}{n} \sum_{j=1}^{L} \sum_{h=1}^{m} \frac{1}{m_jh} \sigma_{j(h)(y,x_i)}$ (6.3.10)

(8) $E_1 C_2(\hat{X}_o(2), \hat{X}_i(2)) = \frac{N}{n} \sum_{j=1}^{L} \sum_{h=1}^{m} \frac{1}{m_jh} \sigma_{j(h)(x_o,x_i)}$

(9) $E_1 C_2(\hat{Y}(2), \hat{X}_o(2)) = \frac{N}{n} \sum_{j=1}^{L} \sum_{h=1}^{m} \frac{1}{m_jh} \sigma_{j(h)(y,x_o)}$

(10) $\sigma^2_{j(h)(y)} = \sum_{r=1}^{M} P_{jhr} \left( \frac{\bar{Y}_{jhr} - \bar{Y}_j}{p_{jhr}} \right)^2$

(11) $\sigma_{j(h)(x_i,x_k)} = \sum_{r=1}^{M} P_{jhr} \left( \frac{x_{i,jhr} - \bar{x}_i}{p_{jhr}} \right) \left( \frac{x_{k,jhr} - \bar{x}_k}{p_{jhr}} \right)$

(12) $S_{ik} = \frac{1}{N-1} \sum_{j=1}^{N} (x_{i,j} - \bar{x}_i)(x_{k,j} - \bar{x}_k)$. 
The unbiased estimators of $a_{ik}$, $b_{ik}$ and $d_{ik}$ may be obtained by substituting in (6.3.6), (6.3.7) and (6.3.8) the expression of unbiased estimators of the results (6.3.10) as given below:

(1) $V_1^2E_2(\hat{Y}_o(2)) = N^2M^2 \frac{1-f}{n} s^2_{by}$, where $s^2_{by} = \frac{1}{n-1} \sum_{j=1}^{M} \left( \frac{1}{M} \check{y}_j - \check{y} \right)^2$

(2) $V_1^2E_2(\hat{X}_o(2)) = N^2M^2 \frac{1-f}{n} s^2_{by}$, where $s^2_{by} = \frac{1}{n-1} \sum_{j=1}^{M} \left( \frac{1}{M} \check{x}_{o,j} - \check{x}_o \right)^2$

(3) $C_1(E_2(\hat{Y}_o(2)),E_2(\hat{X}_o(2))) = N^2M^2 \frac{1-f}{n} s^2_{by}$, where $s^2_{by} = \frac{1}{n-1} \sum_{j=1}^{M} \left( \frac{1}{M} \check{y}_j - \check{y} \right)(\frac{1}{M} \check{x}_{o,j} - \check{x}_o)$

(4) $E_1^2C_2(\hat{X}_o(2),\hat{X}_k(2)) = \frac{N^2}{n^2} \sum_{j=1}^{L} \sum_{h=1}^{L} \hat{\sigma}_{jh}(x_o,x_k)$

(5) $E_1^2C_2(\hat{Y}_o(2),\hat{X}_i(2)) = \frac{N^2}{n^2} \sum_{j=1}^{L} \sum_{h=1}^{L} \hat{\sigma}_{jh}(y,x_i)$

(6) $E_1^2C_2(\hat{X}_o(2),\hat{X}_i(2)) = \frac{N^2}{n^2} \sum_{j=1}^{L} \sum_{h=1}^{L} \hat{\sigma}_{jh}(x_o,x_i)$

(7) $E_1^2C_2(\hat{Y}_o(2),\hat{X}_o(2)) = \frac{N^2}{n^2} \sum_{j=1}^{L} \sum_{h=1}^{L} \hat{\sigma}_{jh}(y,x_o)$

(8) $E_1^2V_2(\hat{Y}(2)) = \frac{N^2}{n^2} \sum_{j=1}^{L} \sum_{h=1}^{L} \hat{\sigma}^2_{jh}(y)$
(9)  \[ E_1V_2(\hat{x}_o(2)) = \frac{N}{n^2} \sum_{j=1}^{L} \sum_{h=1}^{j} \delta_{jh}(x_o) \]

(10)  \[ \delta_{jh}(y) = \frac{1}{(m_{jh}-1)} \sum_{r=1}^{m_{jh}} \left( \frac{y_{jhr}}{p_{jhr}} - \hat{Y}_{jhr} \right)^2 \]

(11)  \[ \delta_{jh}(x_o) = \frac{1}{(m_{jh}-1)} \sum_{r=1}^{m_{jh}} \left( \frac{x_{o,jhr}}{p_{jhr}} - x_{o,jhr} \right)^2 \]

(6.3.11)

(12)  \[ \delta_{jh}(x_i, x_k) = \frac{1}{(m_{jh}-1)} \sum_{r=1}^{m_{jh}} \left( \frac{x_{i,jhr}}{p_{jhr}} - \hat{Y}_{i,jhr} \right) \left( \frac{x_{k,jhr}}{p_{jhr}} - \hat{X}_{k,jhr} \right) \]

(13)  \[ \delta_{jh}(y, x_i) = \frac{1}{(m_{jh}-1)} \sum_{r=1}^{m_{jh}} \left( \frac{y_{jhr}}{p_{jhr}} - \hat{Y}_{jhr} \right) \left( \frac{x_{o,jhr}}{p_{jhr}} - \hat{X}_{o,jhr} \right) \]

(14)  \[ \delta_{jh}(y, x_o) = \frac{1}{(m_{jh}-1)} \sum_{r=1}^{m_{jh}} \left( \frac{y_{jhr}}{p_{jhr}} - \hat{Y}_{jhr} \right) \left( \frac{x_{o,jhr}}{p_{jhr}} - \hat{X}_{o,jhr} \right) \]

(15)  \[ \delta_{jh}(x_o, x_i) = \frac{1}{(m_{jh}-1)} \sum_{r=1}^{m_{jh}} \left( \frac{x_{o,jhr}}{p_{jhr}} - \hat{X}_{o,jhr} \right) \left( \frac{x_{i,jhr}}{p_{jhr}} - \hat{X}_{i,jhr} \right) \]

Now, noting from (5.2.8) that

[\lambda_{o_i}(1) = \frac{E_1C_2(\hat{Y}(2), \hat{x}_i(2))}{E_1V_2(\hat{x}_i(2))}]

and using the results (6.3.10), we get

[\lambda_{o_i}(1) = \frac{N \sum_{j=1}^{L} \sum_{h=1}^{m_{jh}} \sigma_{jh}(y, x_i)}{N \sum_{j=1}^{L} \sum_{h=1}^{m_{jh}} \sigma_{jh}(x_i)^2}]

Further noting that \( \beta_{jh}(y, x_i) = \sigma_{j}(y, x_i) / \sigma_{j}(x_i) \) (6.3.12)

We have

\[
\lambda_{oi}^{(1)} = \sum_{j=1}^{N} \sum_{h=1}^{L_j} a_{jh} \beta_{jh}(y, x_i) 
\]

(6.3.13)

where

\[
\sum_{j=1}^{N} \sum_{h=1}^{L_j} a_{jh} = \sum_{j=1}^{N} \sum_{h=1}^{L_j} m_{jh} \sigma_{j}^2(x_i) 
\]

Similarly we obtain,

\[
\lambda_{oi}^{(2)} = \sum_{j=1}^{N} \sum_{h=1}^{L_j} a_{jh} \beta^{*}_{jh}(x_o, x_i) 
\]

(6.3.14)

where \( \beta^{*}_{jh}(x_o, x_i) = \sigma_{j}(x_o, x_i) / \sigma_{j}(x_i) \) (6.3.15)

In case the above choices of \( \lambda_{oi}^{(1)} \) and \( \lambda_{oi}^{(2)} \) are taken for each \( x_i, i = 1, \ldots, p \), using the results (6.3.10), (6.3.13) and (6.3.14), we get the expression for mean square error of the estimator \( \hat{R}^* \) as (5.2.2), with,
\[ a_{ik} = N^2 \left( \frac{1}{n} - \frac{1}{N} S^2 \right) b_{y} + \frac{N^2}{n} \sum_j \sum_{h=1}^{L_j} \left[ \sigma^2 \right] \frac{1}{m_j h} \rho_{j h}(y) - \frac{N}{L_j} \sum_{j=1}^{L_j} \sum_{h=1}^{m_j} a_{j h} \sigma_{j h}(y, x_i) \]

\[ \frac{N}{L_j} \sum_{j=1}^{L_j} \sum_{h=1}^{m_j} a_{j h} \sigma_{j h}(y, x_k) \]

\[ b_{ik} = N^2 \left( \frac{1}{n} - \frac{1}{N} S^2 \right) b_{x} + \frac{N^2}{n} \sum_j \sum_{h=1}^{L_j} \left[ \sigma^2 \right] \frac{1}{m_j h} \rho_{j h}(x_o) - \frac{N}{L_j} \sum_{j=1}^{L_j} \sum_{h=1}^{m_j} a_{j h} \sigma_{j h}(x_o, x_i) \]

\[ \frac{N}{L_j} \sum_{j=1}^{L_j} \sum_{h=1}^{m_j} a_{j h} \sigma_{j h}(x_o, x_k) \]

\[ c_{ik} = N^2 \left( \frac{1}{n} - \frac{1}{N} S^2 \right) b_{x} + \frac{N^2}{n} \sum_j \sum_{h=1}^{L_j} \left[ \sigma^2 \right] \frac{1}{m_j h} \rho_{j h}(x_i) - \frac{N}{L_j} \sum_{j=1}^{L_j} \sum_{h=1}^{m_j} a_{j h} \sigma_{j h}(x_i, x_k) \]

\[ d_{ik} = N^2 \left( \frac{1}{n} - \frac{1}{N} S^2 \right) b_{x} + \frac{N^2}{n} \sum_j \sum_{h=1}^{L_j} \left[ \sigma^2 \right] \frac{1}{m_j h} \rho_{j h}(x_o) - \frac{N}{L_j} \sum_{j=1}^{L_j} \sum_{h=1}^{m_j} a_{j h} \sigma_{j h}(y, x_o) \]

... (6.3.16)

(6.3.17)
Using the above values of $a_{ik}$, $b_{ik}$ and $d_{ik}$, the expression of the mean square error of the generalized estimator $\hat{R}$ is given as

$$M(\hat{R}) = M(R) - \frac{1}{n} \sum_{i=1}^{N} \sum_{h=1}^{L_j} \frac{1}{m_{jh}} \sigma_j h(y, x_i) +$$

$$-2R \sum_{i=1}^{p} \sum_{k=1}^{w_{ik}(1)} \beta_{i}(y, x_i) \beta_{k}(y, x_k) \sum_{j=1}^{N} \sum_{h=1}^{L_j} \frac{1}{m_{jh}} \sigma_j h(x_o, x_i) -$$

$$-2R \sum_{i=1}^{p} \sum_{k=1}^{w_{ik}(2)} \beta_{i}(y, x_i) \beta_{k}(y, x_k) \sum_{j=1}^{N} \sum_{h=1}^{L_j} \frac{1}{m_{jh}} \sigma_j h(x_i, x_k)$$

$$(6.3.19)$$
where \( \beta(\ ) \) are defined in (6.3.12) and (6.3.15) and \( M(R) \) is given as,

\[
M(R) = N^2 \left( \frac{1}{n} - \frac{1}{N} \right) [S_{by}^2 - 2RS_{by}x_o + R^2S_{x_o}^2] + \frac{N}{n} \sum_{j=1}^{L} \sum_{h=1}^{m_{jh}} \frac{1}{\sigma_{jh}(y)} \cdot \sigma_{jh}(x_o) \\
2R \sigma_{jh}(y,x_o) + R^2 \sigma_{jh}(x_o) \]

(6.3.20)

where \( \hat{R} = \hat{\gamma}(2)/\hat{X}_{0}(2) \)

Further taking \( \lambda^{(1)}_{oi} \) and \( \lambda^{(2)}_{oi} \) in terms of correlation coefficients, we get from (6.3.12), (6.3.13), (6.3.14) and (6.3.15),

\[
\lambda^{(1)}_{oi} = \rho(y,x_i) \frac{\sigma_{*}(y)}{\sigma_{*}(x_i)}
\]

(6.3.21)

and

\[
\lambda^{(2)}_{oi} = \rho(x_o,x_i) \frac{\sigma_{*}(x_o)}{\sigma_{*}(x_i)}
\]

(6.3.22)
Using (6.3.21) to (6.3.23), we get the mean square error of \( \mathcal{R} \) from (5.2.2), with,

\[
a_{ik} = N^2 \left( \frac{1}{n} - \frac{1}{N} \right) S_{by}^2 + \frac{N}{n} \left[ \sigma^2(y) - \rho^2(y, x_i) \sigma^2(y) - \rho^2(y, x_k) \sigma^2(y) \right]
\]

\[
+ \rho(y, x_i) \rho(y, x_k) \rho(x_i, x_k) \sigma^2(y) \]

\[
= N^2 \left( \frac{1}{n} - \frac{1}{N} \right) S_{by}^2 + \frac{N}{n} \sigma^2(y) \left[ 1 - \rho^2(y, x_i) - \rho^2(y, x_k) + \rho(y, x_i) \rho(y, x_k) \rho(x_i, x_k) \right]
\]...

\[
b_{ik} = N^2 \left( \frac{1}{n} - \frac{1}{N} \right) S_{bx_o}^2 + \frac{N}{n} \sigma^2(x_o) \left[ 1 - \rho^2(x_o, x_i) - \rho^2(x_o, x_k) + \rho(x_o, x_i) \rho(x_o, x_k) \rho(x_i, x_k) \right]
\]...

\[
and \quad d_{ik} = N^2 \left( \frac{1}{n} - \frac{1}{N} \right) S_{byx_o} + \frac{N}{n} \sigma^2(x_o) \left[ \rho(y, x_o) - \rho(y, x_i) \rho(x_o, x_i) \right]
\]

\[
- \rho(y, x_k) \rho(x_o, x_k)^* + \rho(y, x_i) \rho(x_o, x_k) \rho(x_i, x_k)^*
\]
6.4. Proposed Sampling Scheme D.<br><br>To estimate the population ratio $R$, when SRSWOR is used to select first stage and second stage units, we consider sampling scheme $D_2$ as given below:<br><br>At first stage a sample $s(l)$ of $n$ first stage units (fsus) is drawn using SRSWOR from a population $U$ of $N$ fsu's, the $j$th fsu containing $M_j$ second stage units. The characters $x_{i}(i=1,\ldots,p)$ are observed yielding the observations ${x_{ijr}}$, $j = 1,\ldots,n$; $r = 1,\ldots,M_j$; $i = 1,\ldots,p$. Now each of the $j$th selected fsu is stratified into $L_j$ strata, the number of the second stage units (ssu's) in $h$th stratum being, say $M_{jh}$ such that<br><br>$$
exists_h M_{jh} = M_j; \quad j = 1,2,\ldots,n.$$<br><br>Further, $m_{jh}$ ssu's are selected from $M_{jh}$ ssu's contained in $h$th stratum of $j$th selected fsu for every $j = 1,2,\ldots,n$, using SRSWOR and the principal characters $y$ and $x_0$ are observed yielding the observations<br><br>$$\{y_{jhr}, x_{0jhr}, x_{ijhr}\}; \quad r = 1,2,\ldots,m_{jh}; \quad h = 1,\ldots,L_j; \quad j = 1,\ldots,n; \quad i = 1,\ldots,p.$$<br><br>Let $Y_{jh} = \sum_{r=1}^{m_{jh}} y_{jhr}$, $j = 1,\ldots,n$; $h = 1,\ldots,L_j$ be the
total of \( y \) in \( h \)th stratum of \( j \)th selected fsu.

Let \( \hat{Y}_{jh} \) be the estimator of \( Y_{jh} \) based on SRSWOR sample of size \( m_{jh} \) from \( h \)th stratum of \( j \)th selected fsu given as,

\[
\hat{Y}_{jh} = \frac{M_{jh}}{m_{jh}} \sum_{r=1}^{m_{jh}} y_{jhr}
\]

where \( E(\hat{Y}_{jh}/h,s(1)) = Y_{jh} \).

We also note that \( \hat{Y}_j = \sum_{h=1}^{L_j} \hat{Y}_{jh} \) is an unbiased estimator of \( Y_j = \sum_{h=1}^{L_j} Y_{jh} \) with \( E(\hat{Y}_j) = Y \).

Similarly, we have

\[
\hat{X}_{ijh} = \frac{M_{ijh}}{m_{ijh}} \sum_{r=1}^{m_{ijh}} x_{ijhr}
\]

where \( E(\hat{X}_{ijh}/h,s(1)) = X_{ijh} \)

and

\[
\hat{X}_{ij} = \sum_{h=1}^{L_j} \hat{X}_{ijh} = \sum_{h=1}^{L_j} \frac{M_{ijh}}{m_{ijh}} \sum_{r=1}^{m_{ijh}} x_{ijhr}
\]

Also, let \( X_{ij} = \sum_{h=1}^{L_j} \sum_{r=1}^{M_{ijh}} x_{ijhr} \) be the total of the auxiliary character \( x_i \), \( i = 1,2,\ldots,p \) for \( j \)th selected fsu.

An unbiased estimator of population total \( X_i \) based on \( s(1) \) is given as
Based on the observations on second stage samples, we can also define the estimators of \( X_i, Y \) and \( X_0 \) as

\[
\hat{X}_i(1) = \frac{N}{n} \sum_{j=1}^{n} X_{ij} = N \bar{X}_i
\]

\[
\hat{X}_i(2) = \frac{N}{n} \sum_{j=1}^{n} \sum_{h=1}^{L_j} \sum_{r=1}^{M_{jhr}} \frac{M_{jhr}}{M} \sum_{m_{jhr}} x_{ijhr}
\]

with \( \text{E}(\hat{X}_i(2)/j \cdot j) = \hat{X}_i(1) = \frac{N}{n} \sum_{j=1}^{n} X_{ij} \)

\[
\hat{Y}(2) = \frac{N}{n} \sum_{j=1}^{n} \sum_{h=1}^{L_j} \sum_{r=1}^{M_{jhr}} \frac{M_{jhr}}{M} \sum_{m_{jhr}} y_{jhr} = \frac{N}{n} \sum_{j=1}^{n} \sum_{j=1}^{n} \hat{Y}_j
\]

and \( \hat{X}_o(2) = \frac{N}{n} \sum_{j=1}^{n} \sum_{h=1}^{L_j} \sum_{r=1}^{M_{jhr}} \frac{M_{jhr}}{M} \sum_{m_{jhr}} x_{ojhr} = \frac{N}{n} \sum_{j=1}^{n} \sum_{j=1}^{n} \hat{X}_o j
\)

Now under sampling scheme \( D_2 \), the bias and mean square error of the estimator may be obtained from (5.2.2a) and (5.2.3), (5.2.4) and (5.2.5) by substituting the values of different components of the expressions (5.2.3), (5.2.4) and (5.2.5) from the following lemma.

**Lemma 6.4.1.**

1. \( \text{Var}_1(E_2(\hat{Y}(2))) = N^2 M'^2 (1-f) \frac{S_{by}^2}{n} ; \quad S_{by}^2 = \frac{1}{N-1} \sum_{j=1}^{n} \frac{M_j}{M} (\bar{Y}_j - \bar{Y})^2 \)

with \( M' = \frac{1}{N} \sum_{j=1}^{N} M_j \) and \( \bar{Y} = \bar{Y}/NM \); \( f = n/N \)
(2) \[ V_1 E_2(\hat{\beta}_o(2)) = N^2 M^2 (1-f) \frac{S_{bxy}}{n} ; S_{bxy}^2 = \frac{1}{N-1} \sum_{j=1}^{N} \frac{M_i \hat{X}_{ij} - \overline{X}_o}{M'} \]

with \[ M' = \frac{1}{N} \sum_{j=1}^{N} M_j \] and \[ \overline{X}_o = \frac{\hat{X}_o}{NM'} \]

(3) \[ E_1 V_2(\hat{Y}_o(2)) = \frac{N}{n} \sum_{j=1}^{N} L_j \sum_{h=1}^{E} M_{jh}^2 (1-f) \frac{S_{ih}(y)}{m_jh} \]

(4) \[ E_1 V_2(\hat{X}_o(2)) = \frac{N}{n} \sum_{j=1}^{N} L_j \sum_{h=1}^{E} M_{jh}^2 (1-f) \frac{S_{ih}(x_0)}{m_jh} \]

(5) \[ E_1 C_2(\hat{Y}_o(2), \hat{X}_i(2)) = \frac{N}{n} \sum_{j=1}^{N} L_j \sum_{h=1}^{E} M_{jh}^2 (1-f) \frac{S_{ih}(y,x_i)}{m_jh} \]

(6) \[ E_1 C_2(\hat{X}_o(2), \hat{X}_i(2)) = \frac{N}{n} \sum_{j=1}^{N} L_j \sum_{h=1}^{E} M_{jh}^2 (1-f) \frac{S_{ih}(x_0,x_i)}{m_jh} \]

(7) \[ E_1 C_2(\hat{Y}_o(2), \hat{X}_o(2)) = \frac{N}{n} \sum_{j=1}^{N} L_j \sum_{h=1}^{E} M_{jh}^2 (1-f) \frac{S_{ih}(y,x_o)}{m_jh} \]

(8) \[ E_1 C_2(\hat{X}_i(2), \hat{X}_k(2)) = \frac{N}{n} \sum_{j=1}^{N} L_j \sum_{h=1}^{E} M_{jh}^2 (1-f) \frac{S_{ih}(x_i,x_k)}{m_jh} \]

(9) \[ C_1(E_2 \hat{Y}_o(2), E_2 \hat{X}_o(2)) = N^2 (1-f) \frac{S_{bxy_0}}{n} \]

with \[ f = \frac{n}{N} \], \[ S_{bxy_0} = \frac{1}{N-1} \sum_{j=1}^{N} \frac{M_i \hat{Y}_{ij} - \overline{Y}}{M'} \left( \frac{M_i \hat{X}_{ij} - \overline{X}_o}{M'} \right) \]

From the above lemma and (5.2.3), (5.2.4) and (5.2.5), we get
\[
\begin{align*}
    a_{ik} &= \frac{N^2(1-f)}{n} S_{by}^2 + \frac{N}{n} \sum_{j=1}^{L_j} \sum_{h=1}^{M_{jh}} (1-f_jh) \frac{S_{jh(y)}}{m_{jh}} \\
    &= \lambda_1^{(1)} \frac{N}{n} \sum_{j=1}^{L_j} \sum_{h=1}^{M_{jh}} (1-f_jh) \frac{S_{jh(y,x_i)}}{m_{jh}} \\
    &= \lambda_2^{(1)} \frac{N}{n} \sum_{j=1}^{L_j} \sum_{h=1}^{M_{jh}} (1-f_jh) \frac{S_{jh(y,x_k)}}{m_{jh}} \\
    &= \lambda_1^{(1)} \lambda_2^{(1)} \frac{N}{n} \sum_{j=1}^{L_j} \sum_{h=1}^{M_{jh}} (1-f_jh) \frac{S_{jh(x_1,x_1)}}{m_{jh}} \\
    &= \lambda_1^{(2)} \lambda_2^{(2)} \frac{N}{n} \sum_{j=1}^{L_j} \sum_{h=1}^{M_{jh}} (1-f_jh) \frac{S_{jh(x_1,x_k)}}{m_{jh}} \\
    \end{align*}
\]

(6.4.5)

\[
\begin{align*}
    b_{ik} &= \frac{N^2(1-f)}{n} S_{bx_0}^2 + \frac{N}{n} \sum_{j=1}^{L_j} \sum_{h=1}^{M_{jh}} (1-f_jh) \frac{S_{jh(x_0)}}{m_{jh}} \\
    &= \lambda_1^{(2)} \frac{N}{n} \sum_{j=1}^{L_j} \sum_{h=1}^{M_{jh}} (1-f_jh) \frac{S_{jh(x_0,x_1)}}{m_{jh}} \\
    &= \lambda_2^{(2)} \frac{N}{n} \sum_{j=1}^{L_j} \sum_{h=1}^{M_{jh}} (1-f_jh) \frac{S_{jh(x_0,x_k)}}{m_{jh}} \\
    &= \lambda_1^{(2)} \lambda_2^{(2)} \frac{N}{n} \sum_{j=1}^{L_j} \sum_{h=1}^{M_{jh}} (1-f_jh) \frac{S_{jh(x_1,x_k)}}{m_{jh}} \\
    \end{align*}
\]

(6.4.6)

and

\[
\begin{align*}
    d_{ik} &= \frac{N^2(1-f)}{n} S_{byx_0}^2 + \frac{N}{n} \sum_{j=1}^{L_j} \sum_{h=1}^{M_{jh}} (1-f_jh) \frac{S_{jh(y,x_0)}}{m_{jh}} \\
    &= \lambda_1^{(1)} \frac{N}{n} \sum_{j=1}^{L_j} \sum_{h=1}^{M_{jh}} (1-f_jh) \frac{S_{jh(y,x_1)}}{m_{jh}} \\
    \end{align*}
\]
The unbiased estimators of $a_{ik}$, $b_{ik}$ and $d_{ik}$ may be obtained by substituting the unbiased estimators of different components from the following lemma.

**Lemma 6.4.2.**

1. \[ \hat{\psi}_1 E_2(\hat{Y}_2) = \frac{N^2(1-f)}{n} M' s_{by}^2 \]
   where \[ s_{by}^2 = \frac{1}{n-1} \sum_{j=1}^{M} \bar{y}_j - \bar{y} \]^2

2. \[ \hat{\psi}_1 E_2(\hat{X}_o(2)) = N^2 M' s_{bxo}^2 \]
   where \[ s_{bxo}^2 = \frac{1}{n-1} \sum_{j=1}^{M} \bar{x}_{oj} - \bar{x} \]^2

3. \[ \hat{C}_1 (E_2 \hat{Y}_2(2), E_2 \hat{X}_o(2)) = N^2 M' \frac{1-f}{n} s_{byx_o} \]
   where \[ s_{byx_o} = \frac{1}{n-1} \sum_{j=1}^{M} \bar{x}_{oj} - \bar{x} \] \[ \frac{1}{M} \bar{y}_j - \bar{y} \] \[ = \frac{1}{M} \bar{y}_j - \bar{y} \]

4. \[ \hat{E}_1 V_2(\hat{Y}_2) = \frac{N^2}{n} \sum_{j=1}^{L_j} \sum_{h=1}^{M_j} (1-f_{jh}) \frac{s_{jh}(y)}{m_{jh}} \]
   where \[ s_{jh}(y) = \frac{1}{m_{jh}-1} \sum_{r=1}^{m_{jh}} \left( y_{jhr} - \bar{y}_{jh} \right)^2 \]
(5) \[ E_{12}^2(\hat{X}_o(2)) = \frac{N^2}{n^2} \sum_{j=1}^{n} \sum_{h=1}^{L} \frac{1}{M_{jh}} \left( 1 - f_{jh} \right) \frac{s_{jh}(x_o)}{m_{jh}} \]

where \[ s_{jh}(x_o) = \frac{1}{m_{jh}-1} \sum_{r=1}^{m_{jh}} (x_{o,jhr} - \bar{x}_{o,jh})^2 \]

(6) \[ E_{12}^2(\hat{X}_o(2), \hat{Y}_i(2)) = \frac{N^2}{n^2} \sum_{j=1}^{n} \sum_{h=1}^{L} \frac{1}{M_{jh}} \left( 1 - f_{jh} \right) \frac{s_{jh}(x_o, y)}{m_{jh}} \]

where \[ s_{jh}(x_o, y) = \frac{1}{m_{jh}-1} \sum_{r=1}^{m_{jh}} (x_{o,jhr} - \bar{x}_{o,jh})(y_{jhr} - \bar{y}_{jh}) \]

(7) \[ E_{12}^2(\hat{Y}_i(2), \hat{X}_i(2)) = \frac{N^2}{n^2} \sum_{j=1}^{n} \sum_{h=1}^{L} \frac{1}{M_{jh}} \left( 1 - f_{jh} \right) \frac{s_{jh}(y, x_i)}{m_{jh}} \]

where \[ s_{jh}(y, x_i) = \frac{1}{m_{jh}-1} \sum_{r=1}^{m_{jh}} (y_{jhr} - \bar{y}_{jh})(x_{i,jhr} - \bar{x}_{i,jh}) \]

(8) \[ E_{12}^2(\hat{X}_o(2), \hat{X}_i(2)) = \frac{N^2}{n^2} \sum_{j=1}^{n} \sum_{h=1}^{L} \frac{1}{M_{jh}} \left( 1 - f_{jh} \right) \frac{s_{jh}(x_o, x_i)}{m_{jh}} \]

where \[ s_{jh}(x_o, x_i) = \frac{1}{m_{jh}-1} \sum_{r=1}^{m_{jh}} (x_{o,jhr} - \bar{x}_{o,jh})(x_{i,jhr} - \bar{x}_{i,jh}) \]

(9) \[ E_{12}^2(\hat{X}_i(2), \hat{X}_k(2)) = \frac{N^2}{n^2} \sum_{j=1}^{n} \sum_{h=1}^{L} \frac{1}{M_{jh}} \left( 1 - f_{jh} \right) \frac{s_{jh}(x_i, x_k)}{m_{jh}} \]

where \[ s_{jh}(x_i, x_k) = \frac{1}{m_{jh}-1} \sum_{r=1}^{m_{jh}} (x_{i,jhr} - \bar{x}_{i,jh})(x_{k,jhr} - \bar{x}_{k,jh}) \]
6.5. **Optimal Sample Sizes**

In this section we obtain the values of \( n \) and \( m_j \) for sampling scheme \( D_1 \) so that the estimated population ratio has maximum precision for given cost has desired precision for minimum cost.

For the given sampling strategy one may adopt the following form of cost function,

\[
C = nC_1 + \Sigma \left( \Sigma m_{jh} C_{jh} \right)_{j=1}^{L_j}
\]

(6.5.1)

where the second term contributes to the cost coming from second stage units, which obviously varies from sample to sample of \( n \) fsu's. We take therefore the average cost of the survey to be as

\[
C = nC_1 + \frac{n}{2} \frac{N}{N} \sum_{j=1}^{L_j} m_j
\]

(6.5.2)

Now the mean square error (6.3.19) of the ratio estimator \( R^* \) can be written for the sake of convenience in the form given below provided we follow proportional allocation, i.e. \( m_j \propto M_j \)

\[
M(R^*) = \frac{1}{n} \left\{ A_1 + N \sum_{j=1}^{L_j} M_j A_{2j} \right\} - A_3
\]

(6.5.3)
where \( A_1, A_{2j} \) (for given \( j \)) and \( A_3 \) are constants independent of \( n \) and \( m_j \).

To obtain optimum \( m_j \), we get from (6.5.2) and (6.4.3),

\[
[M(R^*) + A_3]C = C_1 \left( A_1 N \sum_{j=1}^{M_j} A_{2j} \right) + \\
\frac{C_2}{N} \left( \sum_{j=1}^{M_j} A_{2j} \right) \sum_{j=1}^{M_j} \left( A_1 + N \sum_{j=1}^{M_j} A_{2j} \right)
\]

which on some algebraic simplifications gives

\[
[M(R^*) + A_3]C = \frac{1}{N} \sum_{j=1}^{N} \left( \sqrt{C_2 A_1 m_j} - N \sqrt{\frac{C_1 M_j A_{2j}}{m_j}} \right)^2 \\
+ C_2 \sum_{j=1}^{M_j} \left( \sqrt{\frac{m_j}{M_j A_{2j}}} - \sqrt{\frac{m_j}{M_j A_{2j}}} \right)^2
\]

+ terms free of \( m_j \). \quad (6.5.5)

The above is minimized if both the square terms are zero, thus yielding

\[
\sqrt{C_2 A_1 m_j} = N \sqrt{\frac{C_1 M_j A_{2j}}{m_j}} \quad (6.5.6)
\]

and

\[
\sqrt{\frac{m_j}{M_j A_{2j}}} = \sqrt{\frac{m_j}{M_j A_{2j}}} \quad (6.5.7)
\]

From (6.4.6), the optimum value of \( m_j \) is given by
From (6.4.2) and (6.4.8) we get the optimum \( n \) as

\[
\hat{n} = N \sqrt{\frac{C_1 M_j A_{2j}}{C_2 \Sigma_1}}
\]  

(6.5.8)

and the resultant minimum mean square error of the ratio estimator \( \hat{R} \) from (6.5.3) and (6.4.9) is obtained as,

\[
M_o(\hat{R}^*) = \frac{1}{C} \left[ \sqrt{C_1 A_1} + \sqrt{C_2} \Sigma_{j=1}^{N} \sqrt{M_j A_{2j}} \right] \tag{6.5.10}
\]

Following on the same lines the optimal sample sizes can be obtained easily for sampling scheme \( D_2 \) as well.