3.1. Introduction.

In sample surveys one is frequently interested in estimating population mean, total, proportion, population ratio or the change in their values over different occasions (years or seasons). One may, for example, estimate the total production of milk in a milk survey or the yield rate of a principal crop over different seasons in a crop survey.

The practice of using partial replacement sampling scheme in repeated surveys is quite common. After each sampling occasion a fraction of the units observed on that occasion is replaced by a fresh subsample from the population. This set of replaced (unmatched) units is then observed on the next occasion along with the remaining set of retained (matched) units.

The estimation of population mean $\bar{Y}_2$ of $y$ on second occasion based on the scheme of partial replacement and simple random sampling by using the information on auxiliary characters say $x_1, x_2, \ldots, x_p$ has been considered, among others, by

Raj (1965b) and Tripathi (1970, 1976) considered estimation of population mean on second occasion with probability proportional to size with replacement (PPSWR) sampling and partial replacement of units using a difference type estimator, whereas Tripathi and Srivastava (1979) considered a sampling strategy based on PPSWR sampling with partial replacement scheme and a univariate regression type estimator.

We consider a situation where the information on several auxiliary characters may be readily available at hand or may be made to be available by diverting a part of the survey budget. This information is then used in different ways to improve the precision of estimation. For example in apple production surveys the information about the area of orchards, number of trees, amount of different fertilizers and chemicals used etc. may be used by selecting the orchards with probability proportional to their areas and using the information in constructing the multivariate difference type, ratio type and regression type estimators. Such a scheme was considered by Tripathi et al. (1989), wherein they assumed...
the knowledge of the selection variable $z$ (area of orchard, or number of trees) for estimating the population mean $\bar{V}_2$ of the character $y$ on second occasion. In practice, however, the information on selection variable $z$ may not always be available apriori.

We, therefore, consider modified selection and estimation procedures in which no such assumption is made, and the sample on first occasion is devoted to collect information on $z$ as well, in addition to the information on other related auxiliary characters $x_1, x_2, \ldots, x_p$.

Let $U = \{1, 2, \ldots, N\}$ be a finite population of $N$ units and $\{y_{2j}\}$ be the values of the principal characteristic $y$ on second occasion and $x_{ij}$ and $z_j$ be the values of the auxiliary characters $x_i (i=1, \ldots, p)$ and $z$ for the $j$th unit $(j=1, \ldots, N)$ of the population on first occasion.

Let $\bar{y}_2 = \frac{1}{N} \sum_{j=1}^{N} y_{2j}$, $\bar{x}_i = \frac{1}{N} \sum_{j=1}^{N} x_{ij}$ and $\bar{z} = \frac{1}{N} \sum_{j=1}^{N} z_j$ denote the population means and $p_j = z_j / (NZ)(j=1, \ldots, N)$.

The problem considered is to estimate the population mean $\bar{V}_2$ of $y$ on second occasion using double sampling for PPS estimation. The information on auxiliary characters
(x_1, x_2, ..., x_p) is collected on the first occasion. A sampling strategy based on partial matching is proposed, and comparison with some other strategies is made.

3.2. A General Approach

In this section we present some general results which may be applied in a large number of situations.

On first occasion let a sample s(1) of size n be selected from the population of size N, using any suitable sampling procedure. On the second occasion let m = n\lambda(0 < \lambda < 1) units, say s(2m) be selected from s(1) using some suitable sampling procedure, and a fresh sample of u units, say s(2u) be drawn independently from the population according to a suitable sampling scheme.

Let \( \hat{Y}_m \) and \( \hat{Y}_u \) be two independent estimators for the population mean \( \bar{Y}_2 \) on the second occasion, based on the information on s(1), s(2m) and s(2u), with their variances expressible in the form

\[
V(\hat{Y}_m) = \frac{1}{n} Q_1 + \frac{1}{m} Q_2 + Q_3
\]

\[
V(\hat{Y}_u) = \frac{1}{u} Q_4 + Q_5
\]  

(3.2.1)
where \( Q_i \) (\( i=1,2,\ldots,5 \)) are the quantities not depending on \( n, m \) and \( u \).

Let \( \hat{Y}_2 = \alpha \hat{Y}_m + (1-\alpha) \hat{Y}_u \) \hspace{1cm} (3.2.2)

be a combined estimator of \( Y_2 \), where optimum value of the weight \( \alpha \) is given by,

\[
\alpha_o = \frac{Q_4/u + Q_5}{(Q_1/n + Q_2/m + Q_3 + Q_4/u + Q_5)} \hspace{1cm} (3.2.3)
\]

Let \( n' = m+u = nA+u \) be the size of the sample on the second occasion. The variance of the optimally combined estimator \( \hat{Y}_2 \) is found to be

\[
V(\hat{Y}_2) = \frac{\{Q_4 + (n'-n\lambda)Q_5\}{Q_2+\lambda(Q_1+nQ_3)}}{(n'-n\lambda){Q_2+\lambda(Q_1+nQ_3+nQ_5)}}+n\lambda Q_4 \hspace{1cm} (3.2.4)
\]

In case \( n = n' \), that is the sample sizes are same on both the occasions, we may write

\[
V(\hat{Y}_2) = \frac{1}{n} \frac{\lambda^2 A_1 + \lambda A_2 + A_3}{\lambda^2 B_1 + \lambda B_2 + B_3}
\]

where

\[
A_1 = -nQ_3(Q_1+nQ_3)
\]

\[
A_2 = (Q_1+nQ_3)(Q_4+nQ_5) - nQ_2Q_5
\]

\[
A_3 = Q_2(Q_4+nQ_5)
\]
\[ B_1 = -Q_1 + n(Q_3 + Q_5) \]

\[ B_2 = (Q_1 - Q_2 + nQ_3 + Q_4 + nQ_5) \]

\[ B_3 = Q_2 \]

The value of \( \lambda \) which minimizes \( V(\hat{Y}_2) \) is given by

\[
\lambda_o = \frac{C_2 \pm \sqrt{C_2^2 - C_1 C_3}}{C_1} \quad (3.2.6)
\]

where

\[ C_1 = A_1 B_2 - A_2 B_1 \]

\[ C_2 = A_3 B_1 - A_1 B_3 \]

\[ C_3 = A_2 B_3 - A_3 B_2. \]

In situations where both \( Q_3 \) and \( Q_5 \) are negligible, we obtain (from 3.2.4)

\[
V(\hat{Y}_2) = \frac{Q_4(\lambda Q_1 + Q_2)}{(n^t - n \lambda)[\lambda Q_1 + Q_2] + n \lambda Q_4} \quad (3.2.7)
\]

The value of \( \lambda \) minimizing \( V(\hat{Y}_2) \) in (3.2.7) and the resulting minimum variance are given by,

\[
\lambda_o = \frac{\sqrt{Q_2 Q_4}}{Q_1} [1 - \sqrt{\beta}], \quad (3.2.8)
\]

and
\[ V_0(\hat{\theta}_2) = \frac{Q_1Q_4}{n'Q_1 + nQ_4[1-\sqrt{\theta}]}^2 \] (3.2.9)

where \( \theta = \frac{Q_2}{Q_4} > 0 \).

3.3. Proposed Sampling Strategy and its Properties

In some situations information on the selection variable \( z \) may not be available in advance, and consequently the sample on the first occasion may not be selected according to PPZWR with probabilities \( P_j = \frac{z_j}{\sum_{j=1}^{N} z_j} \), \( (j = 1, \ldots, N) \). We propose below a sampling strategy for estimating \( \bar{Y}_2 \), (the population mean of character \( y \) on second occasion), using multivariate auxiliary information and PPS estimation in such situations.

On the first occasion a sample \( s(1) \) of size \( n \) is drawn from the population \( U \) of \( N \) units with SRSWOR and variates \( (z, x_1, \ldots, x_p) \) are observed, the observations being \( \{z_j, x_{1j}, x_{2j}, \ldots, x_{pj}\}, j = 1, \ldots, n \). Based on the information \( \{z_1, \ldots, z_n\} \) we define the selection probabilities \( p_j = \frac{z_j}{\sum_{j=1}^{n} z_j} \). On the second occasion, a sample \( s(2m) \) of size \( m = n\lambda (0 < \lambda < 1) \) is selected from \( s(1) \) with PPZWR, the probability of selecting jth unit being \( p_j = \frac{z_j}{\sum_{j=1}^{n} z_j} \) at each draw, and a sample \( s(2u) \) of size \( u \) is selected independently from the population using SRSWOR and character
y is observed, yielding the observations \{y_{2j}\}, (j=1,\ldots,m) and \{y_{2j}\}, (j=1,\ldots,u) for matched part s(2m) and unmatched part s(2u) respectively of the sample s(2) = (s(2m),s(2u)) on the second occasion. It may be remarked that one of the variates \(z,x_1,\ldots,x_p\) may be character \(y\) itself observed on the first occasion.

Based on the information on \(s(1)\), we define an unbiased estimate for \(\bar{x}_i\) (\(i=1,\ldots,p\)) as

\[
\bar{x}_i(1) = \frac{1}{n} \sum_{j=1}^{n} x_{ij} = \bar{x}_i
\]

Further on the basis of observations \(\{x_{ij}\}, (j=1,\ldots,m)\) for those units which are common to \(s(2m)\) and \(s(1)\), an unbiased estimator of \(\bar{x}_i\) is

\[
\bar{x}_i(2) = \frac{1}{nm} \sum_{j=1}^{m} x_{ij} / p_j, \quad p_j = z_j / \sum_{j=1}^{n} z_j.
\]

An unbiased estimator for \(\bar{y}_2\) based on observations \(\{y_{2j}\}, (j=1,\ldots,m)\) for units in \(s(2m)\) is defined as,

\[
\bar{y}_2(2) = \frac{1}{nm} \sum_{j=1}^{m} y_{2j} / p_j
\]

The following notations will be used,
Based on the observations collected for $s(1)$ and its retained part $s(2m)$, we define an unbiased multivariate difference estimator for $\bar{Y}_2$ as

$$
\bar{y}_{2md} = \sum_{i=1}^{p} \omega_i \alpha_i, \quad \alpha_i = \bar{y}(2) - \bar{y}(1)(\bar{x}_i(2) - \bar{x}_i(1))
$$

(3.3.3)

where $\alpha_i$'s are suitably chosen constants and $\omega = (\omega_1, \ldots, \omega_p)'$ is the weight vector such that $\sum_{i=1}^{p} \omega_i = 1$.

Further, based on observations made for the independent sample $s(2u)$, an unbiased estimator of $\bar{Y}_2$ is defined as

$$
\bar{y}_{2u} = \frac{1}{u} \sum_{j=1}^{u} y_{2j}
$$

(3.3.4)

Combining the information on matched and unmatched
portions, an unbiased estimator of \( \bar{Y}_2 \) may be defined by,

\[
\hat{\bar{Y}}_2 = a\bar{y}_{2\text{md}} + (1-a) \bar{y}_{2\text{u}}
\]  

(3.3.5)

where \( a \) is a suitably chosen weight.

Following Tripathi (1970, 1976), we obtain

\[
V(\bar{y}_{2\text{md}}) = w'Bw, \quad B = (b_{ik})
\]

\[
b_{ik} = \frac{(n-1)N}{(N-1)nm} \left\{ V(y_2) - \ell_i c(y_2, x_i) - \ell_k c(y_2, x_k) 
+ \frac{1}{n} C(x_i, x_k) \right\} + \left( \frac{1}{n} - \frac{1}{N} \right) S_{2y}^2.
\]  

(3.3.6)

It may be noted that for all practical purposes the term

\[
\frac{(n-1)N}{n(N-1)}
\]

may be replaced by unity, in which case we may write

\[
V(\bar{y}_{2\text{md}}) = \frac{1}{n} Q_1 + \frac{1}{m} Q_2 + Q_3
\]  

(3.3.7)

where \( Q_1 = S_{2y}^2 \), \( Q_2 = w'Dw \), \( Q_3 = -\frac{1}{N} S_{2y}^2 \)

\[
D = (d_{ik})_{p \times p}, \quad d_{ik} = V(y_2) - \ell_i c(y_2, x_i) - \ell_k c(y_2, x_k) 
+ \frac{1}{n} C(x_i, x_k).
\]

We note that \( V(\bar{y}_{2\text{u}}) = \frac{1}{u} Q_4 + Q_5 \)

with \( Q_4 = Q_1 \) and \( Q_5 = Q_3 \).  

(3.3.8)
An unbiased estimator of \( \hat{V}(\bar{y}_{2m}) \) may be given by,

\[
\hat{V}(\bar{y}_{2m}) = w' \hat{B} w, \quad \hat{B} = (\hat{b}_{ik})
\]

with

\[
b_{ik} = \frac{1}{m} \left( \hat{c}(y_2) - \ell_i \hat{c}(y_2, x_i) - \ell_j \hat{c}(y_2, x_k) + \ell_i \ell_j \hat{c}(x_i, x_k) \right)
\]

\[
\quad + \frac{(N-n)}{nN} \hat{s}_{2y}^2
\]

where

\[
\hat{c}(x_i, x_k) = \frac{1}{m-1} \sum_{j=1}^{m} \left( \frac{x_{ij}}{n_j} - \bar{x}_i(2) \right) \left( \frac{x_{kj}}{n_k} - \bar{x}_k(2) \right)
\]

\[
\hat{V}(y_2) = \hat{c}(y_2, y_2) \quad \text{and}
\]

\[
\hat{s}_{2y}^2 = \frac{1}{m(n-1)} \left[ \sum_{j=1}^{m} \frac{y_{2j}^2}{p_j} - \frac{1}{n(n-1)} \sum_{j=k}^{m} \frac{y_{2j}}{p_j} y_{2k} / p_k \right]
\]

is an unbiased estimator of \( s_{2y}^2 \).

Substituting the values of \( Q_i \) (i=1,...,5) in (3.2.4), we obtain

\[
\hat{V}(\bar{y}_2) = \frac{S_{2y}^2 [1-(f'-f) \lambda] [\theta + \lambda(1-f)]}{n(n'-n-\lambda)[\theta + \lambda(1-2f)]+\lambda}
\]  \hspace{1cm} (3.3.9)

where \( f' = n'/N, f = n/N, \theta = Q_2 / S_{2y}^2 = w' Dw / S_{2y}^2 > 0. \)

Henceforth we assume that \( n' = n \), that is the sample size is same at both the occasions. After ignoring the terms like \( \lambda(1-\lambda)f, \lambda f(1-f), \lambda f \theta \) and \( \lambda(1-\lambda)f(1-f), \).
we get

$$V(\hat{Y}_2) = \frac{S^2_{2y}}{n} \left(1-f\right) \frac{\theta + \lambda}{(1-\lambda)(\theta + \lambda) + \lambda}$$ \hspace{1cm} (3.3.10)

3.4. Choice of Matched Proportion

In case all the units of the sample s(1) are retained, an unbiased estimator for $\bar{Y}_2$ is given by

$$\bar{Y}_{2(m=n)} = \frac{1}{n} \sum_{j=1}^{n} y_{2j} \quad j \in s(2m) \quad m = n$$ \hspace{1cm} (3.4.1)

with

$$V(\bar{Y}_{2(m=n)}) = \left(\frac{1}{n} - \frac{1}{N}\right) S^2_{2y}$$ \hspace{1cm} (3.4.2)

Similarly in case all units of s(1) are replaced by an independent sample s(2u)u = n, an unbiased estimator for $\bar{Y}_2$ is given by

$$\bar{Y}_{2(u=n)} = \frac{1}{m} \sum_{j=1}^{m} y_{2j} \quad j \in s(2u)u = n$$ \hspace{1cm} (3.4.3)

with $V(\bar{Y}_{2(u=n)})$ same as given by (3.4.2).

From (3.3.10) and (3.4.2) the gain by using the estimator $\hat{Y}_2$ in (3.3.5) based on arbitrary matching $\lambda (0 < \lambda < 1)$ over complete matching ($\lambda = 1$) and complete replacement ($\lambda = 0$) is given by,
$G = (1-f) \frac{1}{n} S_2^2 y \left( \frac{\lambda(1-\theta-\lambda)}{(1-\lambda)(\theta+\lambda)+\lambda} \right)$

(3.4.4)

It follows that $\hat{y}_2(2)$ would be better than

$\overline{y}_2(m=n)$ and $\overline{y}_2(u=n)$, if $0 < \lambda < 1-\theta$

(3.4.5)

The condition (3.4.5) indicates a policy of matching.

In populations where $\theta = \frac{w'dw}{S_2^2 y} \geq 1$, it would be advisable to

either use $\overline{y}_2(m=n)$ or $\overline{y}_2(u=n)$, while in other situations

it would be preferable to use $\hat{y}_2(2)$ with suitable choice of

$\lambda(0 \leq \lambda < 1)$. Table 3.4.1 gives the percent relative efficiency

(PRE) of $\hat{y}_2(2)$ over $\overline{y}_2(m=n)$ and $\overline{y}_2(u=n)$ for some

values of $\theta$ and $\lambda,$

$$\text{PRE} = \frac{(1-\lambda)(\theta+\lambda)+\lambda}{\theta+\lambda} \times 100.$$  

The table indicates that for $0 < \theta < 0.7$ it would be

preferable to choose $\lambda$ such that $0.2 \leq \lambda \leq 0.3.$

We give the following theorem.

Theorem 3.4.1. The optimum value of $\lambda$ which minimizes $V(\hat{y}_2)$

in (3.3.10) is given by

$$\lambda_o(\theta) = \sqrt{\theta}(1-\sqrt{\theta})$$

(3.4.6)

and the resulting minimum variance is given by
\[ V_o(\hat{Y}_2) = \frac{1}{n} S^2_{2y}[1+(1-\sqrt{\theta})^2]^{-1}, \text{assuming that } \theta < 1. \quad (3.4.7) \]

**Proof.** We get (3.4.6) easily by differentiating (3.3.10) w.r.t \( \lambda \) and equating to zero, whereas (3.4.7) is obtained by substituting optimum value value of \( \lambda_o \) in (3.3.10). Q.E.D.

It may be remarked that the assumption \( \theta < 1 \) is not quite restrictive. For example for \( p = 1, \)

The optimum value \( \hat{\lambda}_{01} = \frac{C(y_2,x_i)}{V(x_i)} = \frac{\delta\sigma(y_2)}{\sigma(x_i)} \) of \( \lambda_1 \) is used or for large \( m \) its estimated value \( \hat{\lambda}_{01} = \frac{\hat{C}(y_2,x_1)}{V(x_1)} \) is used, we have

\[ \theta = \frac{[\sigma^2(y_2) - 2\lambda_1 \delta \sigma(y_2) \sigma(x_1) + \lambda^2 \sigma^2(x_1)]}{S^2_{2y}} \]

which reduces to

\[ \theta = \frac{(1-\delta^2)\sigma(y)}{S^2_{2y}} \quad (3.4.8) \]

From (3.4.8) we note that in all the situations which are favourable to the use of PPSWR sampling compared to the use of simple random sampling, we will have \( \theta < 1. \)

It may be noted that the optimum weight vector \( w, \) in \( Q_2 = w'Dw \) is given by,

\[ w_0 = D^{-1}e/e'D^{-1}e, \quad e = (1, \ldots, 1)' \]
resulting into the minimum value of $Q_2$ as

$$Q_{02} = 1/e^{D-1}e.$$  

From (3.4.6), it is noted that $\lambda_o(\theta)$ assumes its maximum value $\lambda_o = \max_{0<\theta<1} \lambda_o(\theta) = 1/4$ at $\theta = 1/4$. This indicates that whatever be $0 < \theta < 1$, one should not retain more than 25% of the units in $s(1)$ for $s(2m)$. Table (3.4.2) gives optimum matched proportion $\lambda_o$ for various values of $\theta(0<\theta<1)$.

**TABLE 3.4.1: PRE of Arbitrary Matching over Complete Matching and Complete Replacement**

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>$\theta$</th>
<th>0.1</th>
<th>0.2</th>
<th>0.25</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
<th>0.95</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>140.0</td>
<td>123.3</td>
<td>118.6</td>
<td>115.0</td>
<td>110.0</td>
<td>106.7</td>
<td>104.3</td>
<td>102.5</td>
<td>101.1</td>
<td>100.0</td>
<td>99.5</td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>146.7</td>
<td>130.0</td>
<td>124.4</td>
<td>120.0</td>
<td>113.3</td>
<td>108.6</td>
<td>105.0</td>
<td>102.2</td>
<td>100.0</td>
<td>98.2</td>
<td>97.4</td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>146.4</td>
<td>130.6</td>
<td>125.0</td>
<td>120.4</td>
<td>113.5</td>
<td>108.3</td>
<td>104.4</td>
<td>101.3</td>
<td>98.8</td>
<td>96.7</td>
<td>95.8</td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>145.0</td>
<td>130.0</td>
<td>124.5</td>
<td>120.0</td>
<td>112.9</td>
<td>107.5</td>
<td>103.3</td>
<td>100.0</td>
<td>97.2</td>
<td>95.0</td>
<td>94.0</td>
<td></td>
</tr>
<tr>
<td>0.35</td>
<td>142.8</td>
<td>128.7</td>
<td>123.3</td>
<td>118.8</td>
<td>111.7</td>
<td>106.1</td>
<td>101.8</td>
<td>98.3</td>
<td>95.4</td>
<td>93.0</td>
<td>91.9</td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>126.7</td>
<td>126.7</td>
<td>121.5</td>
<td>117.1</td>
<td>110.0</td>
<td>104.4</td>
<td>100.0</td>
<td>96.4</td>
<td>93.3</td>
<td>90.8</td>
<td>89.6</td>
<td></td>
</tr>
<tr>
<td>0.45</td>
<td>136.8</td>
<td>124.2</td>
<td>119.3</td>
<td>115.0</td>
<td>107.9</td>
<td>102.4</td>
<td>97.8</td>
<td>94.1</td>
<td>91.0</td>
<td>88.3</td>
<td>87.1</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>133.3</td>
<td>121.4</td>
<td>116.7</td>
<td>112.5</td>
<td>105.6</td>
<td>100.0</td>
<td>95.4</td>
<td>91.7</td>
<td>88.5</td>
<td>85.7</td>
<td>84.5</td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 3.4.2: Optimum Matched Proportion**

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>0.1</th>
<th>0.2</th>
<th>0.25</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
<th>0.95</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_o$</td>
<td>0.21</td>
<td>0.24</td>
<td>0.25</td>
<td>0.24</td>
<td>0.23</td>
<td>0.20</td>
<td>0.17</td>
<td>0.13</td>
<td>0.09</td>
<td>0.04</td>
<td>0.02</td>
</tr>
</tbody>
</table>
3.5. Some Alternative Estimators

It may be noted that the variances in (3.3.7), (3.3.8) and (3.3.9) are exact for any sample sizes. The estimator \( \hat{V}_2 \) is quite useful in situations where the sample sizes are small. The problem arises, in practice, when the constants \( \lambda_i \) \((i=1,\ldots,p)\) are not chosen in advance and are made to depend on the sampling design. The constant \( \lambda_i \) may be taken as the good guessed value of \( R_i = \bar{Y}_2 / \bar{x}_i \) or \( B_{0i}^* = \frac{C(y_2,x_i)}{V(x_i)} \) or \( R_i^* = \bar{Y}_1 / \bar{x}_1 \) or \( B_{0i}^* = \frac{C(y_1,x_1)}{\sigma^2(x_1)} \) based on pilot survey or past experience where \( \bar{Y}_1 = \frac{1}{N} \sum_{j=1}^{N} Y_{1j} / N \) denotes the population mean of \( y \) on first occasion. In the absence of a suitable choice of these constants, following estimators based on large samples may be considered on the lines of Tripathi (1970, 1976).

A multivariate ratio-type estimator for \( \bar{Y}_2 \) based on the matched portion \( s(2m) \) may be defined as

\[
\bar{y}_{2mr} = \frac{1}{p} \sum_{i=1}^{p} w_i \alpha_i, \quad \alpha_i = \frac{\bar{y}(2)}{\bar{x}(1)} \bar{x}_i(1), \quad \sum_{i=1}^{p} w_i = 1 \quad (3.5.1)
\]

Assuming \( m \) to be large and ignoring the terms of order \( m^{-r} \), \( r > 1 \), the bias and the mean square error (MSE) are given by,

\[
B(\bar{y}_{2mr}) = \frac{(n-1)N}{(N-1)nm} w'b \quad (3.5.2)
\]
and \[ \text{M.S.E.}(\tilde{y}_{2\text{mr}}) = \frac{1}{n} Q_1 + \frac{1}{m} Q_2 + Q_3 \quad (3.5.3) \]

where \[ b = (b_1, b_2, \ldots, b_p)' \; ; \; b_i = \frac{1}{x_i} \left[ R_i V(x_i) - C(y_2, x_i) \right], R_i = \frac{\tilde{y}_2}{x_i}, \]
and \( Q_1, Q_2^* \) and \( Q_3 \) are defined as in (3.3.7) with \( l_i \) replaced by \( R_i = \frac{\tilde{y}_2}{x_i} \).

The estimator for \( \tilde{y}_2 \) may now be defined by (3.3.5) with \( \tilde{y}_{2\text{md}} \) replaced by \( \tilde{y}_{2\text{mr}} \).

Another estimator for \( \tilde{y}_2 \) may be defined by (3.3.5) through replacing \( \tilde{y}_{2\text{md}} \) by \( \tilde{y}_{2\text{mrg}} \), where

\[ \tilde{y}_{2\text{mrg}} = \frac{p}{\sum_{i=1}^{p} w_i a_i}, a_i = \frac{\tilde{y}(2) - \hat{\beta}_{oi}(\tilde{x}_i(2) - \tilde{x}_i(1))}{\sum_{j=1}^{m} \frac{x_{ij}}{np_j} - \tilde{x}_i(2)} \quad (3.5.4) \]

and \[ \hat{\beta}_{oi} = \frac{m}{\sum_{j=1}^{m} \frac{x_{ij}}{np_j} - \tilde{x}_i(2)} \left( \frac{\tilde{y}(2) - \tilde{x}_i(2)}{\sum_{j=1}^{m} \frac{x_{ij}}{np_j} - \tilde{x}_i(2)} \right) - \frac{\sum_{i=1}^{m} \frac{x_{ij}}{np_j} - \tilde{x}_i(2)}{2}. \]

For large \( m \),

\[ V(\tilde{y}_{2\text{mrg}}) = \frac{1}{n} Q_1 + \frac{1}{m} Q_2^** + Q_3 \quad (3.5.5) \]

where \( Q_1, Q_2^** \) and \( Q_3 \) are defined in the same way as in (3.3.7) with \( l_i \) replaced by \( \hat{\beta}_{oi} = C(y_2, x_i)/V(x_i) \).

Defining

\[ \rho_{oi} = \frac{C(y_2, x_i)}{\sigma(y_2) \sigma(x_i)} ; \quad \rho_{ik} = \frac{C(x_i, x_k)}{\sigma(x_i) \sigma(x_k)}, \quad (3.5.6) \]

the term \( Q_2^** \) reduces to
\[ Q_2^{**} = w'D^*w, \quad D^* = (d_{ik}^*)_{p \times p} \]  

\[ d_{ik}^* = [1 - \rho_{oi}^2 - \rho_{ok}^2 + \rho_{oi} \rho_{ok} \rho_{ik}]V(y_2) \]  

(3.5.7)

The other results related to \( \bar{y}_{2mr} \) and \( \bar{y}_{2mrg} \) follow from (3.3.9), (3.3.10) and the discussion in Section 3.4, through replacing \( \theta \) by \( \theta^* = Q_2^*/S_{2y}^2 \) and \( \theta^{**} = Q_2^{**}/S_{2y}^2 \) respectively.