Jour. Ind. Soc. Ag. Statistics
Vol. XXXIX, No. 1 (1987), pp. 34-48

RATIO ESTIMATOR WITH POST STRATIFICATION DESIGN INVOLVING DOUBLE SAMPLING APPROACH

A. S. SETHI
P. A. U. Sugarcane Research Station, Jalandhar
and
A. K. SRIVASTAVA
I.A.S.R.I, New Delhi

(Received : May, 1983)

SUMMARY

With a view to overcome the constraint imposed by the assumption viz., sample mean of auxiliary character should always be less than twice the population mean of the character, on the application of ordinary ratio estimator, an attempt has been made to develop ratio estimator with post stratification design. Approximate expressions for bias and variance of the estimator, based on double sampling approach have been presented for different situations concerning availability or not of strata means, strata sizes and strata totals for the auxiliary variable. Relative efficiency of the esrimator under the different situations has also been investigated through an empirical study.

Keywords : Bias; Variance; Separate ratio estimator with proportional allocation; Finite population correction; Relative efficiency.

Introduction

Consider a finite population of N distinct and identifiable units. Let y be the study character for which the population total Y is to be estimated. An auxiliary character x, highly correlated with y, is assumed to be available for the population. Let a simple random sample of n units be drawn from the population. Ratio estimator, which utilizes auxiliary information at the estimation stage may then be usefully applied for estimating Y.

For the validity of approximate expressions for bias and variance of the ratio estimator, it is required that sample mean of auxiliary character for all possible samples of a given size be less than twice the population mean of the character. But this assumption may sometimes fail to hold good, especially in populations exhibiting sufficiently large variability in x, thereby invalidating the applicability of the estimator. This limitation may be overcome through proper stratification such that within strata variability in x reduces and the required assumption is satisfied for each stratum. In practice, however, such a prior stratification may not be always feasible. We may, therefore, resort to post stratification which ensures more flexibility in the formation of strata after selection of the sample. This technique, commonly adopted in sampling designs (Williams, [7]; Hugh and Matts, [2]; Mehrotra *et al.*, [3]), consists in classifying the selected sample at the estimation stage into strata as if it were a stratified sample, eventually making sample size in each stratum a random variate.

Let L be the number of strata, N_i be the size of ith stratum $\left(\sum_{i=1}^{L} N_i = N\right)$ and n_i be the number of units coming from ith stratum $\left(\sum_{i=1}^{L} n_i = n\right)$. Then the ratio estimator with post stratification design would be

$$\hat{Y} = \sum_i \frac{\bar{y}_i}{\bar{x}_i} X_i \tag{1}$$

where $\bar{y}_i$ and $\bar{x}_i$ are the sample means of y and x, and X_i is the population total of x in ith stratum. This estimator differs from the usual separate ratio estimator due to the difference in the sampling design.

The X_i's are assumed to be known. But in many practical situations these may not be known. In this paper we shall investigate various alternatives to $\hat{Y}$ under different situations depending upon whether the X_i's are known or not.

2. Expressions for Bias and Variance

Keeping in view the randomness in the n_i's, expressions for expectation and variance of $\hat{Y}$ were derived (to the first order of approximation) by first considering the n_i's as fixed and then allowing for the variation over the possible values of the n_i's. The following approximate results, partly due to Stephan [5], were frequently used :

$$E(n_i) = nW_i \tag{2}$$

$$V(n_i) = fnW_i(1 - W_i) \tag{3}$$

$$\text{Cov}\,(n_i, n_j) = -fnW_iW_j \tag{4}$$

$$E(1/n_i) = \frac{1}{nW_i} + \frac{f(1 - W_i)}{n^2 W_i^2} \tag{5}$$

$$V(1/n_i) = \frac{f(1 - W_i)}{n^3 W_i^3} \tag{6}$$

$$\text{Cov}\,(1/n_i,\, 1/n_j) = -\frac{f}{n^3 W_i W_j} \tag{7}$$

where W_i is the relative size of the ith stratum and f is the finite population correction. The notations used subsequently are, in general, common and shall not be defined.

2.1 *When the X_i's are Known*

This situation, encountered most commonly in practice, arises when the information on x is available for all the units in the population. The estimator here, say $\hat{Y}_1$, is the same as given in (1) above. Its expectation would be obtained as

$$E(\hat{Y}_1) = E\{E(\hat{Y}_1 \mid n_i)\} \tag{8}$$

Following Sukhatme and Sukhatme [6], it could be seen that

$$E(\hat{Y}_1 \mid n_i\} = \sum_i N_i \overline{Y}_i \{1 + (1/n_i - 1/N_i)\, \delta_i\} \tag{9}$$

where

$$\delta_i = C_{ix}^2 - \rho_i C_{ix} C_{iy}$$

Therefore

$$E(\hat{Y}_1) = \sum_i N_i \overline{Y}_i [1 + \{E(1/n_i) - 1/N_i\}\, \delta_i]$$

Using (5), we find that the bias in $\hat{Y}_1$ is given by

$$B(\hat{Y}_1) = \frac{Nf}{n} \sum_i W_i \sigma_i - \frac{Nf}{n^2} \sum_i (1 - W_i)\, \sigma_i \tag{10}$$

where

$$\sigma_i = \overline{Y}_i \delta_i / W_i$$

The first term is the bias in the usual separate ratio estimator for Y under proportional allocation, and the second term is the additional component in bias resulting from the randomness in the n_i's.

Further, variance of the estimator would be obtained as

$$V(\hat{Y}_1) = E\{V(\hat{Y}_1 \mid n_i)\} + V\{E(\hat{Y}_1 \mid n_i)\} \tag{11}$$

Now

$$V(\hat{Y}_1 \mid n_i) = \sum_i V(\bar{y}_i/\bar{x}_i \mid n_i)\, X_i^2 + \sum_{\substack{i \\ i \neq j}} \sum_j \mathrm{Cov}(\bar{y}_i/\bar{x}_i, \bar{y}_j/\bar{x}_j \mid n_i\, n_j)\, X_i, X_j$$

The conditional covariance between $\bar{y}_i/\bar{x}_i$ and $\bar{y}_j/\bar{x}_j$ over the samples with fixed n_i and n_j may be seen to be zero, resulting in

$$V(\hat{Y}_1 \mid n_i) = \sum_i N_i^2\ (1/n_i - 1/N_i)\ \varphi_i$$

where

$$\varphi_i = S_{iy}^2 - 2R_i\, S_{iyx} + R_i^2 S_{ix}^2$$

On using (5), we get

$$E\Big\{V(\hat{Y}_1 \mid n_i)\Big\} = \frac{N^2 f}{n} \sum_i W_i\, \varphi_i + \frac{N^2 f}{n^2} \sum_i (1 - W_i)\varphi_i \tag{12}$$

From (9), we have

$$V\{E(\hat{Y}_1 \mid n_i)\} = \sum_i N_i^2\, \bar{Y}_i^2\, \delta_i^2\, V(1/n_i) + \sum_{\substack{i \\ i \neq j}} \sum_j N_i N_j \bar{Y}_i \bar{Y}_j \delta_i \delta_j$$

$$\mathrm{Cov}\,(1/n_i,\ 1/n_j) = \frac{N^2 f}{n^3} \sum_i W_i(\sigma_i - \bar{\sigma})^2 \tag{13}$$

[By using (6) and (7)]

where

$$\bar{\sigma} = \sum_i W_i \sigma_i$$

Hence from (11), (12) and (13), we get

$$V(\hat{Y}_1) = N^2\{(f/n) \sum_i W_i \varphi_i + (f/n^2) \sum_i (1 - W_i)\, \varphi_i + (f/n^3) \sum_i W_i\, (\sigma_i - \bar{\sigma})^2\} \tag{14}$$

Again, the first term on the right hand side of (14) denotes the variance of usual separate ratio estimator for Y under proportional allocation, whereas the second and the third terms are the additional components of variance due to the involvement of post stratification. However, the third term of $O(1/n^3)$ may be ignored for large samples.

2.2 *When the X_i's are not Known*

This situation arises from the lack of information on either strata means of x, or strata sizes, or both. For estimating the unknown X_i's, one may resort to double sampling approach (Han, [1]), in which a preliminary large sample of n' units is first drawn from the population. The main sample of n units is then drawn either as a sub-sample from the preliminary sample or independently from the population itself. The proposed estimator shall be discussed under each of the two sampling schemes, as follows :

2.2.1 MAIN SAMPLE DRAWN AS A SUB-SAMPLE FROM THE PRELIMINARY SAMPLE (SCHEME 1)

Let n_i' be the number of units in the first phase sample coming from the ith stratum, out of which a sub-sample of n_i units is drawn in the second phase ($\sum_i n_i = n$ and $\sum_i n_i' = n'$). Here the n_i's are bounded above by n_i''s which are random variates. The effect of randomness of the n_i may, however, be ignored provided n' is sufficiently large such that the probability of n_i' to be larger than n_i is sufficiently high (Singh, [4]). The expectation and variance of the proposed estimator for each of following sub-cases would, therefore, be obtained as

$$E(\hat{Y}) = E\{E(\hat{Y} \mid n_i')\} \tag{15}$$

$$V(\hat{Y}) = E\{V(\hat{Y} \mid n_i')\} + V\{E(\hat{Y} \mid n_i')\} \tag{16}$$

2.2.1.1 *N_i's are Known but $\bar{X}_i$'s are not Known*

Consider, for instance, a survey for estimating the total consumption of meat products in a multi-religion society. It is just possible that the distribution of religions may be known, but the average consumption of the products in the different religions may not be known. The estimator $\hat{Y}$ in such cases would become

$$\hat{Y}_2 = \sum_i n_i \frac{\bar{y}_i}{\bar{x}_i} \bar{x}_i' \tag{17}$$

where $\bar{x}_i'$ is the usual unbiased estimate of $\bar{X}_i$ from the n_i' units. It may be seen that

$$E(\hat{Y}_2 \mid n_i') = \sum_i N_i \bar{Y}_i \{1 + (1/n_i - 1/n_i') \delta_i\} \tag{18}$$

Since the N_i's are known, we may select the main sample with proportional allocation. Therefore,

$$E(\hat{Y}_2) = \sum_i N_i \bar{Y}_i \left[1 + \left\{ \frac{1}{nW_i} - E\left(\frac{1}{n_i'}\right) \right\} \delta_i \right]$$

Using (5), we see that the bias in $\hat{Y}_2$ is given by

$$B(\hat{Y}_2) = N(f/n - f'/n') \sum_i W_i \sigma_i - \frac{Nf'}{n'^2} \sum_i (1 - W_i) \sigma_i \tag{19}$$

Further, it may be seen that

$$V(\hat{Y}_2 \mid n_i') = \sum_i N_i^2 V\left(\frac{\bar{y}_i}{\bar{y}_i} x_i' \mid n_i' \right) \tag{20}$$

(Covariance terms under fixed n_i' and n_j' vanish)

$$= \sum_i N_i^2 \{(1/n_i - 1/n_i') \varphi_i + (1/n_i' - 1/N_i) S_{iy}^2\} \tag{21}$$

Therefore, taking $n_i = nW_i$ and using (5), we get

$$E\{V(\hat{Y}_2 \mid n_j')\} = \sum_i N_i^2 \left[\frac{f\varphi_i}{nW_i} + \frac{f'}{n'W_i} \left\{ 1 + \frac{1 - W_i}{n'W_i} \right\} \eta_i \right] \tag{22}$$

where

$$\eta_i = 2R_i S_{iyx} - R_i^2 S_{ix}^2$$

As in (13), it may be seen from (18) that

$$V\{E(\hat{Y}_2 \mid n_i')\} = \frac{N^2 f'}{n'^3} \sum_i W_i (\sigma_i - \bar{\sigma})^2 \tag{23}$$

Hence from (16), (22) and (23), we get

$$V(\hat{Y}_2) = N^2 [(f/n) \sum_i W_i \varphi_i + (f'/n') \sum_i W_i \eta_i + (f'/n'^2) \sum_i (1 - W_i) \eta_i + (f'/n'^3) \sum_i W_i (\sigma_i - \bar{\sigma})^2] \tag{24}$$

2.2.1.2 *$\overline{X}_i$'s are Known but N_i's are not Known*

Such a situation is faced, for example, in a survey for estimating average monthly expenditure using income as auxiliary variable. Average income of the individuals in the different income groups may be known, but the number of persons belonging to the different groups may not be known. Since N_i may be estimated unbiasedly by $(N/n')\, n_i'$ so the estimator $\hat{Y}$ would become

$$\hat{Y}_3 = (N/n') \sum_i n_i' (\bar{y}_i/\bar{x}_i)\, \overline{X}_i \tag{25}$$

Since the N_i's are not known, we may select n_i's in proportion to the n_i''s. Therefore,

$$E(\hat{Y}_3 \mid n_i') = (N/n') \sum_i n_i' \overline{Y}_i \left\{ 1 + \left(\frac{n'}{nn_i'} - \frac{1}{N_i} \right) \delta_i \right\} \tag{26}$$

from which it is easy to see that the bias in $\hat{Y}_3$ reduces to

$$B(\hat{Y}_3) = \frac{Nf}{n} \sum_i W_i \sigma_i$$

Now

$$V(\hat{Y}_3 \mid n_j') = (N^2/n'^2) \sum_i n_i'^2 (1/n_i - 1/N_i)\, \varphi_i \tag{27}$$

Therefore, taking $n_i = (n/n')\, n_i'$, we get

$$E\{V(\hat{Y}_3 \mid n_i')\} = (N^2/n'^2) \sum_i \left[(n'/n)\, E(n_i') - \frac{V(n_i') + \{E(n_i')\}^2}{N_i} \right] \varphi_i$$

[Using (2) and (3)]

$$= (N^2 f/n) \sum_i W_i \varphi_i - \frac{Nf'}{n'} \sum_i (1 - W_i)\, \varphi_i \tag{28}$$

From (26) it may be seen that

$$V\{E(\hat{Y}_3 \mid n_i')\} = (N^2/n'^2) \Big[\sum_i (\overline{Y}_i - \sigma_i/N)\, V(n_i')$$

$$+ \sum_{i} \sum_{j \atop i \neq j} (\overline{Y}_i - \sigma_i/N)(\overline{Y}_j - \sigma_j/N) \operatorname{Cov}(n_i', n_j') \Big]$$

[Using (3) and (4)]

$$= (N^2 f'/n') \sum_i W_i(\bar{Y}_i - \bar{Y})^2 - (2Nf'/n') \sum_i W_i(\bar{Y}_i - \bar{Y})(\sigma_i - \bar{\sigma})$$

$$+ (f'/n' \sum_i W_i(\sigma_i - \bar{\sigma})^2 \tag{29}$$

Hence from (16), (28) and (29), we have

$$V(\hat{Y}_3) = N^2 \Big[\frac{f}{n} \sum_i W_i \varphi_i + \frac{f'}{n''} \sum_i W_i(\bar{Y}_i - \bar{Y})^2 - \frac{f'}{n'N} \sum_i (1 - W_i) \varphi_i$$

$$- \frac{2f'}{n'N} \sum_i W_i(\bar{Y}_i - \bar{Y})(\sigma_i - \bar{\sigma}) + \frac{f'}{n'N^2} \sum_i W_i (\sigma_i - \bar{\sigma})^2 \Big] \tag{30}$$

2.2.1.3 *Neither N_i's nor $\bar{X}_i$'s are Known*

Such a situation may occur, for instance, in a survey for estimating total area under vegetables using holding size as auxiliary variable. If the sampling frame of the cultivators consists of their names only, then neither N_i's nor $\bar{X}_i$'s are likely to be known. The estimatior $\hat{Y}$ would thus become

$$\hat{Y}_4 = (N/n') \sum_i n_i' (\bar{y}_i/\bar{x}_i)\, \bar{x}_i' \tag{31}$$

As in (18),

$$E(\hat{Y}_4 \mid n_i') = (N/n') \sum_i n_i' \bar{Y}_i \{1 + (1/n_i - 1/n_i')\, \delta_i\} \tag{32}$$

Again, if n_i's are selected in proportion to the n_i''s, then

$$E(\hat{Y}_4) = (N/n') \sum_i \bar{Y}_i \; \{E(n_i') + (n'/n - 1)\, \delta_i\}$$

which results in

$$B(\hat{Y}_4) = N(f/n - f'/n') \sum_i W_i \sigma_i \tag{33}$$

As in (21), it may be seen that

$$V(\hat{Y}_4 \mid n_i') = (N^2/n'^2) \sum n_i'^2 \{(1/n_i - 1/n_i')\, \varphi_i + (1/n_i' - 1/N_i)\, S_{iy}^2\}$$

Therefore, using (2) and (3),

$$E\{V(\hat{Y}_4 \mid n_i')\} = N^2(f/n - f'/n') \sum_i W_i\varphi_i + (N^2f'/n') \sum_i W_i S_{iy}^2$$

$$- (Nf'/n') \sum_i (1 - W_i) S_{iy}^2 \qquad (34)$$

It may also be seen from (32) that

$$V\{E(\hat{Y}_4 \mid n_i')\} = (N^2f'/n') \sum_i W_i(\overline{Y}_i - \overline{Y})^2 \qquad (35)$$

Hence from (16), (34) and (35), we obtain

$$V(\hat{Y}_4) = N^2\Bigg[\frac{f}{n}\sum_i W_i\varphi_i + \frac{f'}{n'}\sum_i W_i\eta_i + \frac{f'}{n'}\sum_i W_i(\overline{Y}_i - \overline{Y})^2$$

$$- \frac{f'}{n'N}\sum_i (1 - W_i) S_{iy}^2\Bigg] \qquad (36)$$

For fairly large N and n', if we ignore the terms of $O(1/n'^2)$, $O(1/n'^2)$, $O(1/n'N)$ and $O(1/n'N^2)$ in (24), (30) and (36), it follows that

$$V(\hat{Y}) \cong V + \begin{cases} V_2', \text{ if only } \overline{X}_i\text{'s are unknown } (\hat{Y} \equiv \hat{Y}_2) \\ V_3', \text{ if only } N_i\text{'s are unknown } (\hat{Y} \equiv \hat{Y}_3) \\ V_2' + V_3', \text{ if both } \overline{X}_i\text{'s and } N_i\text{'s are unknown } (\hat{Y} \equiv \hat{Y}_4) \end{cases}$$

where

$$V = (N^2f/n) \sum_i W_i\varphi_i, \; V_2' = (N^2f'/n') \sum_i W_i\eta_i, \quad \text{and}$$

$$V_3' = (N^2f'/n') \sum_i W_i(\overline{Y}_i - \overline{Y})^2$$

2.2.2 MAIN SAMPLE DRAWN INDEPENDENTLY FROM THE POPULATION (SCHEME 2)

Such situations are realized when information on auxiliary character collected by an independent agency is utilized. Here both n_i and n_i' are random variates, uncorrelated with each other, and may be seen to satisfy the following approximate relationships :

$$E(n_i'/n_i) = (n'/n)\left\{1 + \frac{f(1 - W_i)}{nW_i}\right\} \tag{37}$$

$$E(n_i'^2/n_i) = (n'^2/n)\{W_i + (f/n + f'/n')(1 - W_i)\} \tag{38}$$

$$V(n_i'/n_i) = (n'^2/n^2)(f/n + f'/n')\left(\frac{1 - W_i}{W_i}\right) \tag{39}$$

$$\text{Cov}(n_i', n_i'/n_i) = (f'n'/n)(1 - W_i) \tag{40}$$

$$\text{Cov}(n_i', n_j'/n_j) = -(f'n'/n)W_i \tag{41}$$

$$\text{Cov}(n_i'/n_i, n_j'/n_j) = -(n'^2/n^2)(f/n + f'/n') \tag{42}$$

Now again we shall investigate the estimator $\hat{Y}$ for each of the three alternative situations using the formulae (15) and (16), but with the difference that the conditional expectation and conditional variance shall be worked out over the samples with fixed n_i's and n_i''s.

2.2.2.1 *N_i's are Known but $\overline{X}_i$'s are not Known*

Here the estimator (say, $\hat{Y}_5$) will be similar to $\hat{Y}_2$. Due to the independence between $\bar{y}_i/\bar{x}_i$ and $\bar{x}_i'$, the conditional expectation of $\hat{Y}_5$ will be the same as given in (9), with the result that $B(\hat{Y}_5) = B(\hat{Y}_1)$. Further, as in (21), it may be seen that

$$V(\hat{Y}_5 \mid n_i, n_i') = \sum_i N_i^2 \{(1/n_i - 1/N_i)\varphi_i + (1/n_i' - 1/N_i)\xi_i\} \tag{43}$$

where

$$\xi_i = R_i^2 S_{ix}^2$$

Therefore,

$$\begin{aligned} E\{V(\hat{Y}_5 \mid n_i, n_i') &= \sum_i N_i^2 [\{E(1/n_i) - 1/N_i\}\varphi_i + \{E(1/n_i') - 1/N_i\}\xi_i] \\ &= N^2[(f/n)\sum_i W_i\varphi_i + (f/n^2)\sum_i (1 - W_i)\varphi_i \\ &\quad + (f'/n')\sum_i W_i\xi_i + (f'/n'^2)\sum_i (1 - W_i)\xi_i] \end{aligned} \tag{44}$$

It may also be seen that

$$V\{E(\hat{Y}_5 \mid n_i, n_i')\} = (N^2 f/n^3)\sum_i W_i(\sigma_i - \bar{\sigma})^2 \tag{45}$$

Hence from (44) and (45),

$$V(\hat{Y}_5) = V(\hat{Y}_1) + N^2 \{(f'/n') \sum_i W_i \xi_i + (f'/n'^2) \sum_i (1 - W_i) \xi_i\} \quad (46)$$

2.2.2.2 *$\overline{X}_i$'s are Known but N_i's are not Known*

Here the estimator, say $\hat{Y}_6$, will be similar to $\hat{Y}_3$. As in (26),

$$E(\hat{Y}_6 \mid n_i, n_i') = (N/n') \sum_i n_i' \overline{Y}_i \{1 + (1/n_i - 1/N_i) \delta_i\} \quad (47)$$

Therefore,

$$E(\hat{Y}_6) = (N/n') \sum_i \overline{Y}_i [E(n_i') + \{E(n_i'/n_i) - E(n_i')/N_i\} \delta_i]$$

Using (2) and (37) it may be seen that the bias in $\hat{Y}_6$ also is the same as that in $\hat{Y}_1$. Further, the conditional variance of $\hat{Y}_6$ is the same as that of $\hat{Y}_3$ given in (27). Thus,

$$E\{V(\hat{Y}_6 \mid n_i, n_i')\} = (N^2/n'^2) \sum_i \{E(n_i'^2/n_i) - E(n_i'^2)/N_i\} \varphi_i$$

[Using (2), (3) and (38)]

$$= N^2\{(f/n) \sum_i W_i \varphi_i + (f/n^2) \sum_i (1 - W_i) \varphi_i + (f/n)(f'/n') \sum_i (1 - W_i) \varphi_i\} \quad (48)$$

From (46),

$$V\{E(\hat{Y}_6 \mid n_i, n_i')\} = (N^2/n'^2) [\sum_i \{(\overline{Y}_i - \sigma_i/N)^2 V(n_i')$$

$$+ 2W_i\sigma_i(\overline{Y}_i - \sigma_i/N) \text{Cov}(n_i', n_i'/n_i)$$

$$+ W_i^2 \sigma_i^2 V(n_i'/n_i)\} + \sum_{i} \sum_{\substack{j \\ i \neq j}} \{(\overline{Y}_i - \sigma_i/N)(\overline{Y}_j - \sigma_j/N)$$

$$\times \text{Cov}(n_i', n_j') + 2W_j\sigma_j(\overline{Y}_i - \sigma_i/N) \text{Cov}(n_i', n_j'/n_j)$$

$$+ W_i W_j \sigma_i \sigma_j \text{Cov}(n_i'/n_i, n_j'/n_j)\}]$$

Using (3), (4), (39), (40), (41) and (42), we get

$$V\{E(\hat{Y}_6 \mid n_i, n_i')\} = N^2\left\{\frac{f'}{n'}\sum_i W_i(\bar{Y}_i - \bar{Y})^2 + \frac{2ff'}{nn'}\sum_i W_i(\bar{Y}_i - \bar{Y})(\sigma_i - \bar{\sigma}) + \frac{f}{n^3}\sum_i W_i(\sigma_i - \bar{\sigma})^2 + \frac{f^2f'}{n^2n'}\sum_i W_i(\sigma_i - \bar{\sigma})^2\right\} \quad (49)$$

Hence using (48) and (49) in (16), we obtain

$$V(\hat{Y}_6) = V(\hat{Y}_1) + N^2\left\{\frac{f'}{n'}\sum_i W_i(\bar{Y}_i - \bar{Y})^2 + \frac{ff'}{nn'}\sum_i (1 - W_i)\varphi_i + \frac{2ff'}{nn'}\sum_i W_i(\bar{Y}_i - \bar{Y})(\sigma_i - \bar{\sigma}) + \frac{f^2f'}{n^2n'}\sum_i W_i(\sigma_i - \bar{\sigma})^2\right\} \quad (50)$$

2.2.2.3 *Neither N_i's nor $\bar{X}_i$'s are Known*

Here the estimator, say $\hat{Y}_7$, would be identical to $\hat{Y}_4$. It may be seen that the conditional expectation of $\hat{Y}_7$ is the same as that of $\hat{Y}_6$ given in (47). Consequently, the bias in $\hat{Y}_7$ again will be the same as that in $\hat{Y}_1$. Moreover, $V\{E(\hat{Y}_7 \mid n_i, n_i')\}$ will be the same as $V\{E(\hat{Y}_6 \mid n_i, n_i')\}$ given in (49). Further, as in (44) and (48),

$$E\{V(\hat{Y}_7 \mid n_i, n_i')\} = N^2\left\{\frac{f}{n}\sum_i W_i\varphi_i + \frac{f}{n^2}\sum_i (1 - W_i)\varphi_i + \frac{ff'}{nn'}\sum_i (1 - W_i)\varphi_i + \frac{f'}{n'}\sum_i W_i\xi_i - \frac{f'}{n'N}\sum_i (1 - W_i)\xi_i\right\} \quad (51)$$

Hence it follows that

$$V(\hat{Y}_7) = V(\hat{Y}_1) + N^2 \left\{ \frac{f'}{n'} \sum_i W_i \xi_i + \frac{f'}{n'} \sum_i W_i (\overline{Y}_i - \overline{Y})^2 + \frac{ff'}{nn'} \sum_i (1 - W_i) \varphi_i + \frac{2ff'}{nn'} \sum_i W_i (\overline{Y}_i - \overline{Y}) (\sigma_i - \bar{\sigma}) - \frac{f'}{n'N} \sum_i (1 - W_i) \xi_i + \frac{f_2 f'}{n^2 n'} \sum_i W_i (\sigma_i - \bar{\sigma})^2 \right\} \quad (52)$$

Again, for sufficiently large N and n', terms of $O(1/n'^2)$, $O(1/n'N)$ and $O(1/n^2 n')$ in (46), (50) and (52) may be neglected. Eventually, variance of the proposed estimator under this sampling scheme may be expressed as

$$V(\hat{Y}) \cong V(\hat{Y}_1) + \begin{cases} V_5', \text{ if only } \overline{X}_i\text{'s are unknown } (\hat{Y} \equiv \hat{Y}_5) \\ V_6', \text{ if only } N_i\text{'s are unknown } (\hat{Y} \equiv \hat{Y}_6) \\ V_5' + V_6' \text{ if both } \overline{X}_i\text{'s and } N_i\text{'s are unknown } (\hat{Y} \equiv \hat{Y}_7) \end{cases}$$

where

$$V_5' = \frac{N^2 f'}{n'} \sum_i W_i \xi_i,$$

and

$$V_6' = \frac{N^2 f'}{n'} \left\{ \sum_i W_i (\overline{Y}_i - \overline{Y})^2 + \frac{f}{n} \sum_i (1 - W_i) \varphi_i + \frac{2f}{n} \sum_i W_i (\overline{Y}_i - \overline{Y}) (\sigma_i - \bar{\sigma}) \right\}$$

Keeping in view the equality of bias in $\hat{Y}_1$, $\hat{Y}_5$, $\hat{Y}_6$ and $\hat{Y}_7$, mean squared errors of these estimators may be similarly expressed.

3. Empirical Study

In order to illustrate relative performance of the different alternatives to $\hat{Y}$ under a small population, data on production (y) and acreage (x)

of sugarcane* for 122 cane growing districts in the sub-tropical zone of India for the year 1983-84 were utilized. The objective was to estimate total cane production (Y) in the zone from a random sample of, say, 30 districts. Though the coefficient of correlation between y and x in the population was very high ($\rho = 0.985$), yet the variation in x (ranging from 10 ha to 178.7 × 10³ ha) was so large that the required condition $0 < \bar{x} < 2\bar{X}$ failed to hold good for some of the samples of size 30, thus disallowing the use of ratio estimator. However, by subjecting the sample and the population to post stratification into 4 categories with N_i's as 20, 45, 41 and 16, and n_i's as 5, 11, 10 and 4, the above condition was satisfied within each stratum. A preliminary sample of 61 districts was selected with n_i''s in the 4 strata as 10, 23, 20 and 8, respectively. The n_i's were randomly selected out of the n_i''s. Such a selection was arbitrarily done for the present illustration so that the assumptions regarding sample allocation made under the sections (2.2.1.1), (2.2.1.2) and (2.2.1.3) were satisfied.

Bias and variance of the proposed estimator were worked out (Table 1) for the different situations from the population values. Precision of the

TABLE 1—RELATIVE PERFORMANCE OF THE RATIO ESTIMATOR WITH POST STRATIFICATION UNDER DIFFERENT SITUATIONS

Estimator	*Estimate* ($x\ 10^6\ t$)	*Bias* ($x\ 10^6$)	*Variance* ($x\ 10^{12}$)	*M.S.E.* ($x\ 10^{12}$)	*Rel. Eff.* w.r.t. $\hat{Y}_{SRS}$	w.r.t. $\hat{Y}_1$	*% S. E.*
$\hat{Y}_{SRS}$	106.7	—	1060.5	1060.5	1.00	0.04	30.5
$\hat{Y}_1$	97.3	−1.00	38.6	39.6	26.78	1.00	6.4
$\hat{Y}_2$	97.9	−0.52	183.2	183.5	5.78	0.22	13.8
$\hat{Y}_3$	96.7	−0.82	237.9	238.5	4.45	0.17	15.8
$\hat{Y}_4$	97.3	−0.55	365.6	365.9	2.90	0.11	19.5
$\hat{Y}_5$	97.9	−1.00	151.9	152.9	6.94	0.26	12.5
$\hat{Y}_6$	96.7	−1.00	239.7	240.7	4.41	0.16	15.9
$\hat{Y}_7$	97.3	−1.00	336.8	337.8	3.14	0.12	18.7

*Source : Agricultural Situation in India, Vol. 40, No. 1 : 58-61.

estimator reduced considerably if the strata totals of x were not known. However, the estimator under each situation was better (judged from its relative efficiency and % standard error), though to varying degrees, than the ordinary simple random sampling estimate ($\hat{Y}_{SRS}$).

ACKNOWLEDGEMENT

We are grateful to the referee for his invaluable suggestions for the improvement of the paper.

REFERENCES

[1] Han, C. P. (1973) : Double sampling with partial information on auxiliary variable, *Jour. Amer. Stat. Assoc.*, 68 (344) : 914-918.

[2] Hugh, R. M. and J. Matts (1983) : Post-stratification in the randomized clinical tests, *Biometrics*, 39 (1) : 217-225.

[3] Mehrotra, P. C., A. K. Srivastava and K. K. Tyagi (1984) : On post-stratification for cluster sampling, *Jour. Indian Soc. Agri. Stat.*, 36 (3) : 98-104.

[4] Singh, B. D. (1962). Use of double sampling in repeated surveys. Unpub. thesis of Dip. in Agri. and Ani. Husb. Stat., *Inst. Agri. Res. Stat., New. Delhi.*

[5] Stephan, F. F. (1945). The expected value and variance of the reciprocal and other negative powers of a positive Bernoullian variate. *Ann. Math. Stat.*, 16 : 50-61.

[6] Sukhatme, P. V. and B. V. Sukhatme (1970) : *Sampling Theory of Surveys with Applications*. Second Ed., Iowa State University Press, Ames (Iowa).

[7] Williams, W. H. (1962) : The variance of an estimator with post-stratified weighing, *Jour. Amer. Stat. Assoc.*, 57 : 622-627.