Jour. Ind. Soc. Ag. Statistics
Vol. XLI, No. 1 (1989), pp. 119-131

ROBUST ESTIMATION IN STRATIFIED SAMPLING WITH MULTI-AUXILIARY VARIABLES

RAJU SINGH, B. V. S. SISODIA* AND R.B. SINGH
Bihar University, Muzaffarpur

(Received : December, 1987)

SUMMARY

Considering the more general linear model, the work of Royall and Herson [2] has been generalised in the light of the theory developed by Holt [4]. It has been found that the two techniques Stratification and Generalized Balanced Sampling provides more efficient protection against the model failure than does Generalised Balanced Sampling alone.

Keywords : Geographical stratification, Super-population probability model, Generalized balanced sample, Generalized stratified balanced sample, Robustness.

Introduction

The stratified sampling under size stratification was considered as an alternative to balanced sampling by Royall and Herson [1], [2]. They showed that stratified sampling together with balanced sampling provided more efficient protection against errors in the model than the balanced sampling could do alone. In many practical situations, the availability of more auxiliary variables related to the variable of interest cannot be ruled out in addition to size variable. Holt [3] has generalised the work of Royall and Herson [1] by considering the general

*Present Address : Department of Statistics, N.D.U.A.T., Kumarganj, Faizabad.

linear model. The present work is the generalisation of the work of Royall and Herson [2] in the light of theory developed by Halt [3]. The stratification based on more auxiliary variables in general possess practical problems. For simplicity we here consider geographical stratification. The effect of misspecification of the model on the estimate of total (or mean) is studied. The choice of sample in which many super population probability model leading to the same optimal estimator is discussed.

2. The General Linear Model

We assure that there are p-auxiliary variables $X_1, X_2 \ldots X_p$ related with character of interest Y and their values, like population totals $T(X_1), T(X_2), \ldots$ are supposed to be known in the finite population of size N labelled $1, 2, \ldots N$. A sample s of n units is to be selected from the finite population and the y-values (Character of interest) along with the values on auxiliary variables associated with the sample units are to be observed in order to estimate the population total $T = \sum_{k=1}^{N} y_k$.

Here the numbers $y_1, y_2, \ldots y_N$ whose sum we want to estimate are treated as realised values of independent random variables $Y_1, Y_2, \ldots Y_N$ such that

$$Y_k = \beta_1 x_{1k} + \beta_2 x_{2k} + \ldots + \beta_p x_{pk} + \varepsilon_k V_k^{\frac{1}{2}}$$

$$k = 1, 2, \ldots, N \qquad (2.1)$$

where $\varepsilon_1, \varepsilon_2, \ldots \epsilon_N$ are independent random variables with mean zero and variance σ^2 and $x_{1k}, x_{2k}, \ldots x_{pk}$ are values of $X_1, X_2, \ldots X_p$ associated with the unit—k. V_k is some form of variance of Y_k depending upon the auxiliary variables. In fact the population total which we want to estimate is an outcome of $Y = \sum_{i=1}^{N} Y_k$. The expected value and variance of Y_k are therefore, $\sum_{i=1}^{p} \beta_i x_{ik}$ and $\sigma^2 V_k$, respectively,

Now, since we assume that the whole population is divided into H geographical strata of sizes $N_1, N_2, \ldots N_H$ such that $\sum_{h=1}^{H} N_h = N$, and samples of sizes $n_1, n_2 \ldots n_H$ are selected independently from strata

1, 2, ... H, such that $\sum_{h=1}^{H} n_h = n$, the model (2.1) can thus for such situation be written as

$$Y_{hk} = \beta_1\, x_{1hk} + \beta_2\, x_{2hk} + \ldots + \beta_p\, x_{phk} + \epsilon_{hk}\, V_{hk}^{\frac{1}{2}}$$
$$h = 1, 2, \ldots H, \; k = 1, 2, \ldots N \qquad (2.2)$$

The expected value and variance of Y_{hk} are, therefore, $\sum_{i=1}^{p} \beta_i\, x_{ihk}$ and $\sigma^2\, V_{hk}$ respectively. Note that slopes are assumed to be same in all the strata. The above model (2.2) in matrix notation is as follows :

$$Y_N = X_{Np}\, \beta_p + \varepsilon_N$$

$$E\,(\varepsilon_N) = 0, \quad E\,(\varepsilon_N\, \varepsilon'_N) = \sigma^2\, \Sigma_{NN} \qquad (2.3)$$

where Y_N is $N \times 1$ random vector of N random variables such that first N_1 are of first stratum followed by N_2 variables of second stratum and so on, Σ_{NN} is $N \times N$ diagonal matrix.

Our objective is to estimate the finite population total

$$T = I'_N\, y_N$$

which is outcome of $I'_N\; Y_N = \sum_{h=1}^{H} I'_{N_h}\; Y_{N_h}$ (2.4)

where I_N is $N \times 1$ unit vector.

Before a survey is conducted Y_N is unknown. In making estimate for the population total the samples are selected from strata independently (not necessary at random) and dependent variables values observed for each. We may think of the sampling procedure as a partitioning of the matrix Σ_{NN} in (2.3) as follows :

$$\Sigma_{NN} = \begin{bmatrix} \Sigma_{nn} & \Sigma_{nm} \\ \Sigma_{mn} & \Sigma_{mm} \end{bmatrix}$$

where $m = N - n = \sum_{h=1}^{H} (N_h - n_h) = \sum_{h=1}^{H} m_h$; m_h is the number of non-sampled units in hth stratum. Also, since random variables ε_{hk} are independent and hence

$$\Sigma_{NN} = \begin{bmatrix} \Sigma_{nn} & 0 \\ 0 & \Sigma_{mm} \end{bmatrix}; \quad \Sigma_{nn} = \begin{bmatrix} \Sigma_{n_h n_h} & 0 \\ 0 & \Sigma_{m_h m_h} \end{bmatrix}$$

where $\Sigma_{n_h n_h}$ and $\Sigma_{m_h m_h}$ are diagonal matrices of known constants V_{hk}

3. The Estimation Strategy

We consider the model (2.3) to be the true model. But the exact specification of the model is however not always possible in practice, The model may be under-specified due to non inclusion of relevant independent variables. On the other hand the model may be over-specified due to inclusion of irrelevant independent variables. Since our objective is to estimate the population total not the co-efficient β_p, the population total of independent variable in the model must be konwn. Let there be a practical situation where some of the columns of X_p are known or mistakenly ignored, i.e., the information is available on $q < p$ auxiliary variables. Naturally the model (2.3) will be reduced to

$$Y_N = X_{Nq}\, b_q + \epsilon_N \tag{3.1}$$

and that the observed variables follows the model

$$Y_n = X_{nq}\, b_q + \varepsilon_n \tag{3.2}$$

whsch is under-specified.

Holt [3] developed an unbiased estimator of the population total in unstratified population under the model (3.1), which is given by

$$\hat{Y} = I_n' Y_n + I_m' X_{mq} \hat{b}_q \tag{3.3}$$

where $\hat{b}_q$ is the weighted least-square estimate of b_q given by

$$\hat{b}_q = (X_{nq}' W_{nn}^{-1} X_{nq})^{-1} X_{nq}' W^{-1} Y_n \tag{3.4}$$

W is a set of weight used in the estimation. Also, $\hat{Y}$ is best linear unbiased estimator (*BLUE*) if $W_{nn} = \Sigma_{nn}$. It is obvious from (3.3) that the observed Y_n are used directly in $\hat{Y}$ and estimate is made up for the m non-sampled elements of the finite population by using weighted least-squares estimate of b and then using the known X_{mq} values to predict the corresponding dependent variable values. In fact the individual values of the q-auxlliary variables are not strictly needed since $I_m' X_{mq}$ is simply the vector of non-sampled population total for each variable.

The estimator $\hat{Y}$ was unbiased under the model (3.1) with the variance equal to

$$\begin{aligned} E\xi (\hat{Y} - T)^2 &= E\xi\, [I_n' Y_n + I_m' X_{mq} \hat{b}_q - I_N' Y_N]^2 \\ &= \sigma^2 [\, I_m' X_{mq} X_{qn}^{-} \Sigma_{nn} (X_{qn}^{-})' X_{mq}' I_m \\ &\quad + I_m' \Sigma_{mm} I_m \,] \end{aligned} \tag{3.5}$$

where $X_{qn}^{-} = (X_{nq}' W_{nn}^{-1} X_{nq})^{-1} X_{n\,q}' W_{nn}^{-1}$ is the generalized inverse of X_{nq}.

If the assumed model (3.1) was correct the question of bias did not arise. We had at our disposal the choice of both W_{nn} and the n members of the sample in order to minimize the variance. The optimal design was to choose $W_{nn} = \Sigma_{nn}$. If the assumed model (3.1) was incorrectly or incompletely specified in terms of the independent variables, i.e. if the true model was (2.3) then the possibility of bias could exist. The bias is therefore given by

$$E\xi\,[\hat{Y} - T] = E\xi\,[I_n' Y_n + I_m' X_{mq} \hat{b}_q - I_N' Y_N]$$

$$= I_m' [X_{mq} X_{nq}^{-} X_{np} - X_{mp}]\, \beta_p \qquad (3.6)$$

Halt [3] showed that if one chooses the weight W used in the estimation of the co-efficient β in such a way that

$$W_{nn} = \text{diag}\left(\sum_{i=1}^{q} a_i x_{i1}, \ \sum_{i=1}^{q} a_i x_{i2}, \ldots \sum_{i=1}^{q} a_i x_{in} \right) \qquad (3.7)$$

x_{ik} is the value of the independent variable X, associated with the kth unit and the constants $a_1, a_2, \ldots a_q$ are any real numbers for which

$$\sum_{i=1}^{q} a_i x_{ik} > 0; \quad k = 1, 2, \ldots N$$

Then the bias given by (3.6) becomes zero under balanced sample and the estimator in this way i.e. with balanced sample becomes

$$\hat{Y} = \frac{N}{n} I_n' Y_n$$

Also under the appropriate variance structure $\Sigma_{nn} = W_{nn}$ this estimator is *BLU* where W is given by (3.7).

The condition for balanced sample is

$$n^{-1} I_n' X_{np} = m^{-1} I_m' X_{mp} \qquad (3.8)$$

That is the first moments of the sampled and non-sampled portions of the finite population are equal for each of the p-auxiliary variables.

Under balanced sample the veriance given by (3.5) reduces to

$$\text{Var}\,(\hat{Y}) = \sigma^2 \left[\left(\frac{m}{n} \right)^2 I_n' \Sigma_{nn} I_n + I_m' \Sigma_{mm} I_m \right]$$

$$= \sigma^2 \left[\left(\frac{m}{n} \right)^2 \Sigma_s \sum_{i=1}^{q} a_i x_{ik} + \Sigma_{\bar{s}} \sum_{i=1}^{q} a_i x_{ik} \right]$$

$$= \sigma^2 \left[(N - n) \frac{N}{n} \left(\sum_{i=1}^{q} a_i \bar{x}_i \right) \right] \quad (3.9)$$

Now, in stratified sampling we make estimate for each stratum total and sum up them to give the estimate of the population total. An unbiased estimator of the hth stratum total $Y_h = I'_{N_h} Y_{N_h}$ under the model (3.1) will be given by

$$\hat{Y}_h = I'_{n_h} Y_{n_h} + I'_{m_h} X_{m_h q} \hat{b}_{h_q} \quad (3.10)$$

where Y_{N_h} is $N_h \times 1$ vactor of variable of the character under study and Y_{n_h} is $n_h \times 1$ vector of the observed variable in the hth stratum. $\hat{b}_{n_q}$ is the weighted leaste square estimate of the co-efficients b_q based on the hth stratum sample and is given by

$$\hat{b}_{h_q} = (X'_{n_h q} W^{-1}_{n_h n_h} X_{n_h q})^{-1} X'_{n_h q} W^{-1}_{n_h n_h} Y_{n_h} \quad (3.11)$$

Also, the $\hat{Y}_h$ will be best linear unbiased estimator of Y_h when

$$W_{n_h n_h} = \Sigma_{n_h n_h}$$

$$W_{n_h n_h} = \text{diag.} \left(\sum_{i=1}^{q} a_i x_{ih1}, \sum_{i=1}^{q} a_i x_{ih2}, \ldots \sum_{i=1}^{q} a_i x_{ihn_h} \right) \quad (3.12)$$

The variance of estimator $\hat{Y}_h$ under the model (3.1) will be given by

$$E\xi \, [\hat{Y}_h - Y_h]^2 = \sigma^2 [I'_{m_h} X_{m_h q} X^{-}_{qn_h} \Sigma_{n_h n_h} X^{-}_{qn_h})' X'_{m_h q} I_{m_h}$$

$$+ I'_{m_h} \Sigma_{m_h m_h} I_{m_h}] \quad (3.13)$$

To obtain the estimate of the population total T we use the estimator $\hat{Y}_{st}$ given by

$$\hat{Y}_{st} = \sum_{h=1}^{H} \hat{Y}_h \quad (3.14)$$

This estimator is unbiased under the model (3.1) with variance equal to Var $(\hat{Y}_{st}) = E\xi [\hat{Y}_{st} - T]^2$

$$= \sigma^2 \sum_{h=1}^{H} \{ I'_{m_h} X_{m_{hq}} \bar{X}'_{qn_h} \Sigma_{n_h n_h} (\bar{X}_{qn_h})' X'_{m_h q} I_m$$

$$+ I'_{m_h} \Sigma_{m_h m_h} I_{m_h} \} \tag{3.15}$$

Since the true model is (2.3), the estimator $\hat{Y}_{st}$ under this model will be biased and the bias will be given by

$$E\xi (\hat{Y}_{st} - T) = E\xi \left[\sum_{h=1}^{H} \hat{Y}_h = \sum_{h=1}^{H} I'_{N_h} Y_{N_h} \right]$$

$$= \sum_{h=1}^{H} \{ ((I'_{m_h} X_{m_{hq}} \bar{X}_{qn_h} X_{n_h p} - I'_{m_h} X_{m_h p}) \beta_p \} \tag{3.16}$$

If the sample from stratum h is balanced i.e. if $n_h^{-1} I'_{nh} X_{nhp} = m_h^{-1} I'_{mh} X_{mhp}$ for all p, then $\hat{Y}_h$ is an unbiased estimate of Y_h under the general linear model (2.3) provided $W_{n_h n_h}$ has the form (3.12). If this is true for each stratum then the bias given by (3.16) is zero and thus the estimator $\hat{Y}_{st}$ becomes unbiased under this general linear model. We refer to such a sample as generalized stratified balanced sample. We denote it by $\hat{S}(p)$.

Under generalized stratified balanced sample i.e. $\bar{x}_{is_h} = \bar{x}_{i\bar{s}_h} = \bar{x}_{ih}$ the variance given by (3.15) reduces to

$$\text{Var} (\hat{Y}_{st}) = \sigma^2 \sum_{h=1}^{H} \left\{ (N_h - n_h) \frac{N_h}{n_h} \left(\sum_{i=1}^{q} a_i \bar{x}_{ih} \right) \right\} \tag{3.17}$$

where $\bar{x}_{is_h}$, $\bar{x}_{i\bar{s}_h}$ and $\bar{x}_{ih}$ are sample mean, non-sample mean and population mean of X_i in the h-th stratum.

4. Optimal Allocation for Stratified Generalized Balance Sampling

Consider a sample survey cost function as given

$$C = C_0 + \sum_{h=1}^{H} c_h n_h \tag{4.1}$$

where C_0 is a fixed amount, C_h is a cost for units sampled in stratum$-h$.

The optimum value of n_h is obtained by minimising $V(\hat{Y}_{st})$ subject to the fixed total cost given by (4.1). It is found that n_h must be proportional to

$$N_h \left(\sum_{i=1}^{q} a_i \bar{x}_{ih} \right)^{\frac{1}{2}}, \text{ i.e.,}$$

$$n_h = n\, N_h \left(\sum_{i=1}^{q} a_i \bar{x}_{ih} \right)^{\frac{1}{2}} \Big/ \sum_{h=1}^{H} \left\{ N_h \left(\sum_{i=1}^{q} a_i \bar{x}_{ih} \right)^{\frac{1}{2}} \right\} \tag{4.2}$$

when the above optimal allocation is used, the variance of $\hat{Y}_{st}$ in stratified grneralized balanced sampling is obtained as follows:

$$E\xi \left[\hat{Y}_{st} - T \right]^2 = \sigma^2 \left[\sum_{h=1}^{H} \left\{ \left(\frac{N_h^2}{n_h} - N_h \right) \left(\sum_{i=1}^{q} a_i \bar{x}_{ih} \right) \right\} \right]$$

$$= \frac{\sigma^2}{n} \left[\left\{ \sum_{h=1}^{H} N_h \left(\sum_{i=1}^{q} a_i \bar{x}_{ih} \right)^{\frac{1}{2}} \right\}^2 \Big/ n - N \sum_{i=1}^{q} a_i \bar{x}_i \right] \tag{4.3}$$

In the following section we shall now compare some strategies under generalized balanced sampling.

5. Comparison of some Strategies

When the variance function V_k and V_{hk} are linear combination of the values of the independent auxiliary variables associated with the unit-k and, if optimum allocation (4.2) is used, then under the more general linear model having p-auxiliary variables the variances of the estimator $\hat{Y}$ with generalized balanced sample and $\hat{Y}_{st}$ with stratified generalized balanced sample are given by (3.9) and (4.2) respectively.

The difference of the variances of these two strategies $[\hat{Y}\,;\ S(p)]$ and $[\hat{Y}_{st}\,;\ \hat{S}(p)]$ is

$$E\xi\, [\hat{Y} - T]^2 - E\xi\, [\hat{Y}_{st} - T]^2$$

$$= \frac{\sigma^2}{n} \left[N^2 \left(\sum_{i=1}^{q} a_i \bar{x}_i \right) - \left\{ \sum_{h=1}^{H} N_h \left(\sum_{i=1}^{q} a_i \bar{x}_{ih} \right)^{\frac{1}{2}} \right\}^2 \right]$$

$$\geqslant \frac{\sigma^2}{n} \left[N^2 \left(\sum_{i=1}^{q} a_i \bar{x}_i \right) \left\{ \left(\sum_{h=1}^{H} N_h \right) \left(\sum_{h=1}^{H} N_h \sum_{i=1}^{q} a_i \bar{x}_{ih} \right) \right\} \right]$$

(By cauchy-schwart inequality)

Thus the difference is non-negative and hence the strategy $[\hat{Y}_{st} ; \hat{S}(p)]$ is more efficient than the strategy $\left[\hat{Y} ; \hat{S}(p)\right]$.

6. Robustness and Efficiency of the Estimators

It is, however, not always possible to find $(p>1)$ auxiliary variables in practice each satisfying the condition (3.8). It is, therefore, important to investigate the robustness and efficiency of the proposed estimator as well as the estimator suggested by Holt [3] under unbalanced sample. To determine the robustness of the estimators a criterion has to be fixed up. We consider change in the amount of M.S.E. with the deviation of the model as a criterion to determine the robustness of the estimators. If the change in the amount of M.S.E. of an estimator with the deviation of the model is nominal, the estimator is said to be robust.

For simplicity we assume $q=1$ and $v=x$. Then the variance of the estimator $\hat{Y}$ and $\hat{Y}_{st}$ given by (3.5) and (3.15) respectively becomes:

$$\text{Var}\,(\hat{Y}) = E\xi\,[\hat{Y} - T]^2 = \sigma^2 N\left(\frac{N}{n} - n\right)\frac{\bar{x}_{\bar{s}1}\,\bar{X}_1}{\bar{x}_{s1}} \tag{6.1}$$

and

$$\text{Var}\,(\hat{Y}_{st}) = E\xi\,[Y_{st} - T]^2 = \sigma^2\left[\sum_{h=1}^{H}\left\{N_h\left(\frac{N_h}{n_h} - 1\right)\frac{\bar{x}_{s\bar{h}_1}\,\bar{X}_{h_1}}{\bar{x}_{sh}}\right\}\right] \tag{6.2}$$

Bias of the estimators $\hat{Y}$ and $\hat{Y}_{st}$ under the model (2.3) for $v = x$ are given respectively by

$$\text{Bias}\,(\hat{Y}) = E\xi\,[\hat{Y} - T]$$

$$= \sum_{j=1}^{P}\beta_j\,(N-n)\left\{\frac{\bar{x}_{\bar{s}1}\,\bar{x}_{sj}}{\bar{x}_{s1}} - \bar{x}_{\bar{s}_j}\right\} \tag{6.3}$$

and

$$\text{Bias}\,(Y_{st})\,E\varepsilon\,[\hat{Y}_{st} - T]$$

$$= \sum_{h=1}^{H}\left[\sum_{j=1}^{p}\beta_j\,(N_h - n_h)\left\{\frac{\bar{x}_{\bar{s}_h1}\,\bar{x}_{s_hj}}{\bar{x}_{s_h1}} - \bar{x}_{\bar{s}_{hj}}\right\}\right] \tag{6.4}$$

As the mean square error (M.S.E.) is sum of the variance and square of the bias, the M.S.E. of the estimators $\hat{Y}$ and $\hat{Y}_{st}$ under the model (2.3) for the variance function $v(x) = x$ is :

$$\text{M.S.E. } (\hat{Y}) = \sum_{j=1}^{p} \beta_j (N-n) \left\{ \frac{\bar{x}_{\bar{s}1} \, \bar{x}_{sj}}{\bar{x}_{s1}} - \bar{x}_{\bar{s}j} \right\}$$

$$+ \sigma^2 N \left(\frac{N}{n} - 1 \right) \frac{\bar{x}_{\bar{s}1} \, \bar{x}_1}{\bar{x}_{s1}} \tag{6.5}$$

$$\text{M.S.E. } (\hat{Y}_{st}) = \left(\sum_{h=1}^{H} \left[\sum_{j=1}^{p} \beta_j (N_h - n_h) \left\{ \frac{\bar{x}_{\bar{s}_h 1} \, \bar{x}_{s_h j}}{\bar{x}_{s_h 1}} - \bar{x}_{\bar{s}_h j} \right\} \right) \right)^2$$

$$+ \sigma^2 \sum_{h=1}^{H} \left\{ N_h \left(\frac{N_h}{n_h} - 1 \right) \frac{\bar{x}_{\bar{s}_h 1} \, \bar{x}_{h1}}{\bar{x}_{s_h 1}} \right\} \tag{6.6}$$

It is difficult to examine theoretically the efficiency and the robustness under deviation of the model of the estimators. For this purpose we consider the following three working models :

Model I : $E\xi\,(Y) = 2\,X_1$

Model II : $E\xi\,(Y) = 2\,X_1 - 1.5\,X_2$

Model III : $E\xi\,(Y) = 2\,X_1 - 1.5\,X_2 + X_3$

Let the population of interest consisting of $N = 30$ units be divided into $H = 3$ strata. Let $N_1 = 8$, $N_2 = 10$, $N_3 = 12$ be strata sizes and $n_1 = 2$, $n_2 = 3$ and $n_3 = 4$ be the sample sizes in the strata N_1, N_2 and N_3 respectively so that total sample is $n=9$.

Without loss of generality we may take stratum mean and two cases of unbalanced sample means of the auxiliary variables X_1, X_2, X_3 as given in the following table :

TABLE 1

auxiliary variable \ *Strata*		*Stratified sample means of strata means of auxiliary variables* 1 ($N_1=8, n_1=2$)	2 ($N_2=10, n_2=3$)	3 ($N_3=12, n_3=4$)	*Unstratified sample means & Population means* ($N=3, n=9$)
X_1	S.S.M. > S.M.	$\bar{x}_{s_1 1}=8$	$\bar{x}_{s_2 1}=9$	$\bar{x}_{s_3 1}=10$	$\bar{x}_{s1}=9.22$
	S.S.M. < S.M.	$\bar{x}_{s_1 1}=3$	$\bar{x}_{s_2 1}=6$	$\bar{x}_{s_3 1}=7$	$\bar{x}_{s1}=5.33$
	S.M.	$\bar{x}_{11}=6$	$\bar{x}_{21}=8$	$\bar{x}_{31}=9$	$\bar{x}_1=7.86$
X_2	S.S.M. > S.M.	$\bar{x}_{s_1 2}=6$	$\bar{x}_{22}=5$	$\bar{x}_{s_3 2}=9$	$\bar{x}_{s2}=7$
	S.S.M. < S.M.	$\bar{x}_{s_1 2}=4$	$\bar{x}_{32}=2$	$\bar{x}_{s_3 2}=6$	$\bar{x}_{s_2}=4.22$
	S.M.	$\bar{x}_{12}=5$	$\bar{x}_{22}=3$	$\bar{x}_{32}=7$	$\bar{x}_2=5.13$
X_3	S.S.M. > S.M.	$\bar{x}s_{1 3}=10$	$\bar{x}s_{2 3}=9$	$\bar{x}s_{3 3}=8$	$\bar{x}s_3=8.78$
	S.S.M. < S.M.	$\bar{x}s_{1 3}=4$	$\bar{x}s_{2 3}=3$	$\bar{x}s_3{}^3=2$	$\bar{x}s_3=2.78$
	S.M.	$\bar{x}_{13}=8$	$\bar{x}_{23}=7$	$\bar{x}_{33}=5$	$\bar{x}_3=6.46$

N.B. : *S.M.* = Stratum Mean, *S.S.M.* = Stratum Sample Mean

x^s_j = Total sample mean of the j-th auxiliary variable

x^s_{hj} = Stratum sample mean of j-th auxiliary variable.

Table 2 shows the difference of the M.S.E. for proposed estimator $\hat{Y}_{st}$ and the estimator $\hat{Y}$ suggested by Holt (1975) under different models for both the situations (sample mean greater than population mean and sample mean less than population mean) of stratified unbalanced sample respectively.

It is obvious from the Table 2 that the proposed estimator $\hat{Y}_{st}$ is more efficient than the estimator $\hat{Y}$ for stratified sample mean less than the population mean under the model I, II, and III. Also for stratified sample mean greater than population mean, it is clear from Table 2 that

TABLE 2 M.S.E. $(\hat{Y}_{st})$ – M.S.E. $(\hat{Y}_{tt})$

Model-I	Model-II	Model-III
$x_s > \bar{x}$ $2.0938\ \sigma^2$	$-166.16 + 2.0938\ \sigma^2$	$-43.17 + 2.0938\ \sigma^2$
$\bar{x}_s' > \bar{x}$ $-2.6730\ \sigma^2$	$-1501.12 - 2.6730\ \sigma^2$	$-6157.65 - 2.6730\ \sigma^2$

the estimator $\hat{Y}_{st}$ is efficient under the above models for $\sigma^2 < 20$ except the model—I.

Table 3 describes the robustness of the estimators $\hat{Y}_{st}$ and $\hat{Y}$. Obviously the difference in M.S.E. of the estimators $\hat{Y}_{st}$ and $\hat{Y}$ under model I and III is small. Such little difference will have negligible effect on the efficiency of the estimators if model I is used instead of model III or vice-versa and hence these estimators are robust for these models. Moreover, the differences in M.S.E. of both the estimators under model I and II and model II and III are considerably high and, therefore, these estimators are not robust in these situations analysed.

Also, it is clear from the Table 3 that in both the situations of unbalanced sample the absolute value of the differences of the M.S.E.'s of the estimator $\hat{Y}_{st}$ under different models is less than the absolute value of the differences of the M.S.E.'s of the estimator $\hat{Y}$ under different corresponding models. Hence the efficiency of the estimator $\hat{Y}_{st}$ is less affected relative to the estimator $\hat{Y}$ with the deviation of the model and so the estimator $\hat{Y}_{st}$ can be more robust than $\hat{Y}$.

TABLE 3—DIFFERENCE OF M.S.E'S OF THE ESTIMATORS UNDER DIFFERENT MODELS
(For $\bar{x}_s > \bar{x}$ and $\bar{x}_s < \bar{x}$)

Estimators \ Model		I-II	I-III	II-III
$\hat{Y}_{st}$	$\bar{x}_s > \bar{x}$	−1255.8200	−2.8600	1252.9606
	$\bar{x}_s < \bar{x}$	−1925.02	−10600.3200	−8675.3000
Y	$\bar{x}_s > \bar{x}$	−1421.9800	−46.0350	1375.9450
	$\bar{x}_s < \bar{x}$	−3426.14	−16757.97	−13331.83

REFERENCES

[1] Royall and Herson (1973) : Robust estimation in finite population I, *J. Amer. Statist. Assoc.*, 68 : 880-889.

[2] Royall and Herson (1973b) : Robust estimation in finite population II. *J. Amer. Statist. Assoc.*, 68 : 890-893.

[3] Holt, D. (1975) : A generalization of balanced sampling, *Sankhya*, Series—C, 37 : 199-203.

[4] Sukhatme, P. V. and Sukhatme, B. V. (1970) : *Sampling Theory of Surveys with Applications*, Indian Soc. Agril. Statistics. New Delhi, India.