Jour. Ind. Soc. Ag. Statistics
Vol. XL, No. 2 (1988), pp. 96-104

AN IMPROVED CLASS OF ESTIMATORS OF POPULATION MEAN USING AUXILIARY INFORMATION

HOUSILA P. SINGH
College of Agricultural Engineering, J. N. Agricultural University, Jabalpur-482 004

(Received : July, 1985)

SUMMARY

An efficient class of estimators for estimating population mean has been considered using supplementary information on two auxiliary variables, one of which is used at the sample selection stage and other for improving the estimator at the estimation stage. This generalizes the estimators suggested by Agarwal and Kumar [1] and Jhajj and Srivastava [2]. Asymptotic expressions for bias and mean squared error (MSE) of the estimators in the class have been obtained.

Keywords : Class of estimators, Auxiliary variables, Bias, Mean squared error.

Introduction

Suppose, information on two auxiliary variables highly correlated with the study variable y is available. Let a sample of size n be drawn with PPS sampling with replacement. Let P_i denote the probability of selection (based on one of the two auxiliary variables) of its unit a_i, $i = 1, 2, \ldots, N$. Let y_i and x_i denote respectively the values of the variable y under study and the auxiliary variable x for the ith unit of the population. Also let $\overline{Y}$ and $\overline{X}$ denote the respective population means.

Write, $u_i = \dfrac{y_i}{NP_i}, v_i = \dfrac{x_i}{NP_i}, \bar{u}_n = \dfrac{1}{n} \sum_{i=1}^{n} u_i, \bar{v}_n = \dfrac{1}{n} \sum_{i=1}^{n} v_i$

and

$$s_v^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (v_i - \bar{v}_n)^2.$$

The following notations are used :

$$\sigma_u^2 = \sum_{i=1}^{N} P_i (u_i - \bar{Y})^2, \ \sigma_v^2 = \sum_{i=1}^{N} P_i (v_i - \bar{X})^2, \ C_u^2 = \sigma_u^2/\bar{Y}^2,$$

$$C_v^2 = \sigma_v^2 / \bar{X}^2, \ \delta_{rs} = \sum_{i=1}^{N} P_i (u_i - \bar{Y})^r (v_i - \bar{X})^s, \ \rho_{uv} = \frac{\delta_{11}}{\sigma_u \sigma_v},$$

$$\psi = \frac{\delta_{12}}{\bar{Y} \sigma_v^2}, \ \beta_2 = \frac{\delta_{04}}{\sigma_v^4}, \ \alpha_1 = \frac{\delta_{03}}{\sigma_v^3} \text{ and } \beta_1 = \alpha_1^2.$$

Defining

$$W = \frac{\bar{v}_n}{\bar{X}}, \ Z = \frac{s_v^2}{\sigma_v^2}, \ \varepsilon = (\bar{u}_n - \bar{Y}), \ \delta = W - 1, \ \eta = Z - 1,$$

we have

$$E(\varepsilon) = E(\delta) = E(\eta) = 0, \ E(\varepsilon^2) = \frac{\bar{Y}^2}{n} C_u^2, \ E(\delta^2) = \frac{C_v^2}{n},$$

$$E(\eta^2) = \frac{1}{n} \left[\beta_2 - \left(\frac{n-3}{n-1}\right)\right] = \frac{(\beta_2 - 1)}{n} \quad \text{for } \frac{n-3}{n-1} = 1,$$

$$E(\varepsilon\delta) = \frac{\bar{Y}}{n} \rho_{uv} C_u C_v$$

$$E(\varepsilon\eta) = \frac{\bar{Y}}{n} \psi C_u \text{ and } E(\delta\eta) = \frac{\alpha_1 C_v}{n}.$$

Using supplementary information on two auxiliary variables : One at the stage of selection of the sample and the other at the estimation stage and then considering the best linear combination of the probability proportional to size (PPS) estimator and the ratio estimator, Agarwal and Kumar [1] reported the following estimator

$$\hat{\bar{Y}}_{AK} = K\bar{u}_n + (1 - K)(\bar{u}_n/\bar{v}_n)\bar{X} \tag{1.1}$$

for estimating population mean $\bar{Y}$ of the study character y, where K is a suitably chosen constant.

The minimum MSE of $\hat{\bar{Y}}_{AK}$ to the first degree of approximation, obtained by Agarwal and Kumar [1] is given by

$$\text{Min. MSE }(\hat{\bar{Y}}_{AK}) = \frac{\bar{Y}^2}{n} C_u^2 (1 - \rho_{uv}^2) \tag{1.2}$$

Following Srivastava and Jhajj [3], Jhajj and Srivastava [2] formulated a class of estimators for $\bar{Y}$ when sampling is done by the method of probability proportional to a suitable size variable which is different from the auxiliary variable used at the estimation stage, as

$$\hat{\bar{Y}}_{HS} = \bar{u}_n\, t(W, Z), \tag{1.3}$$

where $t(W, Z)$ is a function of W and Z such that

$$t(1, 1) = 1 \tag{1.4}$$

and also satisfies certain regularity conditions.

The minimum MSE of $\hat{\bar{Y}}_{HS}$ obtained by Jhajj and Srivastava [2], to the first degree of approximation, is given by

$$\text{Min. MSE }(\hat{\bar{Y}}_{HS}) = \frac{\bar{Y}^2}{n} C_u^2 \left[1 - \rho_{uv}^2 - \frac{(\alpha_1 \rho_{uv} - \psi)^2}{(\beta_2 - \beta_1 - 1)}\right] \tag{1.5}$$

In this paper we have defined a class of estimators of the population mean $\bar{Y}$ when the sampling is done by the method of probability proportional to suitable size variable which is different from the auxiliary variable used at the estimation stage. Asymptotic expressions for bias and MSE of the proposed class of estimators have been obtained. It has been shown that the minimum MSE of the suggested class of estimators is less than those of $\hat{\bar{Y}}_{AK}$ and $\hat{\bar{Y}}_{HS}$ considered by Agarwal and Kumar [1] and Jhajj and Srivastava (2) respectively.

2. The Class of Estimators

Consider the PPS sampling scheme with replacement based on P_i. Then the proposed class of estimators of $\bar{Y}$ is

$$\hat{\bar{Y}}_{HPS} = h\,(\bar{u}_n, W, Z) \tag{2.1}$$

where $h\,(\bar{u}_n, W, Z)$ is a function of $\bar{u}_n$, W and Z's such that

$$h\,(\bar{Y}, 1, 1) = \bar{Y}\, h_0\,(\bar{Y}, 1, 1), \tag{2.2}$$

$h_0\,(\bar{Y}, 1, 1)$ denotes the first order partial derivative with respect to $\bar{u}_n$ about the point $(\bar{Y}, 1, 1)$.

We assume that the function $h\,(\bar{u}_n, W, Z)$ is continuous and has continuous first and second order partial derivatives which are bounded in a closed convex subset, S, of the three dimensional real space containing the point $(\bar{Y}, 1, 1)$.

Expanding the function $h\,(\bar{u}_n, W, Z)$ about the point $(\bar{Y}, 1, 1)$ in a second order Taylor's series, we have

$$\begin{aligned} \hat{\bar{Y}}_{HPS} = {} & h\,(\bar{Y}, 1, 1) + (\bar{u}_n - \bar{Y})\, h_0\,(\bar{Y}, 1, 1) + (W - 1)\, h_1\,(\bar{Y}, 1, 1) \\ & + (Z - 1)\, h_2\,(\bar{Y}, 1, 1) \\ & + \frac{1}{2}\Big[(\bar{u}_n - \bar{Y})^2\, h_{00}\,(\bar{u}_n^*, W^*, Z^*) + 2\,(\bar{u}_n - \bar{Y})\,(W - 1) \\ & \quad h_{01}\,(\bar{u}_n^*, W^*, Z^*) \\ & + (W-1)^2\, h_{11}\,(\bar{u}_n^*, W^*, Z^*) + 2\,(W - 1)\,(Z -)\, h_{12}\,(\bar{u}_n^*, W^*, Z^*) \\ & + 2\,(\bar{u}_n - \bar{Y})\,(Z - 1)\, h_{02}\,(\bar{u}_n^*, W^*, Z^*) + (Z - 1)^2\, h_{22} \\ & \quad (\bar{u}_n^*, W^*, Z^*)\Big], \end{aligned} \tag{2.3}$$

where $\bar{u}_n^* = 1 + \lambda\,(\bar{u}_n - \bar{Y})$, $W^* = 1 + \mu\,(W-1)$, $Z^* = 1 + v\,(Z - 1)$; $0 < \lambda < 1$, $0 < \mu < 1$, $0 < v < 1$ and $h_0\,(\bar{Y}, 1, 1)$, $h_1\,(\bar{Y}, 1, 1)$, $h_2\,(\bar{Y}, 1, 1)$ denote the first order partial derivatives of the function $h\,(\bar{u}_n, W, Z)$ at the point $(\bar{u}_n, W, Z) = (\bar{Y}, 1, 1)$ and $h_{00}\,(\bar{u}_n^*, W^*, Z^*)$, $h_{01}\,(\bar{u}_n^*, W^*, Z^*)$, $h_{02}\,(\bar{u}_n^*, W^*, Z^*)$, $h_{12}\,(\bar{u}_n^*, W^*, Z^*)$, $h_{11}\,(\bar{u}_n^*, W^*, Z^*)$ and $h_{22}\,(\bar{u}_n^*, W^*, Z^*)$ denote its second order partial derivatives at the point $(\bar{u}_n^*, W^*, Z^*)$.

Substituting for $\bar{u}_n$, W and Z in terms of ε, δ and η in (2.3) and using (2.2) we have

$$\hat{\bar{Y}}_{HPS} = (\bar{Y} + \varepsilon)\, h_0 (\bar{Y}, 1, 1) + \delta\, h_1 (\bar{Y}, 1, 1) + \eta\, h_2 (\bar{Y}, 1, 1)$$
$$+ \frac{1}{2}\Big[\bar{Y}^2\, \varepsilon^2\, h_{00}\, (\bar{u}_n^*, W^*, Z^*) + 2\, \bar{Y}\, \varepsilon\delta\, h_{01}\, (\bar{u}_n^*, W^*, Z^*)$$
$$+ \delta^2\, h_{11}\, (\bar{u}_n^*, W^*, Z^*) + 2\, \eta\delta\, h_{12}\, (\bar{u}_n^*, W^*, Z^*)$$
$$+ 2\, \bar{Y}\, \varepsilon\eta\, h_{02}\, (\bar{u}_n^*, W^*, Z^*)$$
$$+ \eta^2\, h_{22}\, (\bar{u}_n^*, W^*, Z^*)\Big] \qquad (2.4)$$

Taking expectation of both sides of (2.4), it is easily found that

$$(\hat{\bar{Y}}_{HPS}) = \bar{Y}\, h_0(\bar{Y}, 1, 1) + 0\, (n^{-1}). \qquad (2.5)$$

The mean squared error upto terms of order n^{-1} of $\hat{\bar{Y}}_{HPS}$ is given by

$$\text{MSE}\, (\hat{\bar{Y}}_{HPS}) = E\, (\hat{\bar{Y}}_{HPS} - \bar{Y})^2$$
$$= E\, [\, (\bar{Y}^2 + \varepsilon^2 + 2\bar{Y}\varepsilon)\, h_0^2\, (\bar{Y}, 1, 1) + \delta^2\, h_1^2\, (Y, 1, 1)$$
$$+ \eta^2\, h_2^2\, (\bar{Y}, 1, 1) + \bar{Y}^2 + 2\, (\bar{Y} + \varepsilon)\, \delta\, h_0\, (\bar{Y}, 1, 1)\, h_1\, (\bar{Y}, 1, 1)$$
$$+ 2\, \eta\delta\, h_1\, (\bar{Y}, 1, 1)\, h_2\, (Y, 1, 1) + 2\, (\bar{Y} + \varepsilon)\, \eta$$
$$\times\, h_0\, (\bar{Y}, 1, 1)\, h_2\, (\bar{Y}, 1, 1) - 2\, \bar{Y}\{\, (\bar{Y} + \epsilon)\, h_0\, (\bar{Y}, 1, 1)$$
$$+ \delta\, h_1\, (\bar{Y}, 1, 1) + \eta\, h_2\, (\bar{Y}, 1, 1)\}], \qquad (2.6)$$
$$= [\bar{Y}^2\, \{(1 + n^{-1}\, C_u^2)\, h_0^2\, (\bar{Y}, 1, 1) - 2\, h_0\, (\bar{Y}, 1, 1) + 1\}$$
$$+ \frac{1}{n}\Big\{\, C_v^2\, h_1^2\, (\bar{Y}, 1, 1) + (\beta_2 - 1)\, h_2^2\, (\bar{Y}, 1, 1)$$
$$+ 2\, \bar{Y}\, \rho_{uv}\, C_u\, C_v\, h_0\, (\bar{Y}, 1, 1)\, h_1\, (\bar{Y}, 1, 1) + 2\, \alpha_1\, C^v$$
$$\times\, h_1\, (\bar{Y}, 1, 1)\, h_2\, (\bar{Y}, 1, 1) + 2\, \bar{Y}\, \psi\, C_u\, h_0\, (\bar{Y}, 1, 1)\, h_2\, (\bar{Y}, 1, 1)\Big\}]$$

which is minimized for (2.7)

$$h_0\, (\bar{Y}, 1, 1) = \frac{D_1}{D}, \qquad (2.8)$$

$$h_1\,(\bar{Y}, 1, 1) = \frac{R\bar{X}\,C_u\,D_2}{DC_v}, \tag{2.9}$$

$$h_2\,(\bar{Y}, 1, 1) = \frac{R\bar{X}\,C_u\,D_3}{D}, \tag{2.10}$$

where $R = \frac{\bar{Y}}{\bar{X}}$, $D_1 = (\beta_2 - \beta_1 - 1)$, $D_2 = [\,\psi\alpha_1 - (\beta_2 - 1)\,\rho_{uv}]$,

$D_3 = (\rho_{uv}\,\alpha_1 - \Psi)$

and $D = [\beta_2 - \beta_1 - 1 + n^{-1}\,C_u^2\,\{(\beta_2 - \beta_1 - 1)\,(1 - \rho_{uv}^2) - (\alpha_1\,\rho_{uv} - \psi)^2\}]$

Substituting from (2.8), (2.9) and (2.10) in (2.7), the minimum mean squared error of $\hat{\hat{Y}}_{HPS}$, up to terms of order n^{-1}, is given by

Min $\cdot$ MSE $(\hat{\hat{Y}}_{HPS})$

$$= \frac{\bar{Y}^2}{n} \cdot \frac{C_u^2\,[(1 - \rho_{uv}^2)\,(\beta_2 - \beta_1 - 1) - (\rho_{uv}\alpha_1 - \psi)^2]}{[\beta_2 - \beta_1 - 1 + n^{-1}\,C_u^2\,\{(1 - \rho_{uv}^2)\,(\beta_2 - \beta_1 - 1) - (\rho_{uv}\,\alpha_1 - \psi)^2\}]} \tag{2.11}$$

The class of estimators defined in (2.1) is very large. Any parametric function $h\,(\bar{u}_n, W, Z)$ satisfying (2.2) can generate an estimator of the class. If the parameters in $h\,(\bar{u}_n, W, Z)$ are so chosen that they satisfy (2.8), (2.9) and (2.10), then the resulting estimator will have the asymptotic mean squared error given by (2.11).

3. Theoretical Comparisons

It is well known that the variance of $\bar{u}_n$ is

$$\text{Var}\,(\bar{u}_n) = \bar{Y}^2\,\frac{C_u^2}{n} \tag{3.1}$$

From (1.2) and (3.1) we have

$$\text{Var}\,(\bar{u}_n) - \text{Min}\cdot\text{MSE}\,(\hat{Y}_{AK}) = \frac{\bar{Y}^2}{n}\,C_u^2\,\rho_{uv}^2 \tag{3.2}$$
$$> 0$$

From (1.2) and (1.5) we have

$$\text{Min} \cdot \text{MSE} (\hat{\hat{Y}}_{AK}) - \text{Min} \cdot \text{MSE} (\hat{\hat{Y}}_{HS}) = \frac{\bar{Y}^2}{n} \cdot \frac{C_u^2 (\alpha_1 \rho_{uv} - \psi)^2}{(\beta_2 - \beta_1 - 1)}, \tag{3.3}$$

$$> 0, \quad \text{since } \beta_2 - \beta_1 - 1 > 0$$

and from (1.5) and (2.11) we have

$$\text{Min} \cdot \text{MSE} (\hat{Y}_{HS}) - \text{Min} \cdot \text{MSE} (\hat{\hat{Y}}_{HPS})$$

$$= \frac{\bar{Y}^2}{n} \cdot \frac{[(\beta_2 - \beta_1 - 1)(1 - \rho_{uv}^2) C_u^2 - (\alpha_1 \rho_{uv} - \psi)^2]^2}{(\beta_2 - \beta_1 - 1)^2 [1 + n^{-1} C_u^2 \{1 - \rho_{uv}^2 - (\alpha_1 \rho_{uv} - \psi)^2 (\beta_2 - \beta_1 - 1)^{-1}\}}$$

$$> 0 \tag{3.4}$$

Thus we have the following inequality :

$$\text{Min} . \text{MSE} (\hat{\hat{Y}}_{HPS}) \leq \text{Min} \cdot \text{MSE} (\hat{\hat{Y}}_{HS}) \leqslant \text{Min} \cdot \text{MSE} (\hat{\hat{Y}}_{AK}) \leqslant \text{Var} (\bar{u}_n) \tag{3.5}$$

It follows from (3.5) that the proposed class of estimators $\hat{\hat{Y}}_{HPS}$ is more efficient than those of $\hat{\hat{Y}}_{AK}$ and $\hat{\hat{Y}}_{HS}$ envisaged by Agarwal and Kumar [1] and Jhajj and Srivastava [2], and the usual estimator $\bar{u}_n$.

It is to be noted that the minimum MSE of $\hat{\hat{Y}}_{HS}$ suggested by Jhajj and Srivastava [2] would be less than that of the usual regression estimator $\hat{Y}_{lr} = \bar{u}_n + b (\bar{X} - \bar{y}_n)$, b being the sample regression coefficient of y on x and the estimator $\hat{\hat{Y}}_{AK}$ considered by Agarwal and Kumar [1] if and only if $\rho_{uv} \; C_u \; \alpha_1 \neq \psi$, where as the minimum MSE of the proposed class of estimators $\hat{\hat{Y}}_{HPS}$ is less than that of $\hat{Y}_{lr}$ and $\hat{\hat{Y}}_{AK}$ even when $\rho_{uv} \; C_u \; \alpha_1 = \psi$.

It can also easily be seen that the estimators of the form :

$$t_1 = \bar{u}_n + \alpha_1^* (W - 1) + \alpha_2^* (Z - 1) \tag{3.6}$$

and

$$t_2 = \alpha_0^* \bar{u}_n + \alpha_1^* (W - 1) + \alpha_2^* (Z - 1) \tag{3.7}$$

are the members of the proposed class $\hat{\bar{Y}}_{HPS}$ in (2.1) but not of the class $\hat{\bar{Y}}_{HS}$ in (1.3) reported by Jhajj and Srivastava [2], where α_0^*, α_1^* and α_2^* are suitably chosen constants.

The conclusion is that the proposed class of estimators $\hat{\bar{Y}}_{HPS}$ is more efficient and wider than the classes of estimators $\hat{\bar{Y}}_{AK}$ in (1.1) and $\hat{\bar{Y}}_{HS}$ in (1.3) forwarded by Agarwal and Kumar [1] and Jhajj and Srivastava [2] respectively.

4. The Bias

To obtain the bias of $\hat{\bar{Y}}_{HPS}$, we assume that the third partial derivatives of $h(\bar{u}_n, W, Z)$ also exist and are continuous and bounded in a closed convex subset, S, of three dimensional real space containing the point $(\bar{Y}, 1, 1)$. Then expanding $h(\bar{u}_n, W, Z)$ around, $(\bar{Y}, 1, 1)$ to third order Taylor's series and taking expectations, we obtain the bias of $\hat{\bar{Y}}_{HPS}$ up to terms of order n^{-1}

$$\begin{aligned}\text{Bias } (\hat{\bar{Y}}_{HPS}) = \frac{\bar{Y}}{2n} [2n \{h_0(\bar{Y}, 1, 1) - 1\} &+ R\bar{X}\, C_u^2\, h_{00}\, (\bar{Y}, 1, 1) \\ &+ 2\rho_{uv} C_u C_v\, h_{01} (\bar{Y}, 1, 1) + 2\psi\, C_u\, h_{02}\, (\bar{Y}, 1, 1) \\ &+ (R\bar{X})^{-1} \{C_v^2 h_{11} (\bar{Y}, 1, 1) + 2\,\alpha_1 C_v\, h_{12}\, (\bar{Y}, 1, 1) \\ &+ (\beta_2 - 1)\, h_{22} (\bar{Y}, 1, 1)\}] \qquad (4.1)\end{aligned}$$

where $h_{00}(\bar{Y}, 1, 1)$, $h_{01}(\bar{Y}, 1, 1)$, $h_{02}(\bar{Y}, 1, 1)$, $h_{11}(\bar{Y}, 1, 1)$, $h_{12}(\bar{Y}, 1, 1)$ and $h_{22}(\bar{Y}, 1, 1)$ denote second partial derivatives of the function $h(\bar{u}_n, W, Z)$ at the point $(\bar{Y}, 1, 1)$. Thus it is seen that the bias of the estimator $\hat{\bar{Y}}_{HPS}$ also depends upon the second partial derivatives of the function $h(\bar{u}_n, W, Z)$ at the point $(\bar{Y}, 1, 1)$ and will be different for different asymptotic bias of any estimator of the class (2.1).

ACKNOWLEDGEMENTS

Author is thankful to the referee for his valuable suggestions/comments regarding the improvement of the paper. Author is also grateful to Professor Lal Chand, Head, Department of Mathematics and Statistics, J. N. K. V. V., Jabalpur for his kind perusal of revised version of the paper.

REFERENCES

[1] Agarwal, S. K. and Kumar, P. (1980): Combination of ratio and PPS estimators, *Jour. Ind. Soc. Agri. Statist.*, 32 : 81-86.

[2] Jhajj, H. S. and Srivastava, S. K. (1983) : A class of PPS estimators of population mean using auxiliary information, *Jour. Ind. Soc. Agri. Statist.*, 35(1) : 57-61.

[3] Srivastava, S. K. and Jhajj, H. S. (1981) : A class of estimators of the population mean in survey sampling using auxiliary information, *Biometrika*, 68 : 341-43.