Aligarh Journal of Statistics
Vol. 27 (2007), 11-24

SOME ESTIMATORS BETTER THAN REGRESSION ESTIMATOR USING SHRINKAGE INTERVAL TECHNIQUE

Vyas Dubey and T.P. Tripathi

ABSTRACT

In a case of bivariate finite populations where the mean $\overline{X}$ of an auxiliary characteristics x is known, it is customary to define ratio, regression, product and difference estimators for estimating mean $\overline{Y}$ of a principal variable y. It is well known that for large samples the mean squared error of regression estimator is smaller than those of other estimators mentioned above. In this paper, we make a search for some estimators whose MSE may be smaller than that of regression estimator. For estimating $\overline{Y}$ we have considered several estimators of the form $d=(1-w)\,\overline{y}_{rg}+wt$, where $\overline{y}_{rg}$ is well known regression estimator, w is a suitably chosen weight and t is a function of y and x values in the sample. We have obtained optimum choices of weights w and corresponding minimum mean squared errors. The results are illustrated for bivariate normal populations. The relative efficiencies of the proposed estimators compared to that of regression estimator have been obtained for a natural population data.

1. INTRODUCTION

It is well known in sample survey literature that auxiliary information, if used intelligibly, increases precision of estimator of a parameter. In case of a finite population $U=\{1,2,\cdots,N\}$, if the population mean $\overline{X}$ of an auxiliary variable x is known, it is customary to use ratio, product, difference or regression estimators defined by

$$\hat{\overline{Y}}_r=\hat{\overline{Y}}\,(\overline{X}/\hat{\overline{X}}),$$

$$\hat{\overline{Y}}_p=\hat{\overline{Y}}\,(\hat{\overline{X}}/\overline{X}),$$

$$\hat{\overline{Y}}_D=\hat{\overline{Y}}+K\,(\overline{X}-\hat{\overline{X}}) \tag{1.1}$$

and $\hat{\overline{Y}}_{rg}=\hat{\overline{Y}}+\hat{\beta}\,(\overline{X}-\hat{\overline{X}})$

for estimating the population mean $\overline{Y}$ of the principal variable y under study, where $\hat{\overline{Y}}$ and $\hat{\overline{X}}$ are unbiased estimators of $\overline{Y}$ and $\overline{X}$ respectively based on any sampling design and $\hat{\beta}$ is an unbiased estimate of $\beta = \text{Cov}(\hat{\hat{Y}}, \hat{\hat{X}}) / V(\hat{\hat{X}})$. It is found that in class of estimators

$$e_1 = \hat{\hat{Y}} (\overline{X} / \hat{\hat{X}})^{\alpha}$$

and $e_2 = \hat{\hat{Y}} h(u)$, $\quad u = \hat{\hat{X}} / \overline{X}$

considered by Srivastava (1967, 1971) and the class of estimators

$$e_3 = \hat{\hat{Y}} + t_0 (\overline{X} - \hat{\hat{X}})$$

considered by Tripathi (1970, 1980) , none of the estimators has MSE smaller than

$$M(\hat{\hat{Y}}_{rg}) = (1 - \rho^2) V(\hat{\hat{Y}}) \tag{1.2}$$

which is MSE of regression estimator $\hat{\hat{Y}}_{rg}$, ρ is correlation coefficient between $\hat{\hat{Y}}$ and $\hat{\hat{X}}$. It is also found that optimum estimators in the class

$$e_4 = \alpha \hat{\hat{Y}} + (1 - \alpha) \hat{\hat{Y}}_r$$

$$e_5 = \alpha \hat{\hat{Y}} + (1 - \alpha) \hat{\hat{Y}}_p$$

defined by Ray *et al*. (1979), Vos (1980) and by Chaubey *et al*. (1984) have the MSE same as in (1.2). It is further interesting to note that all the estimators in the class

$$e_6 = \frac{\hat{\hat{Y}} + t_1 (\overline{X} - \hat{\hat{X}})}{[t_2 \overline{X} + (1 - t_2) \hat{\hat{X}}]^{\alpha}} \overline{X}^{\alpha}$$

discussed by Das and Tripathi (1979,1980) are also unable to have MSE smaller than that of regression type estimator $\hat{\hat{Y}}_{rg}$ defined in (1.1).

In an effort to improve Searls (1964) type estimator

$$e_7 = \lambda \hat{\hat{Y}}$$

Das and Tripathi (1980) and Das (1988) considered the estimators

$$e_8 = W_1 \hat{Y} + W_2 (\overline{X} - \hat{X})$$

and $e_9 = W \hat{\hat{Y}}_{rg}$,

assuming the knowledge of $\overline{X}$ and C_y, the coefficient of variation of y. They found that the optimum values of weights W_1, W_2 and W are given by

$$W_{01} = 1/[1+(1-\rho^2)\, C^2(\hat{\overline{Y}})]\,,$$

$$W_{02} = \beta W_{01}$$

and $\quad W_0 = W_{01}$

respectively. The minimum MSE of above estimators e_8 and e_9 is equal, which is given by

$$M_0(e_8) = M\,(\hat{\overline{Y}}_{rg})/[1+(1-\rho^2)\, C^2(\hat{\overline{Y}})]\,. \tag{1.3}$$

We note that for above optimum, exactly or approximately, the customary regression estimator $\hat{\overline{Y}}_{rg}$ will be improved. In this paper, we consider several estimators as an improvement over the regression estimator by choosing weights suitably and by shrinking the interval of preference. This interval contains values of the weights for which suggested estimator is more precise than $\hat{\overline{Y}}_{rg}$. First we give some general results and then we consider several special estimators.

2. SOME GENERAL RESULTS

Let $\hat{\theta}$ be an estimator of a parameter θ of interest based on any sampling design and t be a suitably chosen statistic which provides some information about θ directly or indirectly, then a weighted estimator of θ may be proposed as

$$d = (1-\omega)\,\hat{\theta} + \omega t\,, \tag{2.1}$$

where ω is a suitably chosen random weight (which in particular may be a constant) such that its mean $E(\omega)$ exists. The exact expressions of bias and MSE of d are given by

$$B\,(d) = (1-E\omega)B\,(\hat{\theta}) + (E\omega)\,(E\,t-\theta) - Cov\,(\hat{\theta},\omega) + Cov\,(t,\omega) \tag{2.2}$$

and

$$M\,(d) = [(1-E\omega)B\,(\hat{\theta}) + (E\omega)\,(E\,t-\theta)]^2 + (1-E\omega)^2\,V\,(\hat{\theta})$$
$$+\,(E\omega)^2\,V(t) + 2\,(E\omega)\,(1-E\omega)\,Cov\,(\hat{\theta},t) + Q_1 + Q_2\,, \tag{2.3}$$

where,

$$B\,(\hat{\theta}) = (E\hat{\theta}-\theta)\,,$$

$$Q_1 = [B(\hat{\theta}) - (Et - \theta)]^2 V(\omega) - 2\,[2(1-E\omega)B(\hat{\theta})$$
$$- (1-2E\omega)(Et-\theta)]Cov(\hat{\theta},\omega)$$
$$-2[(1-2E\omega)B(\hat{\theta}) + 2E\omega(Et-\theta)]\;Cov(t,\omega),$$
$$Q_2 = E[\delta_\omega^2(\delta_{\hat{\theta}} - \delta_t)^2] - 2(1-E\omega)E(\delta_{\hat{\theta}}^2\,\delta_\omega)$$
$$+2(1-2E\omega)E(\delta_{\hat{\theta}}\,\delta_\omega\,\delta_t) + 2(E\omega)E(\delta_t^2\,\delta_\omega)$$
$$+2[B(\hat{\theta}) - (Et-\theta)]\;E[\delta_\omega^2(\delta_{\hat{\theta}} - \delta_t)],$$
$$\delta_{\hat{\theta}} = \hat{\theta} - E\hat{\theta}, \quad \delta_t = t - Et.$$

In general, it is difficult to find optimum value of random weight ω which minimizes the MSE as it is confounded with $M(d)$ through the terms such as $V(\omega)$, $Cov(\hat{\theta},\omega)$ and $Cov(t,\omega)$. Consequently, we confine our discussion mainly to the case when $\omega = W$ (a constant). In this situation $Q_1 = 0$ and $Q_2 = 0$ and expressions (2.2) and (2.3) are considerably simplified and reduces to

$$B(d) = (1-W)\,B(\hat{\theta}) + W(Et-\theta) \tag{2.4}$$

and

$$M(d) = [(1-W)\,B(\hat{\theta}) + W(Et-\theta)]^2 + (1-W)^2 V(\hat{\theta})$$
$$+W^2 V(t) + 2W(1-W)\,Cov(\hat{\theta},t). \tag{2.5}$$

It may be shown that the optimum weight W which minimizes $M(d)$ in (2.5) is given by

$$W_0 = [M(\hat{\theta}) - Cov(\hat{\theta},t) - B(\hat{\theta})(Et-\theta)]/D, \tag{2.6}$$

where

$$D = M(\hat{\theta}) + V(t) - 2\,Cov(\hat{\theta},t) - 2B(\hat{\theta})(Et-\theta) + (Et-\theta)^2 \tag{2.7}$$
$$= E\{\hat{\theta} - t\}^2 > 0.$$

The resulting minimum MSE is given by

$$M_0(d) = V(\hat{\theta}) + [B(\hat{\theta})]^2 - W_0^2 D. \tag{2.8}$$

Again

$$M(d) = M(\hat{\theta}) + (W^2 - 2WW_0)D. \tag{2.9}$$

Therefore estimator d will be more efficient than $\hat{\theta}$ if

either $0 < W < 2W$ (2.10)

or $2W_0 < W < 0$.

It is noted that optimum weight as well as the interval of preference $(0, 2W_0)$ over $\hat{\theta}$ may depend upon unknown population values. However, in practice, some good guessed values W_0^* of W_0 may be found based on Census data or previous surveys or a pilot survey. If W_0^* is a quantity such that $W_0^* < W_0$ if W_0 is positive and $W_0^* > W_0$ if W_0 is negative, then the shrunken interval of preference would be given by choosing W between zero and $2W_0^*$.

Further in practice, for a given t, one may use $\hat{W}_0$ for ω in (2.1) where $\hat{W}_0$ is such that

$$E\hat{W}_0 = W_0 + \text{terms of order } O(n^{-r}), \qquad r > 0. \tag{2.11}$$

Such a $\hat{W}_0$ may customarily be obtained by replacing various unknown parameters in (2.6) through their estimated values. From (2.3), (2.5) and (2.6), it is noted that such a choice $\hat{W}_0$ would be near optimum for the situations in which $Q_2 = 0$.

For obtaining the estimators better than regression estimator, we consider the class of estimators

$$d = (1-W)\,\bar{y}_{rg} + Wt \tag{2.12}$$

for estimating $\bar{Y}$ in the next section and confine our discussion to simple random sampling without replacement (*SRSWOR*) in which case $\hat{\bar{Y}} = \bar{y}$ and $\hat{\bar{X}} = \bar{x}$ are means of sample of size n, selected from a population of size N and

$$\bar{y}_{rg} = \bar{y} + b\,(\bar{X} - \bar{x}),$$

where

$$b = s_{yx} / s_x^2, \qquad s_{yx} = \sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})/(n-1), \; s_x^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2/(n-1).$$

Ignoring third and higher order terms and using the results given above, we obtain

$$B(d) = (1-W)\,B(\bar{y}_{rg}) + W\,(Et - \bar{Y}) \tag{2.13}$$

and

$$M(d) = (1-W)^2 M(\bar{y}_{rg}) + 2W(1-W)\{Cov(\bar{y},t) - \beta\, Cov(\bar{x},t)$$
$$+ (Et - \bar{Y})B(\bar{y}_{rg})\} + W^2\{V(t) + (E(t) - \bar{Y})^2\}, \qquad (2.14)$$

where

$$B(\bar{y}_{rg}) = -\left(\frac{1-f}{n}\right)\left(\frac{N}{N-2}\right)[\psi_{12}(y,x) - \rho_{y,x}\sqrt{\beta_1(x)}]\bar{Y}\, C_y\,,$$

$$M(\bar{y}_{rg}) = \left(\frac{1-f}{n}\right)(1-\rho_{yx}^2)\, S_y^2\,,$$

$$\psi_{ij}(y,x) = \frac{\mu_{ij}(y,x)}{\sqrt{\mu_{20}^i(y,x)\,\mu_{02}^j(y,x)}}\,, \quad (i,j) = 1,2\,.$$

$$\mu_{rs}(y,x) = E(y-\bar{Y})^r (x-\bar{X})^s\,,$$

$$\beta_1(x) = \frac{\mu_{03}^2(y,x)}{\mu_{02}^3(y,x)}\,, \quad S_y^2 = \frac{N\mu_{20}(y,x)}{(N-1)}\,, \quad C_y = \frac{S_y}{\bar{Y}}\,,$$

ρ_{yx} is correlation coefficient between y and x, $f = n/N$. In this case expressions (2.6), (2.7) and (2.8) reduce to

$$W_0(t) = [M(\bar{y}_{rg}) - Cov(\bar{y},t) + \beta\, Cov(\bar{x},t) - (Et - \bar{Y})B(\bar{y}_{rg})]/D(t)$$

$$D(t) = M(\bar{y}_{rg}) + V(t) - 2\{Cov(\bar{y},t) - \beta\, Cov(\bar{x},t)$$
$$+ (Et - \bar{Y})B(\bar{y}_{rg})\} + (Et - \bar{Y})^2$$

and

$$M_0(d) = M(\bar{y}_{rg}) - W_0^2(t)D(t)\,. \qquad (2.15)$$

It is clearly noted that even if the chosen statistic t is uncorrelated with $\bar{y}$ and $\bar{x}$, one would obtain $W_0(t) \neq 0$ and thus $\bar{y}_{rg}$ may be improved through d in (2.12) by choosing W between 0 and $2W_0(t)$ for that choice of t. However t has to be chosen such that $W_0^2(t)D(t)$ is large so that corresponding reduction in MSE is appreciable. From (2.15) we observe that the choice $t = t_1$ in (2.12) would be preferable over $t = t_2$ if

$$[W_0(t_1)/W_0(t_2)]^2 > \{D(t_2)/D(t_1)\}\,. \qquad (2.16)$$

3. PROPOSED ESTIMATORS

In this section we consider various choices of t in the class of estimators defined in (2.12) for $\overline{Y}$. The resulting estimators are

$$d_1 = (1-W_1)\,\bar{y}_{rg} + W_1\,\bar{y}\,(S_x^2 / s_x^2) \tag{3.1}$$

$$d_2 = (1-W_2)\,\bar{y}_{rg} + W_2\,\bar{y}\,(s_x^2 / S_x^2) \tag{3.2}$$

$$d_3 = (1-W_3)\,\bar{y}_{rg} + W_3\,(s_y^2 / \bar{y}) \tag{3.3}$$

$$d_4 = (1-W_4)\,\bar{y}_{rg} + W_4\,(s_y / C_y) \tag{3.4}$$

$$d_5 = (1-W_5)\,\bar{y}_{rg} + W_5\,(s_y / s_x)\,\overline{X}\,, \tag{3.5}$$

where W_i, $i = 1, 2, \cdots, 5$ are suitably chosen constants and $S_x^2 = N\,\mu_{02}(y,x)/(N-1)$. The estimators d_1, d_2 are motivated by the situations where population mean and variance of auxiliary variable x have proportionate relationship, e.g. population mean and variance of Poisson distribution are equal. In the situations where mean and standard deviation have proportionate relationship, the estimators d_3, d_5 may be found to be suitable. The estimator d_4 has been proposed for the situations where C_y is known.

4. BASIC RESULTS ABOUT PROPOSED ESTIMATORS

We assume that sample size n is large and only principal terms, i.e. the terms upto $O(n^{-1})$, are considered. We use the symbols

$$A = \frac{n^2}{(n-1)^2} - \frac{2\,n\,(N-2n)}{(n-1)^2 (N-2)} + \frac{N^2 + N - 6\,Nn + n^2}{(n-1)^2 (N-2)\,(N-3)}\,,$$

$$B = \frac{n\,N^2}{(n-1)\,(N-1)\,(N-2)} - \frac{3\,N\,(N-n-1)}{(n-1)\,(N-2)\,(N-3)}\,, \quad F = \frac{N}{(N-2)}\left(\sqrt{\frac{(N-1)}{N}}\right),$$

$$\beta_1(y) = \frac{\mu_{30}^2(y,x)}{\mu_{20}^3(y,x)}\,, \quad \beta_2(y) = \frac{\mu_{40}(y,x)}{\mu_{02}^2(y,x)}\,, \; \beta_2(x) = \frac{\mu_{04}(y,x)}{\mu_{02}^2(y,x)}\,, \; C_x = \frac{S_x}{\overline{X}}\,,$$

$$K = \frac{C_y}{C_x}\,, \quad \beta_2^*(y) = A\,\beta_2(y) - B\,, \quad \beta_2^*(x) = A\,\beta_2(x) - B\,,$$

$$\psi_{22}^*(y,x) = A\,\psi_{22}(y,x) - B\,, \quad \delta_{12}(y,x) = \psi_{12}(y,x) - \rho_{yx}\sqrt{\beta_1(x)}\,,$$

$$\delta_{21}(y,x) = \rho_{yx}\,\psi_{21}(y,x) - \sqrt{\beta_1(y)}\,.$$

Under above assumptions, bias of proposed estimators are given by

$$B(d_1)=\frac{1-f}{n}\bar{Y}[W_1\{\beta_2^*(x)-F\,C_y\sqrt{\beta_1(x)}\}-F\,C_y\,\delta_{12}(y,x) \tag{4.1}$$

$$B(d_2)=\frac{1-f}{n}\bar{Y}\,F\,C_y\,[W_2\,\psi_{12}(y,x)-(1-W_2)\,\delta_{12}(y,x)] \tag{4.2}$$

$$B(d_3)=\bar{Y}\Big[W_3(C_y^2-1)-\frac{1-f}{n}\{(1-W_3)\,F\,C_y\delta_{12}(y,x)$$

$$-W_3C_y^2(C_y^2-F\,C_y\sqrt{\beta_1(y)}\}] \tag{4.3}$$

$$B(d_4)=-\frac{1-f}{8n}\bar{Y}\,C_y\,[8\,F\,(1-W_4)\,\delta_{12}(y,x)+W_4\,\beta_2^*(y)] \tag{4.4}$$

$$B(d_5)=(K-1)W_5\,\bar{Y}-\frac{1-f}{8n}\bar{Y}[8\,(1-W_5)\,F\,C_y\,\delta_{12}(y,x)$$

$$+W_5\,K\{\beta_2^*(y)-3\,\beta_2^*(x)+2\,\psi_{22}^*\,(y,x)\}] \tag{4.5}$$

MSE of the estimators are

$$M(d_1)=M(\bar{y}_{rg})+\frac{1-f}{n}\bar{Y}^2\,[W_1^2\{\rho_{yx}^2\,C_y^2+\beta_2^*(x)-2\,F\rho_{yx}\,C_y\sqrt{\beta_1(x)}\}$$

$$-2W_1\,F\,C_y\,\delta_{12}(y,x)] \tag{4.6}$$

$$M(d_2)=M(\bar{y}_{rg})+\frac{1-f}{n}\bar{Y}^2\,[W_2^2\{\rho_{yx}^2\,C_y^2+\beta_2^*(x)-2\,F\rho_{yx}\,C_y\sqrt{\beta_1(x)}\}$$

$$+2\,W_2\,F\,\delta_{12}(y,x)] \tag{4.7}$$

$$M(d_3)=M(\bar{y}_{rg})+\bar{Y}^2\Bigg[W_3^2\Bigg\{(C_y^2-1)^2+\frac{1-f}{n}\Bigg\{(1+2\,C_y^2)\,C_y^2(1-\rho_{yx}^2)$$

$$+3\,C_y^6+(\beta_2^*(y)-2)\,C_y^4+2\,F\,C_y\,(C_y^2\rho_{yx}\,\psi_{21}(y,x)+(C_y^2-1)\,\delta_{12}(y,x)$$

$$-2\,C_y^4\sqrt{\beta_1(y)})\Bigg\}\Bigg\}-2\frac{1-f}{n}W_3\,\Bigg\{(1+C_y^2)\,C_y^2\,(1-\rho_{yx}^2)$$

$$+F\,C_y\,(C_y^2\,\delta_{21}(y,x)+(C_y^2-1)\,\delta_{12}(y,x))\Bigg\}\Bigg] \tag{4.8}$$

$$M(d_4)=M(\bar{y}_{rg})+\frac{1-f}{n}\bar{Y}^2\,[W_4^2\,\{C_y^2\,(1-\rho_{yx}^2)+(\beta_2^*(y)/4)$$

$$+F\,C_y\delta_{21}(y,x)\}-2\,W_4\,\{C_y^2\,(1-\rho_{yx}^2)+(F\,C_y\delta_{21}(y,x)/2)\}] \tag{4.9}$$

$$M(d_5) = M(\bar{y}_{rg}) + W_5^2\,\bar{Y}^2\Bigg[(K-1)^2 + \frac{1-f}{n}\Big\{C_y^2\,(1-\rho_{yx}^2) + \frac{K}{4}(\beta_2^*(y)$$

$$+ (4K-3)\,\beta_2^*(x) - 2(2K-1)\,\psi_{22}^*(y,x)) + F\,C_y\Big\{K\,\delta_{21}(y,x)$$

$$+ (3K-2)\,\delta_{12}(y,x)\Big\} - 2W_5\,\frac{1-f}{n}\{C_y^2\,(1-\rho_{yx}^2)$$

$$+ (F\,C_y/2)(K\,\delta_{21}(y,x) + (3K-2)\,\delta_{12}(y,x)\Big\}\Bigg]. \tag{4.10}$$

The best values of W_i for which MSE of the proposed estimators d_i, $i = 1, 2, \cdots, 5$ will be minimum are respectively given by

$$W_{01} = F\,C_y\,\delta_{12}(y,x)/D_1 \tag{4.11}$$

$$W_{02} = -F\,C_y\,\delta_{12}(y,x)/D_2 \tag{4.12}$$

$$W_{03} = [(1+C_y^2)\,C_y^2(1-\rho_{yx}^2) + F\,C_y^3\,\delta_{21}(y,x)$$

$$+ F\,C_y\,(C_y^2 - 1)\,\delta_{12}(y,x)]/D_3 \tag{4.13}$$

$$W_{04} = \{C_y^2(1-\rho_{yx}^2) + (F\,C_y\,\delta_{21}(y,x)/2)\}/D_4 \tag{4.14}$$

$$W_{05} = [C_y^2(1-\rho_{yx}^2) + (F\,C_y/2)\{K\,\delta_{21}(y,x)$$

$$+ (3K-2)\,\delta_{12}(y,x)\}]/D_5, \tag{4.15}$$

where

$$D_1 = \rho_{yx}^2\,C_y^2 + \beta_2^*(x) - 2F\,\rho_{yx}\,C_y\sqrt{\beta_1(x)} \tag{4.16}$$

$$D_2 = \rho_{yx}^2\,C_y^2 + \beta_2^*(x) - 2F\,\rho_{yx}\,C_y\sqrt{\beta_1(x)} \tag{4.17}$$

$$D_3 = \frac{n\,(C_y^2 - 1)^2}{1-f} + [(1+2C_y^2)\,C_y^2(1-\rho_{yx}^2) + 3C_y^6$$

$$+ (\beta_2^*(y) - 2)\,C_y^4 + 2F\,C_y\,(C_y^2\,\rho_{yx}\,\psi_{12}(y,x)$$

$$+ (C_y^2 - 1)\,\delta_{12}(y,x) - 2C_y^4\sqrt{\beta_1(y)})] \tag{4.18}$$

$$D_4 = C_y^2(1-\rho_{yx}^2) + \{\beta_2^*(y)/4\} + F\,C_y\,\delta_{21}(y,x) \tag{4.19}$$

$$D_5 = \frac{n(K-1)^2}{1-f} + C_y^2(1-\rho_{yx}^2) + \frac{K}{4}\Big\{\beta_2^*(y) + (4K-3)\,\beta_2^*(x)$$

$$-\,2(2K-1)\,\psi_{22}^*(y,x)\Big\} + F\,C_y\{K\,\delta_{21}(y,x) + (3K-2)\,\delta_{12}(y,x)\} \tag{4.20}$$

Following result (2.12), the corresponding minimum MSE of estimator d_i is obtained as

$$M_0(d_i) = M(\bar{y}_{rg}) - \left(\frac{1-f}{n}\right)\bar{Y}^2\,W_{0i}^2\,D_i, \qquad i = 1,2,\cdots,5\,. \tag{4.21}$$

Using the result (2.10), we have

$$M\,(d_i) < M\,(\bar{y}_{rg}),\ \text{if} \quad 0 < W_i < 2W_{0i}\,,\ i = 1,2,\cdots,5\,. \tag{4.22}$$

5. THE CASE OF BIVARIATE NORMAL POPULATIONS

In general it is difficult to compare efficiency of the proposed estimators d_1 to d_5 among themselves as they have been proposed under different situations. To get an idea we consider the case of bivariate normal populations in which case $\delta_{12}(y,x) = 0 = \delta_{21}(y,x)$, $\beta_2(x) = 3 = \beta_2(y)$, $\psi_{22}(y,x) = 1 + 2\rho_{yx}^2$. For simplicity of presentation, henceforth we assume that the sampling fraction $f = (n/N)$ is ignored, leading the terms A and B to be replaced by unity. Thus we obtain minimum MSE of d_1 to d_5 as

$$M_0^*(d_1) = M\,(\bar{y}_{rg}) = M_0^*(d_2) \tag{5.1}$$

$$M_0^*(d_3) = M\,(\bar{y}_{rg})\left[1 - \frac{C_y^2(1-\rho_{yx}^2)\,(1+C_y^2)^2}{C_y^2\,(1-\rho_{yx}^2)(1+2C_y^2) + 3\,C_y^6 + n\,(C_y^2-1)^2}\right] \tag{5.2}$$

$$M_0^*(d_4) = M\,(\bar{y}_{rg})\left[1 - \frac{2C_y^2(1-\rho_{yx}^2)}{2C_y^2(1-\rho_{yx}^2)+1}\right] \tag{5.3}$$

$$M_0^*(d_5) = M\,(\bar{y}_{rg})\left[1 - \frac{C_y^2(1-\rho_{yx}^2)}{\{C_y^2 + 2K^2 - K)(1-\rho_{yx}^2) + n\,(K-1)^2}\right] \tag{5.4}$$

From above expressions of $MSE's$, it is clear that all the proposed estimators are more efficient than usual regression estimator. If $K=1$, $M_0^*(d_5)$ reduces to

$$M_0^*(d_5)=\frac{1}{n}S_y^2(1-\rho_{yx}^2)\left(1-\frac{C_y^2}{1+C_y^2}\right) \tag{5.5}$$

which will be smaller than $M_0^*(d_4)$ if $\rho_{yx}^2 \le 1/2$. If $K=1/2$,

$$M_0^*(d_5)=\frac{1}{n}S_y^2(1-\rho_{yx}^2)\left[1-\frac{C_y^2(1-\rho_{yx}^2)}{C_y^2(1-\rho_{yx}^2)+(n/4)}\right] \tag{5.6}$$

which will be greater than $M_0^*(d_4)$ for sufficiently large value of n. Further more, we find that

$$M_0^*(d_3)<M(\bar{y}_{rg}), \quad \text{if} \quad C_y \ne 0 \tag{5.7}$$

For $C_y = 1$, we have

$$M_0^*(d_4)<M_0^*(d_3). \tag{5.8}$$

6. SHRUNKEN RANGE OF INTERVALS

The value of W_{0i} may not be known in advance, therefore it may be difficult to find the interval $(0,\, 2W_{0i})$ in which W_{0i} may take the values so that corresponding estimator is efficient than regression estimator. In shrunken range technique, we find an interval $(0,\, 2W_{0i}^*)$ of smaller width than $(0,\, 2W_{0i})$, where W_{0i}^* is known. For simplicity we consider the case of bivariate normal populations with *fpc* ignored. Here values of W_{0i}, $i=1,2,\cdots,5$, are given by

$$W_{01}=0=W_{02} \tag{6.1}$$

$$W_{03}=\frac{C_y^2(1+C_y^2)(1-\rho_{yx}^2)}{C_y^2(1+2C_y^2)(1-\rho_{yx}^2)+3C_y^6+n(C_y^2-1)^2} \tag{6.2}$$

$$W_{04}=\frac{2C_y^2(1-\rho_{yx}^2)}{1+2C_y^2(1-\rho_{yx}^2)} \tag{6.3}$$

$$W_{05}=\frac{C_y^2(1-\rho_{yx}^2)}{(C_y^2+2K^2-K)(1-\rho_{yx}^2)+n(K-1)^2} \tag{6.4}$$

Let $C_y^{(1)}$ and $C_y^{(0)}$ be maximum and minimum values of C_y, $\rho_{yx}^{(1)}$ be maximum value of ρ_{yx}. We consider the value of W_{03} and W_{05} as

$$W_{03}^{*} = \frac{C_y^{(0)^2}(1 + C_y^{(0)})^2(1 - \rho_{yx}^{(1)^2})}{C_y^{(1)^2}(1 + 2C_y^{(1)})^2(1 - \rho_{yx}^{(1)^2}) + 3C_y^{(1)^6} + n(C_y^{(1)^2} - 1)^2} \tag{6.5}$$

$$W_{05}^{*} = \frac{C_y^{(0)^2}(1 - \rho_{yx}^{(1)^2})}{\{C_y^{(1)^2} + 2K^{(1)}(K^{(1)} - 1)\}(1 - \rho_{yx}^{(1)^2}) + n(K^{(1)} - 1)^2} \tag{6.6}$$

respectively, where $K^{(1)} = C_y^{(1)} / C_x$. If auxiliary variable x is considered as past data, where $C_y = C_x$, W_{05}^{*} reduces to

$$W_{05}^{*} = \frac{C_y^{(0)^2}}{1 + C_y^{(1)^2}} \tag{6.7}$$

The estimator d_4 is proposed under the knowledge of C_y. Therefore we consider W_{04}^{*} as

$$W_{04}^{*} = \frac{2C_y^2(1 - \rho_{yx}^{(1)^2})}{1 + 2C_y^2(1 - \rho_{yx}^{(1)^2})}. \tag{6.8}$$

7. NUMERICAL EXAMPLE

Let us consider the data which relates to the villages wise population of rural area under Police Station, Singur, Dist. Hooghly as obtained from District Census Handbook, 1981, published by Government of India. The population consists of 96 villages with

$y =$ the number of agricultural labourers in village and $x =$ the area the village.

$\bar{Y} = 137.9271$, $\bar{X} = 144.8278$, $C_y = 1.316262$, $C_x = 0.807529$,

$\beta_1(y) = 21.58889$, $\beta_1(x) = 5.438926$, $\beta_2(y) = 32.96591$,

$\beta_2(x) = 9.27468$, $\psi_{12}(y,x) = 2.81763$, $\psi_{21}(y,x) = 2.11855$,

$\psi_{22}(y,x) = 11.81294$, $\rho_{yx} = 0.77323$.

For this data relative efficiency (RE) of proposed estimators with respect to $\bar{y}$, defined as $\frac{V(\bar{y})}{M_0(d_i)} \times 100$, $i = 1, 2, \cdots, 5$, are given in Table 1:

Table 1:

Estimator	Relative Efficiency		
	$n=10$	$n=20$	$n=30$
d_1	504.19	516.48	516.86
d_2	301.75	302.25	302.28
d_3	298.72	286.74	283.06
d_4	466.77	465.81	465.76
d_5	295.84	255.09	251.61

The RE of $\bar{y}_{rg}$ *w.r.t.* $\bar{y}$ is 248.6%. Table 1 shows that the estimators d_1, d_2, d_3, d_4 and d_5, are more than efficient than $\bar{y}_{rg}$.

8. CONCLUSION

It is concluded that the estimators of $\bar{Y}$ generated through (2.12) with suitably chosen weights are more precise than $\bar{y}_{rg}$. The weights may be chosen using shrinkage interval technique, discussed in section 6. Similarly, parameters like variance and coefficient of variation could be estimated more efficiently using general class of estimators (2.1).

Acknowledgement

Authors are thankful to the referee for his suggestions leading to improvement in the paper.

REFERENCES

Chaubey, Y.P., Singh, M. and Dwivedi, T.D. (1984): A note on an optimality property of the regression estimator. *Biom. J.*, **26**, 465-467.

Das, A.K. (1988): *Contribution to the Theory of Sampling Strategies based on Auxiliary Information*. Ph.D. thesis submitted to B.C.K.V., Mohanpur, Nadia, West Bengal.

Das, A.K. and Tripathi, T.P. (1979): A class of estimators of population mean when the mean of an auxiliary character is known. Stat-Math. Tech. Report No. 22/79, I.S.I., Calcutta. Abs. No.3, *J. Indian Soc. Agricultural Statist*., **31**, 69.

Das, A.K. and Tripathi, T.P. (1980): Sampling strategies for population mean when the coefficient of variation of an auxiliary character is known. *Sankhyā, Ser C*, **42**, 76-86.

Ray, S.K., Sahai, A. and Sahai, A. (1979): A note on ratio and product estimators. *Ann. Inst. Statist. Math.*, **31**, 141-144.

Searls, D.T. (1964): The utilization of a known coefficient of variation in the estimation procedure. *J. Amer. Statist. Assoc.*, **59**, 1225-1226.

Srivastava, S.K. (1967): An estimator using auxiliary information in sample surveys. *Calcutta Statist. Assoc. Bull.*, **16**,121-132.

Srivastava, S.K. (1971): A generalized estimators for the population mean of a finite population using multi-auxiliary information. *J. Amer. Statist. Assoc.*, **66**, 404-407.

Tripathi, T.P. (1970): *Contributions to the Sampling Theory using Multivariate Information*. Ph. D. thesis submitted to Punjabi University, Patiala.

Tripathi, T.P. (1980): A general class of estimators of population ratio. *Sankhyā, Ser C*, **42**, 63-75.

Vos, J.W.E. (1980): Mixing of direct, ratio and product method estimators. *Statist. Neerlandica*, **34**, 209-218.

Received : 26-08-2004

Revised : 02-09-2006

Vyas Dubey
School of Studies in Statistics,
Pt. Ravishankar Shukla University, Raipur, India.
e-mail: dubey_vyas@rediffmail.com

T.P. Tripathi
Indian Statistical Institute, Kolkata.