STATISTICA, anno LXXII, n. 4, 2012

AN IMPROVEMENT OVER REGRESSION METHOD OF ESTIMATION

H.P. Singh, A. Rathour, R.S. Solanki

1. INTRODUCTION

It is well established in sample surveys that auxiliary information is often used to improve the precision of estimators of population parameters. It is well known that for estimating population mean $\bar{Y}$ of the study variable y using information on auxiliary variable x (say population mean $\bar{X}$ of x is known in advance) at the estimation stage, the ratio, product, difference and regression methods are extensively used. The properties of these methods have been dealt in detail in different books [e.g. (Murthy, 1967), (Cochran, 1977) and (Singh, 2003)]. The ratio method of estimation which has been developed by (Cochran, 1940) is used to estimate the population mean $\bar{Y}$ of the study variable y which is positively (high) correlated with an auxiliary variable x. For example

(i) y may be the production of wheat and x is the area cultivated.

(ii) y may denote the number of bullocks on a holding and x its area in acres.

(iii) y may be the total production and x is the number of workers.

However, when auxiliary variable x is negatively correlated (high) with the study variable y the product method of estimation envisaged by (Robson, 1957) and rediscovered by (Murthy, 1964) can be employed quite effectively. For example

(i) Education and years in jail people who have more years of education tend to have fewer years in jail (or phrased as people with more years in jail) crying and being held among babies, those who are held more tend to cry less (or phrased as babies who are held less tend to cry more) are other examples of negative correlations.

(ii) Price and demand of a commodity.

(iii) The volume and pressure of a gas.

(Hansen *et al.*, 1953) have suggested the difference method of estimation in the same condition when a study variable y is correlated with an auxiliary variable x.

The ratio, product and difference methods are considered be the most practicable in many situations, but they have the limitations of having at the most same efficiency as that of the usual linear regression method of estimation. In the situation where the relation between the study variable y and the auxiliary variable x is a straight line and passing through the origin the ratio and product estimators have efficiencies equals to the regression estimator and difference estimator perform equals to the regression estimator at their optimum condition. This encourage authors to develop a class of estimators (for estimating population mean of the study variable) which is more efficient than the usual regression estimator by using information on an auxiliary variable.

Consider a finite population $U=(U_1,U_2,...,U_N)$ of N identifiable units. Let y and x denote the study variable and auxiliary variable taking values y_i and x_i respectively on the i^{th} unit U_i of the population U and a sample of size $n(<N)$ is drawn by simple random sampling without replacement (SRSWOR) from the population U. Let $\bar{y}[=n^{-1}\sum_{i-1}^{n}y_i]$ and $\bar{x}[=n^{-1}\sum_{i-1}^{n}x_i]$ denote the sample mean of the study variable y and auxiliary variable x respectively. In this article we have proposed a new class of estimators for population mean $\bar{Y}$ of the study variable y by using information on an auxiliary variable x which is more efficient than the usual linear regression estimator and other existing estimators.

The remaining part of the paper is organized as follows: Sec. 2 give a brief review of some estimators for the population mean of the study variable. In Sec. 3, a new class of estimator is described and the expressions for its asymptotic bias and the mean square error are obtained. Sec. 4 discusses a particular case of proposed class of estimators with its properties under large sample approximation. Sec. 5 addresses the problem of efficiency comparisons, while in Sec. 6 an empirical study is carried out to evaluate the performance of different estimators using some natural population data sets. Sec. 7 concludes the paper with final remarks.

2. REVIEWING ESTIMATORS OF THE POPULATION MEAN

Much literature has been produced on a sampling form finite population to address the issue of the efficient estimation of the mean of the study variable when auxiliary variables are available for instance see (Diana 1993), (Upadhyaya and Singh, 1999), (Singh and Tailor, 2005), (Singh and Karpe, 2009), (Singh and Solanki, 2012). Our analysis refers to simple random sampling without replacement (SRSWOR) and considers the case when only a single auxiliary variable is used.

It is very well known that sample mean $\bar{y}$ is an unbiased estimator of population mean $\bar{Y}$ and under SRSWOR its variance is given by

$$Var(\bar{y})=\theta S_y^2=\theta\bar{Y}^2C_y^2, \tag{1}$$

where $\theta = n^{-1}(1-f)$, $f = (n/N)$ (sample fraction), $S_y^2 = (N-1)^{-1}\sum_{i=1}^{N}(y_i - \bar{Y})^2$ (population mean square of y) and $C_y^2 = (S_y^2 / \bar{Y}^2)$ (coefficient of variation of y).

The usual ratio [(Cochran, 1940)] and product [(Robson, 1957) and (Murthy, 1964)] estimators of population mean $\bar{Y}$ have been defined respectively as

$$\bar{y}_R = \bar{y}\left(\frac{\bar{X}}{\bar{x}}\right),$$

$$\bar{y}_P = \bar{y}\left(\frac{\bar{x}}{\bar{X}}\right),$$

where $\bar{X}$, the population mean of auxiliary variable x is assumed to be known.

To the first degree of approximation the mean square errors (*MSE*s) of the ratio estimator $\bar{y}_R$ and the product estimator $\bar{y}_P$ are respectively given by

$$MSE(\bar{y}_R) = \theta \bar{Y}^2[C_y^2 + C_x^2 - 2\rho C_y C_x], \tag{2}$$

$$MSE(\bar{y}_P) = \theta \bar{Y}^2[C_y^2 + C_x^2 + 2\rho C_y C_x], \tag{3}$$

where $C_x = (S_x / \bar{X})$ (coefficient of variation of x), $S_x^2 = (N-1)^{-1}\sum_{i=1}^{N}(x_i - \bar{X})^2$ (population mean square of x), $\rho = (S_{xy} / S_x S_y)$ (correlation coefficient between y and x) and $S_{xy} = (N-1)^{-1}\sum_{i=1}^{N}(x_i - \bar{X})(y_i - \bar{Y})$ (covariance between x and y).

The difference estimator due to (Hansen *et al.*, 1953) has been defined as

$$\bar{y}_d = [\bar{y} + d(\bar{X} - \bar{x})], \tag{4}$$

where d is suitably chosen constant.

To the first degree of approximation the *MSE* of the difference estimator $\bar{y}_d$ is given by

$$MSE(\bar{y}_d) = \theta[S_y^2 + d^2 S_x^2 - 2dS_{xy}],$$

which is minimum when

$$d = \frac{S_{xy}}{S_x^2} = \beta. \tag{5}$$

Substitution of d at (5) in (4), yields the optimum estimator known as the usual linear regression estimator [(Watson, 1937)]

$$\bar{y}_\beta = [\bar{y} + \beta(\bar{X} - \bar{x})],$$

where β [Population regression coefficient of y on x] is assumed to be known.

To the first degree of approximation the MSE [or variance] of $\bar{y}_\beta$ [or minimum MSE of $\bar{y}_d$] is given by

$$MSE(\bar{y}_\beta) = Var(\bar{y}_\beta) = MSE_{\min}(\bar{y}_d) = \theta\bar{Y}^2(1-\rho^2)C_y^2. \tag{6}$$

It follows from (6) that the difference estimator $\bar{y}_d$ at its optimum condition is equal efficient as that of the usual linear regression estimator $\bar{y}_\beta$.

Some modifications over difference estimator $\bar{y}_d$ have been given by (Bedi and Hajela, 1984), (Jain, 1987), (Rao, 1991) and (Dubey and Singh, 2001). These estimators are respectively designed as

$\bar{y}_{BH} = w[\bar{y} + \beta(\bar{X} - \bar{x})]$, [(Bedi and Hajela, 1984)]

$\bar{y}_J = [\alpha_1\bar{y} + \alpha_2(\bar{X} - \bar{x})]$, $(\alpha_1 + \alpha_2) = 1$, [(Jain, 1987)]

$\bar{y}_{RG} = [k_1\bar{y} + k_2(\bar{X} - \bar{x})]$, [(Rao, 1991)]

$\bar{y}_{DS} = [m_1\bar{y} + m_2\bar{x} + (1 - m_1 - m_2)\bar{X}]$, [(Dubey and Singh, 2001)]

where β is as same as defined earlier and w, (α_1, α_2), (k_1, k_2) and (m_1, m_2) are suitably chosen constants to be determined such that the MSEs of $\bar{y}_{BH}$, $\bar{y}_{RG}$, $\bar{y}_J$ and $\bar{y}_{DS}$ are minimum respectively.

To the first degree of approximation, the MSEs of $\bar{y}_{BH}$, $\bar{y}_{RG}$, $\bar{y}_J$ and $\bar{y}_{DS}$ are respectively given by

$$MSE(\bar{y}_{BH}) = \bar{Y}^2[1 + w^2\{\theta C_y^2(1-\rho^2) + 1\} - 2w],$$

$$MSE(\bar{y}_J) = \bar{Y}^2[1 + \theta^2 R^{-2} C_x^2 - 2\alpha_1\{1 + \theta R^{-1}(R^{-1}C_X^2 + \rho C_y C_x)\}$$
$$+\alpha_1^2\{1 + \theta(C_y^2 + R^{-2}C_x^2 + 2R^{-1}\rho C_y C_x)\}],$$

$$MSE(\bar{y}_{RG}) = \bar{Y}^2[1 + k_1^2\{1 + \theta C_y^2\} + k_2^2 R^{-2} C_x^2\theta - 2k_1k_2R^{-1}\rho C_y C_x\theta - 2k_1],$$

$$MSE(\bar{y}_{DS}) = [\bar{Y}^2\theta[(m_1^2C_y^2 + m_2^2R^{-2}C_x^2 + 2m_1m_2R^{-1}\rho C_y C_x)] + (m_1 - 1)^2(\bar{Y} - \bar{X})^2],$$

where $R = (\bar{Y}/\bar{X})$.

The *MSE*s of $\bar{y}_{BH}$, $\bar{y}_J$, $\bar{y}_{RG}$ and $\bar{y}_{DS}$ are respectively minimized for

$$w = [1+\theta C_y^2(1-\rho^2)]^{-1} = w^*, \text{ (say)}$$

$$\alpha_1 = \frac{[1+\theta R^{-1}(R^{-1}C_x^2+\rho C_y C_x)]}{[1+\theta\{C_y^2+R^{-1}(R^{-1}C_x^2+2\rho C_y C_x)\}]} = \alpha_1^*, \text{ (say)}$$

$$\left.\begin{aligned} k_1 &= [1+\theta C_y^2(1-\rho^2)]^{-1} = k_1^* \\ k_2 &= [RK\{1+\theta C_y^2(1-\rho^2)\}^{-1}] = k_2^* \end{aligned}\right\}, \text{ (say)}$$

$$\left.\begin{aligned} m_1 &= [1+MSE(\bar{y}_\beta)(\bar{Y}-\bar{X})^{-2}]^{-1} = m_1^* \\ m_2 &= -\beta\, m_1^* = m_2^* \end{aligned}\right\}, \text{ (say)}$$

where $K = \rho(C_y/C_x)$ and β is same as defined earlier.

Thus the resulting minimum *MSE*s of $\bar{y}_{BH}$, $\bar{y}_J$, $\bar{y}_{RG}$ and $\bar{y}_{DS}$ are respectively given by

$$MSE_{\min}(\bar{y}_{BH}) = \frac{\bar{Y}^2\theta C_y^2(1-\rho^2)}{[1+\theta C_y^2(1-\rho^2)]} = \frac{MSE(\bar{y}_\beta)}{[1+\theta C_y^2(1-\rho^2)]}, \tag{7}$$

$$MSE_{\min}(\bar{y}_J) = \frac{\bar{Y}^2\theta C_y^2[1+\theta R^{-2}C_x^2(1-\rho^2)]}{[1+\theta\{C_y^2+R^{-1}(R^{-1}C_x^2+2\rho C_y C_x)\}]}, \tag{8}$$

$$MSE_{\min}(\bar{y}_{RG}) = \frac{\bar{Y}^2\theta C_y^2(1-\rho^2)}{[1+\theta C_y^2(1-\rho^2)]} = \frac{MSE(\bar{y}_\beta)}{[1+\theta C_y^2(1-\rho^2)]} = MSE_{\min}(\bar{y}_{BH}), \tag{9}$$

$$MSE_{\min}(\bar{y}_{DS}) = \frac{MSE(\bar{y}_\beta)}{[1+MSE(\bar{y}_\beta)(\bar{Y}-\bar{X})^{-2}]}. \tag{10}$$

We would like to mention here that, the estimators which we have discussed earlier such as usual ratio $(\bar{y}_R)$, usual product $(\bar{y}_P)$, usual difference $(\bar{y}_d)$, usual linear regression $(\bar{y}_\beta)$, (Bedi and Hajela, 1984) $(\bar{y}_{BH})$, (Jain, 1987) $(\bar{y}_J)$, (Rao, 1991) $(\bar{y}_{RG})$ and (Dubey and Singh, 2001) $(\bar{y}_{DS})$ used only population mean $\bar{X}$ as a auxiliary information on variable x. In the next section we have proposed a new class of estimators which have an improvement over usual linear regression estimator $\bar{y}_\beta$ by using several parameters of an auxiliary variable x such as mean $(\bar{X})$, mean square (S_x^2), coefficient of variation (C_x) etc.

3. PROPOSED CLASS OF ESTIMATORS

We have proposed the following class of estimators for estimating population mean $\bar{Y}$ of study variable y as

$$T = \left[w_1 \bar{y}_\beta + w_2 \bar{y} \left(\frac{\eta \bar{X} + \lambda}{\eta \bar{x} + \lambda} \right)^\alpha \right], \tag{11}$$

where (w_1, w_2) are suitably chosen scalars such that mean square error (*MSE*) of proposed class of estimators is minimum, (η, λ) are either constants or function of some known population parameters such as population mean $(\bar{X})$, population mean square (S_x^2), coefficient of variation (C_x) and population coefficient of kurtosis $(\beta_2(x))$ of the auxiliary variable x and α being constant which take finite values for designing the different estimators. Some members of proposed class of estimators [for different values of (η, λ, α)] have been summarized in Table 1.

To obtain the bias and *MSE* of proposed class of estimators T, we define

$$\bar{y} = \bar{Y}(1 + e_0),$$

$$\bar{x} = \bar{X}(1 + e_1),$$

such that

$$E(e_0) = E(e_1) = 0$$

and to the first degree of approximation

$$E(e_0^2) = \theta C_y^2, \; E(e_1^2) = \theta C_x^2, \; E(e_0 e_1) = \theta \rho C_y C_x .$$

Expressing (11) in terms of e's, we have

$$T = [w_1 \bar{Y}(1 + e_0 - K e_1) + w_2 \bar{Y}(1 + e_0)(1 + \tau e_1)^{-\alpha}], \tag{12}$$

where K is as same as defined earlier and $\tau = \eta \bar{X}(\eta \bar{X} + \lambda)^{-1}$.

TABLE 1

Some members of proposed class of estimators T

Estimator	Values of constant		
	α	η	λ
$T_1 = \left[w_1\bar{y}_\beta + w_2\bar{y}\left(\frac{\bar{X}}{\bar{x}}\right)\right]$	1	1	0
$T_2 = \left[w_1\bar{y}_\beta + w_2\bar{y}\left(\frac{(C_x\bar{X}-\bar{X})}{(C_x\bar{x}-\bar{X})}\right)\right]$	1	C_x	$-\bar{X}$
$T_3 = \left[w_1\bar{y}_\beta + w_2\bar{y}\left(\frac{(C_x^2\bar{x}+S_x)}{(C_x^2\bar{X}+S_x)}\right)\right]$	-1	C_x^2	S_x
$T_4 = \left[w_1\bar{y}_\beta + w_2\bar{y}\left(\frac{(C_x\bar{X}+\rho)}{(C_x\bar{x}+\rho)}\right)\right]$	1	C_x	ρ
$T_5 = \left[w_1\bar{y}_\beta + w_2\bar{y}\left(\frac{(S_x\bar{x}+\rho)}{(S_x\bar{X}+\rho)}\right)\right]$	-1	S_x	ρ
$T_6 = \left[w_1\bar{y}_\beta + w_2\bar{y}\left(\frac{(S_x-\rho\bar{x})}{(S_x-\rho\bar{X})}\right)\right]$	-1	- ρ	S_x
$T_7 = \left[w_1\bar{y}_\beta + w_2\bar{y}\left(\frac{(\rho\bar{X}-C_x^2)}{(\rho\bar{x}-C_x^2)}\right)\right]$	1	ρ	- C_x^2
$T_8 = \left[w_1\bar{y}_\beta + w_2\bar{y}\left(\frac{(\bar{x}-S_x)}{(\bar{X}-S_x)}\right)\right]$	-1	1	- S_x

We assume $|\tau e_1| < 1$, so that the term $(1+\tau e_1)^{-\alpha}$ is expandable. Thus by expanding the right hand side of (12) and neglecting the terms of e's having power greater than two, we have

$$T = \bar{Y}[w_1(1+e_0-Ke_1)+w_2(1+e_0)\{1-\alpha\tau e_1+(\alpha(\alpha+1)/2)\tau^2 e_1^2\}]$$

or

$$(T-\bar{Y}) = \bar{Y}[w_1(1+e_0-Ke_1)+w_2\{1+e_0-\alpha\tau e_1-\alpha\tau e_0 e_1+(\alpha(\alpha+1)/2)\tau^2 e_1^2\}-1]. \tag{13}$$

Taking expectation on both sides of (13) we get the bias of T to the first degree of approximation as

$$B(T) = \bar{Y}[w_1+w_2\{1+\theta((\alpha\tau C_x^2/2))((\alpha+1)\tau-2K)\}-1]. \tag{14}$$

Squaring both sides of (13) and neglecting terms of e's having power greater than two, we have

$$(T-\bar{Y})^2 = \bar{Y}^2[1+w_1^2(1+2e_0-2Ke_1+e_0^2+K^2e_1^2-2Ke_0e_1)$$

$$+w_2^2\{1+e_0^2+2e_0-2\alpha\tau e_1+\alpha\tau^2(2\alpha+1)e_1^2-4\alpha\tau e_0 e_1\}$$

$$+2w_1w_2\{1+2e_0-(K+\alpha\tau)e_1+e_0^2-(K+2\alpha\tau)e_0e_1+\alpha\tau(K+((\alpha+1)/2)\tau)e_1^2\}$$

$$-2w_1(1+e_0-Ke_1)-2w_2\{1+e_0-2\tau e_1-\alpha\tau e_0e_1+(\alpha(\alpha+1)/2)\tau^2e_1^2\}]. \quad (15)$$

Taking expectation on both sides of (15) we get the MSE of proposed class of estimators T to the first degree of approximation as

$$MSE(T)=\bar{Y}^2[1+w_1^2A+w_2^2B+2w_1w_2C-2w_1-2w_2D], \quad (16)$$

where

$$A=[1+\theta C_y^2(1-\rho^2)],$$

$$B=[1+\theta\{C_y^2+\alpha\tau C_x^2((2\alpha+1)\tau-4K)\}],$$

$$C=[1+\theta\{C_y^2(1-\rho^2)+\alpha\tau C_x^2(((\alpha+1)/2)\tau-K)\}],$$

$$D=[1+\theta\alpha\tau C_x^2\{((\alpha+1)/2)\tau-K\}].$$

Differentiating (16) with respect to w_1 and w_2 and equating them to zero, we get the following normal equations

$$\begin{bmatrix} A & C \\ C & B \end{bmatrix}\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}=\begin{bmatrix} 1 \\ D \end{bmatrix}. \quad (17)$$

Solving (17) we get the optimum values of w_1 and w_2 as

$$w_1=\frac{(B-CD)}{(AB-C^2)}=w_1^*, \text{ (say)} \quad (18)$$

$$w_2=\frac{(AD-C)}{(AB-C^2)}=w_2^*, \text{ (say)}. \quad (19)$$

Substituting (18) and (19) in (16) we get the minimum MSE of proposed class of estimators T as

$$MSE_{\min}(T)=\bar{Y}^2\left[1-\frac{(B-2CD+AD^2)}{(AB-C^2)}\right]. \quad (20)$$

Thus we established the following theorem.

Theorem 1: To the first degree of approximation

$$MSE(T) \geq \bar{Y}^2 \left[1 - \frac{(B - 2CD + AD^2)}{(AB - C^2)} \right]$$

with equality holding if

$$w_1 = w_1^* \text{ and } w_2 = w_2^* \ .$$

We would like to mention here that the proposed class of estimator T is reduced to some known estimators of population mean $\bar{Y}$ by putting a different values of $(w_1, w_2, \eta, \lambda, \alpha)$. i.e.

$(w_1, w_2, \eta, \lambda, \alpha) = (0, 1, 1, 0, 1)$; $T \to \bar{y}_R$,

$(w_1, w_2, \eta, \lambda, \alpha) = (0, 1, 1, 0, -1)$; $T \to \bar{y}_P$,

$(w_1, w_2, \eta, \lambda, \alpha) = (1, 0, -, -, -)$; $T \to \bar{y}_\beta$,

$(w_1, w_2, \eta, \lambda, \alpha) = (w_1, 0, -, -, -)$; $T \to \bar{y}_{BH}$.

In the next section we have considered a particular case of proposed class of estimators T by making the constraint on scalars w_1 and w_2 .

4. PARTICULAR CASE

For $(w_1 + w_2 = 1)$, the suggested class of estimators T is reduced to the new class of estimators defined as

$$T^* = \left[w_1 \bar{y}_\beta + (1 - w_1) \bar{y} \left(\frac{\eta \bar{X} + \lambda}{\eta \bar{x} + \lambda} \right)^\alpha \right],$$

where w_1 , η , λ and α are as same as defined earlier.

We can easily get for the first degree of approximation the bias and MSE of T^* by putting $(w_2 = (1 - w_1))$ in (14) and (16) respectively as

$$B(T^*) = \bar{Y}[w_1 + (1 - w_1)\{1 + \theta((\alpha \tau C_x^2 / 2))((\alpha + 1)\tau - 2K)\} - 1],$$

$$MSE(T^*) = \bar{Y}^2[1 + B - 2D + w_1^2(A + B - 2C) - 2w_1(1 + B - C - D)]. \quad (21)$$

Differentiating (21) with respect to w_1 and equating to zero we get the optimum value of w_1 as

$$w_1 = \frac{(1+B-C-D)}{(A+B-2C)}. \tag{22}$$

Inserting (22) in (21) we get the minimum MSE of T^* as

$$MSE_{\min}(T^*) = \bar{Y}^2\left[1+B-2D-\frac{(1+B-C-D)^2}{(A+B-2C)}\right] = \theta\bar{Y}^2(1-\rho^2)C_y^2 \tag{23}$$

$$= MSE(\bar{y}_\beta) \text{ [or } MSE_{\min}(\bar{y}_d)].$$

Thus we established the following theorem.

Theorem 2: To the first degree of approximation

$$MSE(T^*) \geq \bar{Y}^2\left[1+B-2D-\frac{(1+B-C-D)^2}{(A+B-2C)}\right]$$

with equality holding if

$$w_1 = w^{**}.$$

It clearly indicates from (23) that the particular case T^* of proposed class of estimators T is as efficient as that of the usual linear regression estimator $\bar{y}_\beta$ at its optimum condition.

5. EFFICIENCY COMPARISONS

From (1), (2), (3), (6) and (23) we have

$$[Var(\bar{y}) - MSE(\bar{y}_\beta) \text{ [or } MSE_{\min}(\bar{y}_d) \text{ or } MSE_{\min}(T^*)]] = \theta\bar{Y}^2C_y^2\rho^2 \geq 0, \tag{24}$$

$$[MSE(\bar{y}_R) - MSE(\bar{y}_\beta) \text{ [or } MSE_{\min}(\bar{y}_d) \text{ or } MSE_{\min}(T^*)]] = \theta\bar{Y}^2C_y^2(1-K)^2 \geq 0, \tag{25}$$

$$[MSE(\bar{y}_P) - MSE(\bar{y}_\beta) \text{ [or } MSE_{\min}(\bar{y}_d) \text{ or } MSE_{\min}(T^*)]] = \theta\bar{Y}^2C_y^2(1+K)^2 \geq 0. \tag{26}$$

From (24)-(26), it is observed that the usual regression estimator $\bar{y}_\beta$, usual difference estimator $\bar{y}_d$ (at their optimum condition) and class of estimators T^* (at their optimum condition) are more efficient than

(i) the usual unbiased estimator $\bar{y}$ provided $\rho \neq 0$. For $\rho = 0$, the estimators $\bar{y}$, $\bar{y}_\beta$, $\bar{y}_d$ and T^* are equally efficient.

(ii) the usual ratio estimator $\bar{y}_R$ provided $K \neq 1$ [or $R \neq \beta$]. In case $K = 1$ [or $R = \beta$] the estimators $\bar{y}$, $\bar{y}_R$, $\bar{y}_\beta$, $\bar{y}_d$ and T^* are equally efficient.

(iii) the usual product estimator $\bar{y}_P$ provided $K \neq -1$ [or $R \neq -\beta$]. In case $K = -1$ [or $R = -\beta$] the estimators $\bar{y}$, $\bar{y}_P$, $\bar{y}_\beta$, $\bar{y}_d$ and T^* are equally efficient.

From (6), (7), (9), (10), (20) and (23) we have

$$[MSE(\bar{y}_\beta) \text{ [or } MSE_{\min}(\bar{y}_d) \text{ or } MSE_{\min}(T^*)] - MSE_{\min}(\bar{y}_{BH}) \text{ [or } MSE_{\min}(\bar{y}_{RG})]] = \frac{MSE_{\min}(\bar{y}_{BH}) MSE(\bar{y}_\beta)}{\bar{Y}^2} \geq 0, \quad (27)$$

$$[MSE(\bar{y}_\beta) \text{ [or } MSE_{\min}(\bar{y}_d) \text{ or } MSE_{\min}(T^*)] - MSE_{\min}(\bar{y}_{DS})] = [MSE(\bar{y}_\beta)]^2 (\bar{Y} - \bar{X})^{-2} \geq 0, \quad (28)$$

$$[MSE(\bar{y}_\beta) \text{ [or } MSE_{\min}(\bar{y}_d) \text{ or } MSE_{\min}(T^*)] - MSE_{\min}(T)] = \frac{\bar{Y}^2 [B - C - A(B-D) - C(D-C)]^2}{(AB - C^2)(A + B - 2C)} \geq 0. \quad (29)$$

It is observed from (27)-(29) that the estimators $\bar{y}_{BH}$ (Bedi and Hajela, 1984), $\bar{y}_{RG}$ (Rao, 1991), $\bar{y}_{DS}$ (Dubey and Singh, 2001) and the proposed class of estimators T are more efficient (at their optimum conditions) than the usual regression estimator $\bar{y}_\beta$ and hence more efficient than the estimators $\bar{y}$, $\bar{y}_R$, $\bar{y}_P$, $\bar{y}_d$ and class of estimators T^*.

From (7), (8), (9) and (10) we have

$$MSE_{\min}(\bar{y}_{DS}) < MSE_{\min}(\bar{y}_{BH}) \text{ [or } MSE_{\min}(\bar{y}_{RG})] \text{ if } [1 < 2R],$$

$$MSE_{\min}(\bar{y}_{DS}) < MSE_{\min}(\bar{y}_J) \text{ if } \left[\rho^2 > \frac{2\theta K C_x^2 R^{-1}}{2\theta K C_x^2 R^{-1} + 1}\right].$$

From (7), (9) and (20) we have

$$[MSE_{\min}(\bar{y}_{BH}) \text{ [or } MSE_{\min}(\bar{y}_{RG})] - MSE_{\min}(T)] = \frac{\bar{Y}^2 [(C - AD)]^2}{A(AB - C^2)} \geq 0.$$

In the next section we have made an empirical study to judge the theoretical results practically considering two natural population data sets.

6. EMPIRICAL STUDY

To evaluate the performance of estimators T_i, (i = 1, 2, ..., 8) which are members of proposed class of estimators T over other competitors, we have considered two population data sets. The descriptions of population data sets are as follows.

Population I: It consists of 106 villages of Marmara region of Turkey in 1999. The study variable y and auxiliary variable x being the level of apple production and the number of apple trees respectively, (Kadilar and Cingi, 2003).

Population II: It consists of 106 villages of Aegean region of Turkey in 1999. The study variable y and auxiliary variable x being the level of apple production and the number of apple trees respectively, (Kadilar and Cingi, 2004).

The required values of parameters are summarized in Table 2.

TABLE 2

The required parameters of populations under study

Population	Parameter								
	N	n	$\bar{Y}$	$\bar{X}$	C_y	C_x	ρ	θ	R
I	106	20	1536.77	24375.59	4.18	2.02	0.82	0.04	0.06
II	106	20	2212.59	27421.70	5.22	2.10	0.86	0.04	0.08

We have computed the percent relative efficiencies (*PRE*s) of different estimators t_0, with respect to the usual unbiased estimator $\bar{y}$ as

$$PRE(t_0, \bar{y}) = \frac{Var(\bar{y})}{MSE_{\min}(t_0)} * 100$$

and the results are displayed in Table 3.

It is observed form Table 3 that

- the usual linear regression $\bar{y}_\beta$ which is equally efficient as that of the usual difference estimator $\bar{y}_d$ (at its optimum condition) and the class of estimators T^* (at its optimum condition) performed better than the usual unbiased estimator $\bar{y}$, usual ratio estimator $\bar{y}_R$ (Cochran, 1940) and usual product estimator $\bar{y}_P$ (Robson, 1957) and (Murthy, 1967).
- the estimators T_i, (i = 1, 2, ..., 8) (members of proposed class of estimators T), $\bar{y}_{BH}$ (Bedi and Hajela, 1984), $\bar{y}_J$ (Jain, 1987), $\bar{y}_{RG}$ (Rao, 1991) and $\bar{y}_{DS}$

(Dubey and Singh, 2001) are more efficient than the usual linear regression estimator $\bar{y}_\beta$ and hence more efficient than the estimators $\bar{y}$, $\bar{y}_R$, $\bar{y}_P$, $\bar{y}_d$ and T^*.

- the estimators T_i, (i = 1, 2, ..., 8) which are members of proposed class of estimators T performed better (i.e. having larger *PRE*s) than the estimators $\bar{y}_{BH}$, $\bar{y}_{RG}$, $\bar{y}_J$, and $\bar{y}_{DS}$.
- the estimators T_8 (utilizes the information on $\bar{X}$ and S_x) and T_2 (utilizes the information on $\bar{X}$ and C_x) are best in the sense of having largest *PRE*s ($PRE(T_8, \bar{y})$ = 964.27 and $PRE(T_2, \bar{y})$ = 701.67) among all the estimators discussed here for population data sets I and II respectively.

Finally we have concluded that the proposed class of estimators T which utilizes the information on several population parameters of auxiliary variable x has an improvement over regression method of estimation and other existing estimators of population mean $\bar{Y}$ which utilize the information only on population mean of auxiliary variable x.

TABLE 3

PREs of different estimators with respect to $\bar{y}$

Estimator (t_0)	$PRE(t_0, \bar{y})$	
	Population – I	Population – II
$\bar{y}$	100.00	100.00
$\bar{y}_R$	226.76	212.82
$\bar{y}_P$	49.36	53.94
$\bar{y}_d$, $\bar{y}_\beta$, T^*	305.25	384.02
$\bar{y}_{BH}$, $\bar{y}_{RG}$	376.13	494.56
$\bar{y}_J$	356.92	478.99
$\bar{y}_{DS}$	305.57	384.88
T_1	390.19	518.64
T_2	507.29	701.67*
T_3	378.94	500.05
T_4	390.19	518.64
T_5	381.10	504.57
T_6	395.73	520.24
T_7	390.20	518.65
T_8	964.27*	584.12

* Indicates that the estimator has largest efficiency.

7. CONCLUSION

In this article we have suggested a class of estimators for the population mean of the study variable when information on an auxiliary variable is known in advance. The asymptotic bias and mean square error formulae of proposed class have been obtained. The asymptotic optimum estimator in the proposed class has been identified with its properties. Some new members of proposed class have been generated by putting different values of scalars involved in proposed class. It has been identified theoretically that the proposed method of estimation (at its optimum condition) is better than the traditional methods of estimation such as usual unbiased, ratio, product, difference, linear regression and methods proposed by (Bedi and Hajela, 1984) and (Rao, 1991). An empirical study is carried out to judge the merits of proposed class over other competitors by using two natural population data sets with satisfying all theoretical results. It has been shown empirically that the proposed method also has an improvement over methods suggested by (Jain, 1987) and (Dubey and Singh, 2001). Thus the proposed class of estimators has been recommended for its use in practice. However these conclusions cannot be extrapolated in general due to limited empirical study.

School of Studied in Statistics
Vikram University Ujjain-456010, M.P., India

HOUSILA P. SINGH
ANJANA RATHOUR
RAMKRISHNA S. SOLANKI

ACKNOWLEDGEMENT

Authors are thankful to the learned referee for his valuable suggestions regarding improvement of the paper.

REFERENCES

C. KADILAR, H. CINGI (2003), *A study on the chain ratio-type estimator*, "Hacett. Jour. Math. Statist.", 32, pp. 105-108.

C. KADILAR, H. CINGI (2004), *Ratio estimators in simple random sampling*, "Appl. Math. Comp.", 151, 3, pp. 893-902.

D.J. WATSON (1937), *The estimation of leaf area in field crops*, "Jour. Agric Sci.", 27, pp. 474-483.

D.S. ROBSON (1957), *Application of multivariate polykays to the theory of unbiased ratio type estimators*, "Jour. Amer. Statist. Assoc.", 52, pp. 511-522.

G. DIANA (1993), *A class of estimators of the population mean in stratified random sampling*, "Statistica", 53, 1, pp. 59-66.

H.P. SINGH, N. KARPE (2009), *On the estimation of ratio and product of two population means using supplementary information in presence of measurement errors*, "Statistica", 69, 1, pp. 27-47.

H.P. SINGH, R. TAILOR (2005), *Estimation of finite population mean using known correlation coefficient between auxiliary characters*, "Statistica", 65, 4, pp. 407-418.

H.P. SINGH, R.S. SOLANKI (2012), *Improved estimation of population mean in simple random sampling using information on auxiliary attribute*, "Appl. Math. Comp.", 218, pp. 7798-7812.

L.N. UPADHYAYA, H.P. SINGH (1999), *Use of transformed auxiliary variable in estimating the finite population mean*, "Biometrical Jour.", 41, 5, pp. 627-636.

M.N. HANSEN, W.N. HURWITZ, W.G. MADOW (1953), *Sample Survey Methods and Theory*, John Wiley and Sons, New York, USA.

M.N. MURTHY (1964), *Product method of estimation*, "Sankhya", 26, pp. 69-74.

M.N. MURTHY (1967), *Sampling Theory and Methods*. Statistical Publishing Society, Calcutta, India.

P.K. BEDI, D. HAJELA (1984), *An estimator for population mean utilizing know coefficient of variation and auxiliary variable*, "Jour. Statist. Res.", 18, pp. 29-33.

R.K. JAIN (1987), *Properties of estimators in simple random sampling using auxiliary variable*, "Metron", 45, 1-2, pp.265-271.

S. SINGH (2003), *Advanced Sampling Theory with Applications How Michael 'selected' Amy*, Kluwer Academic Publishers, The Netherlands.

T.J. RAO (1991), *On certain methods of improving ratio and regression estimators*, "Commun. Statist. Theo. Meth.", 20, 10, pp. 3325-3340.

V. DUBEY, S.K. SINGH (2001), *An improved regression estimator for estimating population mean*, "Jour. Ind. Soc. Agri. Statist.", 54, pp. 179-183.

W.G. COCHRAN (1940), *The estimation of the yields of cereal experiments by sampling for the ratio gain to total produce*, "Jour. Agric. Soc.", 30, pp. 262-275.

W.G. COCHRAN (1977), *Sampling Techniques*, John Wiley and Sons, New York, USA.

SUMMARY

An improvement over regression method of estimation

This paper suggested a class of estimators for the population mean of the study variable using information on an auxiliary variable with its properties under large sample approximation. The asymptotic optimum estimator in the proposed class has been identified with its properties. In addition, some existing estimators have been founded members of proposed class. It has been identified theoretically that the proposed class of estimators is better than the some traditional methods of estimation. An empirical study is carried out to judge the merits of proposed class over other competitors by using two natural population data sets.