Pak. J. Statist.
2010 Vol. 26(2), 427-443

ON THE FAMILY OF ESTIMATORS OF POPULATION MEAN IN STRATIFIED RANDOM SAMPLING

Nursel Koyuncu and **Cem Kadilar**
Hacettepe University, Department of Statistics, Beytepe, Ankara, Turkey
Email: [1]nkoyuncu@hacettepe.edu.tr and [2]kadilar@hacettepe.edu.tr

ABSTRACT

Diana (1993) introduced a family of estimators in the stratified random sampling to estimate the population mean. Following the articles of Diana (1993) and Kadilar and Cingi (2003), we propose a new family of estimators in the stratified random sampling that includes the estimators suggested by Kadilar and Cingi (2003), Shabbir and Gupta (2005), Singh et al. (2008). Up to the first and second order of approximations, we obtain the mean square error (MSE) and the optimum case is discussed. Also an empirical study is carried out to show the properties of the proposed estimators.

KEY WORDS

Ratio estimator; Auxiliary information; Mean square error; Efficiency; Stratified random sampling.

Mathematics Subject Classification: Primary 62D05

1. INTRODUCTION

The use of the auxiliary information in sample surveys results in considerable improvement in the precision of estimators of the population mean. Whenever there is auxiliary information available, the researchers want to utilize it in the method of estimation to obtain the most efficient estimator. Ratio method is used to obtain more efficient estimates for the population mean by taking the advantage of the correlation between an auxiliary variable and study variable. In the stratified random sampling scheme, Hansen *et al.* (1946) and Kaur (1985) proposed ratio estimators using the population mean of the auxiliary variable. In simple random sampling, Srivastava (1971) suggested general classes of estimators which contain many estimators in the literature. They prove that the best estimator in their proposed classes is always a regression-type estimator and for the estimators belonging to this general class, no further improvement for their performances is possible. Because of this reason, Diana (1992) calculated the MSE of the estimators using the k-th order approximation for the Taylor series to find the most efficient estimator. Diana (1993) extended this class of estimators of the population mean which includes some other estimators proposed in the literature in the stratified sampling and examined the *MSE* equation of these estimators up to the *k*-th order approximation. Kadilar and Cingi (2005) extended Prasad (1989) estimator, Singh and Vishwakarma (2006) extended Sahai (1979) estimator for the stratified random sampling. Shabbir and Gupta (2006) suggested a new estimator using Bedi (1996) transformation.

Singh and Vishwakarma (2008) suggested a family of estimators using transformation in the stratified random sampling. Moreover, Kadilar and Cingi (2003), Shabbir and Gupta (2005), Singh *et al.* (2008), Koyuncu and Kadilar (2009) suggested various estimators using knowledge of population parameters of the auxiliary information in the stratified random sampling. In this study, motivated by Diana (1993) and Kadilar and Cingi (2003), we propose a new family of estimators in the stratified random sampling which includes some new estimators and some others proposed in literature such as Kadilar and Cingi (2003), Shabbir and Gupta (2005), Singh *et al.* (2008).

Consider a finite population $U=\left(u_1,u_2,...,u_N\right)$ of size *N* and let *y* and *x*, respectively, be the study and auxiliary variables associated with each unit u_j $\left(j=1,2,....,N\right)$ of the population. Let the population of size, *N*, is stratified into *L* strata with *h*-th stratum containing N_h units, where $h=1,2,...,L$ such that $\sum_{h=1}^{L} N_h = N$. A simple random sample of size n_h is drawn without replacement from the *h*-th stratum such that $\sum_{h=1}^{L} n_h = n$. Let $\left(y_{hi},x_{hi}\right)$ denote the observed values of *y* and *x* on the *i*-th unit of the *h*-th stratum, where $i=1,2,...,N_h$ and $h=1,2,...,L$. Moreover, let $\bar{y}_h=\sum_{i=1}^{n_h}\frac{y_{hi}}{n_h}$, $\bar{y}_{st}=\sum_{h=1}^{L} W_h \bar{y}_h$, and $\bar{Y}_h=\sum_{i=1}^{N_h}\frac{y_{hi}}{N_h}$, $\bar{Y}=\bar{Y}_{st}=\sum_{h=1}^{L} W_h \bar{Y}_h$ be the sample and population means of *y*, respectively, where $W_h=\frac{N_h}{N}$ is the stratum weight. Similar expressions for *x* can also be defined.

To obtain the bias and the *MSE*, let us define $e_0=\left(\bar{y}_{st}-\bar{Y}\right)\big/\bar{Y}$ and $e_1=\left(\bar{x}_{st}-\bar{X}\right)\big/\bar{X}$. Using these notations,

$$E\left(e_0\right)=E\left(e_1\right)=0,$$

$$V_{r,s}=\sum_{h=1}^{L} W_h^{r+s}\frac{E\left[\left(\bar{x}_h-\bar{X}_h\right)^r\left(\bar{y}_h-\bar{Y}_h\right)^s\right]}{\bar{X}^r\bar{Y}^s}. \tag{1.1}$$

From (1.1), we can write

$$E\left(e_0^2\right)=\frac{\sum_{h=1}^{L} W_h^2\gamma_h S_{yh}^2}{\bar{Y}^2}=V_{0,2},\quad E\left(e_1^2\right)=\frac{\sum_{h=1}^{L} W_h^2\gamma_h S_{xh}^2}{\bar{X}^2}=V_{2,0},$$

$$E\left(e_0e_1\right)=\frac{\sum_{h=1}^{L} W_h^2\gamma_h S_{xyh}}{\bar{X}\bar{Y}}=V_{1,1},$$

where

$$S_{yh}^2 = \frac{\sum_{i=1}^{N_h}\left(y_{hi}-\bar{Y}_h\right)^2}{N_h-1}, \ S_{xh}^2 = \frac{\sum_{i=1}^{N_h}\left(x_{hi}-\bar{X}_h\right)^2}{N_h-1}, \ S_{xyh} = \frac{\sum_{i=1}^{N_h}\left(y_{hi}-\bar{Y}_h\right)\left(x_{hi}-\bar{X}_h\right)}{N_h-1},$$

$\gamma_h = \dfrac{1-f_h}{n_h}$, and $f_h = \dfrac{n_h}{N_h}$.

In the stratified random sampling, when information on an auxiliary variable is available, a well known estimator for the population mean is the combined ratio estimator defined as

$$\bar{y}_{RC} = \left(\frac{\bar{y}_{st}}{\bar{x}_{st}}\right)\bar{X}. \tag{1.2}$$

Using the notation (1.1), the bias and the *MSE* of this estimator, to the first-order of approximation, are respectively given by

$$Bias\left(\bar{y}_{RC}\right) = \bar{Y}\left(V_{2,0} - V_{1,1}\right), \tag{1.3}$$

$$MSE\left(\bar{y}_{RC}\right) = \bar{Y}^2\left(V_{0,2} + V_{2,0} - 2V_{1,1}\right). \tag{1.4}$$

Diana (1993) suggested a family of estimators for the population mean in the stratified random sampling as

$$\bar{y}_{CST} = \bar{y}_{ST}\left(\frac{\bar{x}_{ST}}{\bar{X}}\right)^{\delta}\left[w + (1-w)\left(\frac{\bar{x}_{ST}}{\bar{X}}\right)^{\varepsilon}\right]^{\eta} \tag{1.5}$$

where δ, ε, η, and w can take finite values. When these four parameters are conveniently chosen, many estimators are obtainable. Also, when one parameter is considered as "free parameter", it is possible to obtain some subclasses of estimators. Here, free parameter means that one of the four parameters is such a scalar that the mean square error of $\bar{y}_{CST}$ gets the minimum value. In other words, we minimize the MSE equation according to the free parameter. Some estimators, which are generated from (1.5) for different combinations of δ, ε, η, and w, are given in Table 1. Rewriting (1.5) in terms of *e*'s, we have

$$\bar{y}_{CST} = \bar{y}_{ST}\left(1+e_1\right)^{\delta}\left[w + (1-w)\left(1+e_1\right)^{\varepsilon}\right]^{\eta}. \tag{1.6}$$

To obtain the bias and the *MSE*, the term $\left(1+e_1\right)^{\delta}$ is expandable as

$$\left(1+e_1\right)^{\delta} = a_0 + a_1 e_1 + a_2 e_1^2 + a_3 e_1^3 + a_4 e_1^4 + \ldots \tag{1.7}$$

where

$$a_i = \begin{cases} \frac{\delta(\delta-1)...(\delta-i+1)}{i!} & i = 1,2,3,... \\ 1 & i = 0 \end{cases} \tag{1.8}$$

and similarly the term $\left[w+(1-w)(1+e_1)^{\varepsilon}\right]^{\eta}$ in (1.6) is expandable as

$$\left[w+(1-w)(1+e_1)^{\varepsilon}\right]^{\eta} = b_0 + b_1 e_1 + b_2 e_1^2 + b_3 e_1^3 + b_4 e_1^4 + ... \tag{1.9}$$

where

$$b_i = \begin{cases} \frac{1}{i!}\frac{\partial^i}{\partial(e_1)^i}\left[w+(1-w)(1+e_1)^{\varepsilon}\right]^{\eta}\Big|_{e_1=0} & i = 1,2,3,... \\ 1 & i = 0 \end{cases} . \tag{1.10}$$

Up to the first order of approximation, the bias and the *MSE* of the estimator $\overline{y}_{CST}$ are respectively given by

$$Bias_I\left(\overline{y}_{CST}\right) = \overline{Y}\left(c_1 V_{1,1} + c_2 V_{2,0}\right), \tag{1.11}$$

$$MSE_I\left(\overline{y}_{CST}\right) = \overline{Y}^2\left(V_{0,2} + c_1^2 V_{2,0} + 2c_1 V_{1,1}\right), \tag{1.12}$$

where $c_i = \sum_{j=0}^{i} \alpha_j b_{i-j}$.

Note that the optimum value of c_1 is obtained as $c_{1(opt)} = -V_{1,1}/V_{2,0}$. Using this optimal value, the minimum *MSE* of the estimator $\overline{y}_{CST}$ is

$$MSE_{I\min}\left(\overline{y}_{CST}\right) = \overline{Y}^2\left(V_{0,2} - \frac{V_{1,1}^2}{V_{2,0}}\right) = \sum_{h=1}^{L} W_h^2 \lambda_h S_{yh}^2 \left\{1-\rho_c^2\right\} = MSE\left(\overline{y}_{lrc}\right) \tag{1.13}$$

which is equal to the *MSE* of the combined regression estimator. This means that members of this family of estimators, which are generated by considering one parameter free, have the same *MSE* with the combined regression estimator. To find the most efficient estimator among $\overline{y}_{CST}$, it is useful to find their *MSE* equations up to the second order of approximation. Up to the second order of approximation, the bias and the *MSE* of the estimator $\overline{y}_{CST}$ are respectively given by

$$Bias_{II}\left(\bar{y}_{CST}\right)=Bias_{1}\left(\bar{y}_{CST}\right)+\bar{Y}\left(c_{2}V_{2,1}+c_{3}\left(V_{3,0}+V_{3,1}\right)+c_{4}V_{4,0}\right), \tag{1.14}$$

$$MSE_{II}\left(\bar{y}_{CST}\right)=MSE_{1}\left(\bar{y}_{CST}\right)+\bar{Y}^{2}\left[\left(c_{2}^{2}+2c_{1}c_{3}\right)V_{4,0}+2c_{1}c_{2}V_{3,0}+2\left(2c_{1}c_{2}+c_{3}\right)V_{3,1}\right.$$
$$\left.+2\left(c_{1}^{2}+c_{2}\right)V_{2,1}+\left(c_{1}^{2}+2c_{2}\right)V_{2,2}+2c_{1}V_{1,2}\right]. \tag{1.15}$$

The optimum value of "free parameter" is obtained by minimizing the $MSE_{II}\left(\bar{y}_{CST}\right)$. Solution for the determination of this value is obtained by using the "fminbnd" function in Matlab. To obtain the *MSE* equations for the second order expressions, we have used some results given in Sukhatme *et al.* (1984).

2. SUGGESTED FAMILY OF ESTIMATORS

Following Diana (1993) and using some known population parameters, we can define a general family of estimators for the population mean as

$$\bar{y}_{N}=\bar{y}_{ST}\left(\frac{\bar{x}_{st}}{\bar{X}}\right)^{\delta}\left[w+\left(1-w\right)\left(\frac{\bar{x}_{st}A_{st}+B_{st}}{\bar{X}A_{st}+B_{st}}\right)^{\varepsilon}\right]^{\eta}, \tag{2.1}$$

where $A_{st}=\sum_{h=1}^{L}W_{h}A_{h}$, $B_{st}=\sum_{h=1}^{L}W_{h}B_{h}$. Here A_h and B_h may be the population information of the auxiliary variable for the *h*-th stratum such as S_{xh}, coefficient of variation C_{xh}, skewness $\beta_{1(x)h}$, kurtosis $\beta_{2(x)h}$, correlation coefficient $\rho_{h(xy)}$. Note that Koyuncu and Kadilar (2009) use these definitions in their family of estimators. δ, ε, η, and w can be defined as in Diana (1993). Many new estimators, which are generated from (2.1) for different combinations of δ, ε, η, and w, are given in Table 2.

Expressing the estimator, $\bar{y}_N$, in terms of e_i (i = 0, 1), we can write (16) as

$$\bar{y}_{N}=\bar{Y}\left(1+e_{0}\right)\left(1+e_{1}\right)^{\delta}\left[w+\left(1-w\right)\left(1+t_{st}e_{1}\right)^{\varepsilon}\right]^{\eta}, \tag{2.2}$$

where $t_{st}=\dfrac{\bar{X}A_{st}}{\bar{X}A_{st}+B_{st}}$.

Up to the first order of approximation, the bias and the *MSE* of the estimator $\bar{y}_N$ are respectively given by

$$Bias_{1}\left(\bar{y}_{N}\right)=\bar{Y}\left(c_{1}^{*}V_{1,1}+c_{2}^{*}V_{2,0}\right), \tag{2.3}$$

$$MSE_{I}\left(\bar{y}_{N}\right)=\bar{Y}^{2}\left(c_{1}^{*2}V_{2,0}+V_{0,2}+2c_{1}^{*}V_{1,1}\right), \tag{2.4}$$

where $c_{i}^{*}=\sum_{j=0}^{i}\alpha_{j}b_{i-j}\left(t_{st}\right)^{i-j}$ and optimum value of c_1^* is obtained as $c_{1(opt)}^{*}=-V_{1,1}/V_{2,0}$. Then, substituting this value in (2.4), we can get the minimum *MSE* as

$$MSE_{I(\min)}\left(\bar{y}_N\right)=\sum_{h=1}^{L} W_h^2 \lambda_h S_{yh}^2\left\{1-\rho_c^2\right\}=MSE\left(\bar{y}_{lrc}\right) \tag{2.5}$$

which is equal to (1.13). This means that subclasses of this family have the same *MSE* with the combined regression estimator up to the first order of approximation, whereas up to the second order approximation, we can give *MSE* as

$$MSE_{II}\left(\bar{y}_N\right)=\bar{Y}^2\left(c_1^{*2}V_{2,0}+V_{0,2}+2c_1^*V_{1,1}+2c_1^*c_2^*V_{3,0}+2\left(c_1^{*2}+c_2^*\right)V_{2,1}+2c_1^*V_{1,2}\right.$$
$$\left.+\left(c_1^{*2}+2c_2^*\right)V_{2,2}+\left(c_2^{*2}+2c_1^*c_3^*\right)V_{4,0}+2\left(c_3^*+2c_1^*c_2^*\right)V_{3,1}\right). \tag{2.6}$$

To have efficient estimates, we decide to use the Kadilar and Cingi (2003) definition in (2.1) and by this way we define a new family as

$$\bar{y}_K=\bar{y}_{ST}\left(\frac{\bar{x}_{st}}{\bar{X}}\right)^{\delta}\left[w+(1-w)\left(\frac{a_{st}^*+B_{st}}{A_{st}^*+B_{st}}\right)^{\varepsilon}\right]^{\eta}, \tag{2.7}$$

where $A_{st}^*=\sum_{h=1}^{L} W_h\bar{X}_hA_h$, $B_{st}=\sum_{h=1}^{L} W_hB_h$, $a_{st}^*=\sum_{h=1}^{L} W_h\bar{x}_hA_h$. Some new estimators and suggested estimators in the stratified random sampling literature, which are generated from (2.7) for different combinations of δ, ε, η, w, A_h, and B_h, are given in Table 3. To obtain the *MSE*, we can define a new *e* term as

$$e_1^*=\frac{a_{st}^*-A_{st}^*}{A_{st}^*}=\frac{\sum_{h=1}^{L} W_hA_h\left(\bar{x}_h-\bar{X}_h\right)}{A_{st}^*}. \tag{2.8}$$

Expressing the estimator, $\bar{y}_K$, in terms of e_i (i = 0, 1) and e_1^*, we can rewrite (2.7) as

$$\bar{y}_K=\bar{Y}\left(1+e_0\right)\left(1+e_1\right)^{\delta}\left[w+(1-w)\left(1+t_{st}^*e_1^*\right)^{\varepsilon}\right]^{\eta}, \tag{2.9}$$

where $t_{st}^*=\dfrac{A_{st}^*}{A_{st}^*+B_{st}}$.

Expanding terms $\left(1+e_1\right)^{\delta}$ and $\left[w+(1-w)\left(1+t_{st}^*e_1^*\right)^{\varepsilon}\right]^{\eta}$ in (2.9), we can get

$$\bar{y}_K=\bar{Y}\left(1+e_0\right)\left(a_0+a_1e_1+a_2e_1^2+a_3e_1^3+a_4e_1^4+\ldots\right)\left(b_0+b_1t_{st}^*e_1^*\right.$$
$$\left.+b_2t_{st}^{*2}e_1^{*2}+b_3t_{st}^{*3}e_1^{*3}+b_4t_{st}^{*4}e_1^{*4}+\ldots\right), \tag{2.10}$$

where the terms a_i and b_i are defined as (1.8) and (1.10), respectively. Up to the first order of approximation, we can write

$$\left(\bar{y}_K-\bar{Y}\right)^2=\bar{Y}^2\left(a_1^2e_1^2+e_0^2+b_1^2t_{st}^{*2}e_1^{*2}+2a_1e_1e_0+2a_1b_1t_{st}^*e_1e_1^*+2b_1t_{st}^*e_0e_1^*\right). \quad (2.11)$$

Considering the following equations,

$$E\left(\left(e_1\right)^s\left(e_1^*\right)^t\left(e_0\right)^r\right)=\frac{\sum_{h=1}^{L}W_h^2\left(A_h\right)^tE\left[\left(\bar{x}_h-\bar{X}_h\right)^{s+t}\left(\bar{y}_h-\bar{Y}_h\right)^r\right]}{\left(A_{st}^*\right)^t\bar{X}^s\bar{Y}^r}, \quad (2.12)$$

$$V_{s,t,r}^*=\frac{\sum_{h=1}^{L}W_h^2\left(A_h\right)^tE\left[\left(\bar{x}_h-\bar{X}_h\right)^{s+t}\left(\bar{y}_h-\bar{Y}_h\right)^r\right]}{\bar{X}^{s+t}\bar{Y}^r}, \quad (2.13)$$

we obtain the *MSE* of the estimator, $\bar{y}_K$, up to the first order of approximation, by

$$MSE_I\left(\bar{y}_K\right)=\bar{Y}^2\left(a_1^2V_{200}^*+V_{002}^*+b_1^2k_{st}^2V_{020}^*+2a_1V_{101}^*+2a_1b_1k_{st}V_{110}^*+2b_1k_{st}V_{011}^*\right), \quad (2.14)$$

where $k_{st}=\dfrac{\bar{X}}{A_{st}^*+B_{st}}$. Optimum values of a_1 and b_1 are also obtained as

$$a_{1(opt)}=\frac{-V_{101}^*V_{020}^*+V_{011}^*V_{110}^*}{V_{200}^*V_{020}^*-V_{110}^{*2}}, \qquad b_{1(opt)}=\frac{V_{101}^*V_{110}^*-V_{011}^*V_{200}^*}{k_{st}\left(V_{020}^*V_{200}^*-V_{110}^{*2}\right)}. \quad (2.15)$$

Substituting these optimum values in (2.14), we can get the minimum *MSE* of $\bar{y}_K$ as

$$MSE_{I(\min)}\left(\bar{y}_K\right)=\bar{Y}^2\left(V_{002}^*-\frac{V_{101}^{*2}V_{020}^*+V_{200}^*V_{011}^{*2}-2V_{101}^*V_{011}^*V_{110}^*}{V_{200}^*V_{020}^*-V_{110}^{*2}}\right). \quad (2.16)$$

From (2.16), we can say that various usages of A_h can affect the $MSE_{I(\min)}\left(\bar{y}_K\right)$. Up to the second order of approximation, the *MSE* equation is given by

$$\begin{aligned}MSE_{II}\left(\bar{y}_K\right)=\bar{Y}^2\Big\{&a_1^2V_{200}^*+V_{002}^*+b_1^2k_{st}^2V_{020}^*+2a_1V_{101}^*+2a_1b_1k_{st}V_{110}^*+2b_1k_{st}V_{011}^*\\&+2a_1a_2V_{300}^*+6a_1b_1k_{st}V_{111}^*+2b_1b_2k_{st}^3V_{030}^*+2\left(a_1^2+a_2\right)V_{201}^*+2\left(a_1^2b_1+a_2b_1\right)k_{st}V_{210}^*\\&+2\left(a_1b_2+a_1b_1^2\right)k_{st}^2V_{120}^*+2a_1V_{102}^*+2b_1k_{st}V_{012}^*+2\left(b_2+b_1^2\right)k_{st}^2V_{021}^*\\&+\left(a_1^2+2a_2\right)V_{202}^*+\left(a_1^2b_1^2+2a_1^2b_2+2a_2b_1^2+2a_2b_2\right)k_{st}^2V_{220}^*+\left(b_1^2+2b_2\right)k_{st}^2V_{022}^*\\&+2\left(2a_1^2b_1+3a_2b_1\right)k_{st}V_{211}^*+2\left(3a_1b_2+2a_1b_1^2\right)k_{st}^2V_{121}^*+4a_1b_1k_{st}V_{112}^*\\&+\left(a_2^2+2a_1a_3\right)V_{400}^*+\left(b_2^2+2b_1b_3\right)k_{st}^4V_{040}^*+4a_1a_2V_{301}^*+2\left(2a_1a_2b_1+a_3b_1\right)k_{st}V_{310}^*\\&+2\left(a_1b_3+2a_1b_1b_2\right)k_{st}^3V_{130}^*+2a_3V_{301}^*+2\left(b_3+2b_1b_2\right)k_{st}^3V_{031}^*\Big\}\end{aligned} \quad (2.17)$$

We find the efficiency condition for the proposed family of estimators, $\bar{y}_k$, as follows:

$$MSE_{I(\min)}\left(\bar{y}_K\right) < MSE_{I\min}\left(\bar{y}_{CST}\right) = MSE_{I(\min)}\left(\bar{y}_N\right) = MSE\left(\bar{y}_{lrc}\right)$$

$$\bar{Y}^2\left(V_{002}^* - \frac{V_{101}^{*2}V_{020}^* + V_{200}^*V_{011}^{*2} - 2V_{101}^*V_{011}^*V_{110}^*}{V_{200}^*V_{020}^* - V_{110}^{*2}}\right) < \bar{Y}^2\left(V_{0,2} - \frac{V_{1,1}^2}{V_{2,0}}\right)$$

$$\frac{V_{1,1}^2}{V_{2,0}} < \frac{V_{101}^{*2}V_{020}^* + V_{200}^*V_{011}^{*2} - 2V_{101}^*V_{011}^*V_{110}^*}{V_{200}^*V_{020}^* - V_{110}^{*2}} \tag{2.18}$$

If the condition (33) is satisfied, $\bar{y}_K$ is more efficient than $\bar{y}_{CST}$, $\bar{y}_N$, and $\bar{y}_{lrc}$.

3. NUMERICAL EXAMPLE

In this section, we use the data concerning the number of teachers as study variable and the number of students as auxiliary variable in both primary and secondary schools for 923 districts at 6 regions (as 1: Marmara 2: Agean 3: Mediterranean 4: Central Anatolia 5: Black Sea 6: East and Southeast Anatolia) in Turkey in 2007 (Source: Ministry of Education, Republic of Turkey). The summary statistics of the data are given in Table 4. We used the Neyman allocation for allocating the samples to different strata (Cochran, 1977).

Up to the first order of approximation, we compute the minimum *MSEs* of Diana (1993) and suggested family of estimators, $\bar{y}_N$, $\bar{y}_K$, using (1.13), (2.5), and (2.16), respectively. By this way, we get the minimum *MSE* of Diana (1993) and suggested family of estimators, $\bar{y}_N$, as 194.283 which is equal to the *MSE* of the combined regression estimator, whereas we get the minimum *MSE* of the suggested estimator, $\bar{y}_K$, different from the combined regression estimator. When we use A_h as kurtosis, coefficient of variation, skewness and correlation coefficient we get the minimum *MSE* values as 190.421, 182.742, and 188.445, 131.7843 respectively. From these results, we can say that the suggested family of estimators, $\bar{y}_K$, is more efficient than the estimators of Diana (1993), suggested family of estimators, $\bar{y}_N$, and the traditional combined regression estimator when we define A_h as correlation coefficient.

As mentioned in the previous sections, when a parameter is defined as a "free parameter" for the family of estimators in Diana (1993) and suggested family of estimators, $\bar{y}_N$, the minimum *MSE* of these estimators is equal to the *MSE* of the combined regression estimator. Therefore, to investigate the efficiency among subclasses of these families, we decide to compute the *MSE* up to the second order of approximation. Using (1.15), we obtain the second order *MSE* values of members of Diana (1993) estimators, given in Table 1, and these *MSE* values are shown in Table 5. From Table 5, we observe that $\bar{y}_{Gu(st)}$ is the most efficient estimator for this data set.

Similarly, using (2.6), we obtain the second order *MSE* values of members of $\bar{y}_N$ estimators, given in Table 2, with various known population parameters and these *MSE* values are shown in Table 6. From Table 6, we observe that $\bar{y}_{N2}$ is the most efficient estimator for this data set and also we note that using different known population parameters of the auxiliary information has approximately no effect on the *MSE* values.

Up to the first order of approximation, the estimators $\bar{y}^*_{st(SD)}$, $\bar{y}^*_{st(SK)}$, $\bar{y}^+_{st(SD)}$, $\bar{y}^+_{st(SD)}$, and $\bar{y}^+_{st(SK)}$ have the same minimum *MSE* which is equal to the combined regression estimator. Note that $\bar{y}^*_{st(US2)}$ and $\bar{y}_{R(\delta_{st})}$ are the most efficient estimators.

As seen in Table 7, up to the first order of approximation, some estimators have the same *MSE*. For this reason, to find the most efficient estimator among $\bar{y}_K$, we decide to find their *MSE* equations up to the second order of approximation by (2.17). Up to the second order of approximation we find that $\bar{y}^+_{st(US2)}$ is the most efficient estimator.

4. CONCLUSION

In this paper, the properties of estimators in Diana (1993) are discussed in the stratified random sampling and we propose two new families of estimators using some known population parameters of the auxiliary variable. Firstly, we use Koyuncu and Kadilar (2009) definition in Diana (1993) family of estimators and we find that up to the first order of approximation we get the minimum MSE equal to the MSE of the combined regression estimator and using auxiliary information does not affect the MSE much, whereas up to the second order of approximation it affects the MSE slightly. Secondly, we suggest one more family using the definition in Kadilar and Cingi (2003) and finally, we find that up to the first order of approximation using the auxiliary information affects the MSE enormously as seen in Table 7. When we define A_h as a correlation coefficient in $\bar{y}_K$, we get the minimum MSE as 131.784 which is quite smaller than MSE of the combined regression that is 194.283. Second family of estimators also includes some other estimators proposed by many authors in literature such as Kadilar and Cingi (2003), Shabbir and Gupta (2005), Singh *et al.* (2008). Moreover, from this family many new estimators can also be obtained. We examine the effect of various transformations of the auxiliary information on the families of estimators. We also study the second order of approximation on the proposed families of estimators and we see that the minimum *MSE* values of the $\bar{y}_K$ can change according to definition of A_h. We show that the proposed family of estimators, $\bar{y}_K$, can be more efficient than the combined regression estimator. However, the *MSE* of the proposed family of estimators, $\bar{y}_N$, is equal to the *MSE* of the combined regression estimator such as family of estimators in Diana (1993).

REFERENCES

1. Bedi, P.K. (1996). Efficient utilization of auxiliary information at estimation stage. *Biometrical Journal*, 38(8), 973-976.
2. Cochran, W.G. (1977). *Sampling Techniques*. John Wiley and Sons, New-York.
3. Diana, G. (1993). A class of estimators of the population mean in stratified random sampling. *Statistica*, 1, 59-66.
4. Diana, G. (1992). A study of k-th order approximation of some ratio type strategies, *Metron*, 1(2), 19-32.
5. Hansen, M.H., Hurwitz, W.N. and Gurney, M. (1946). Problems and methods of the sample survey of business. *J. Amer. Statist. Assoc.*, 41, 173-189.
6. Kadilar, C. and Cingi, H. (2003). Ratio estimator in stratified sampling. *Biometrical Journal*, 45(2), 218-225.
7. Kadilar, C. and Cingi, H. (2005). A new estimator in stratified random sampling. *Commun. in Statist.: Theory and Meth.*, 34, 597-602.
8. Kaur, P. (1985). On the estimation of population mean in stratified sampling. *Biometrical Journal*, 27(1), 101-105.
9. Koyuncu, N. and Kadilar, C. (2009). Ratio and product estimators in stratified random sampling*. J. Statist. Plann. and Infer.*, 139(8), 2552-2558.
10. Prasad, B. (1989). Some improved ratio type estimators of population mean and ratio in finite population sample surveys. *Commun. in Statist.: Theory and Meth.*, 18(1), 379-392.
11. Sahai, A. (1979). An efficient variant of the product and ratio estimators. *Statistica Neerlandica*, 33, 27-35.
12. Shabbir, J. and Gupta, S. (2005). Improved ratio estimators in stratified sampling. *Amer. J. Math. and Mngt. Scien.*, 25, 3-4, 293-311.
13. Shabbir, J. and Gupta, S. (2006). A new estimator of population mean in stratified sampling. *Commun. in Statist.: Theory and Meth.*, 35, 1201-1209.
14. Singh, H.P., Tailor, R., Singh, S. and Kim, J.M. (2008). A modified estimator of population mean using power transformation. *Statistical Papers*, 49, 37-58.
15. Singh, H.P. and Vishwakarma, G.K. (2006). An efficient variant of the product and ratio estimators in stratified random sampling. *Statistics in Transition*, 7(6), 1311-1325.
16. Singh, H.P. and Vishwakarma, G.K. (2008). A family of estimators of population mean using auxiliary information in stratified sampling. *Commun. in Statist.-Theory and Meth.*, 37, 1038-1050.
17. Srivastava, S.K. (1971). A generalized estimator for mean of a finite population using multi-auxiliary information. *JASA,* 66, 404-407.
18. Sukhatme, P.V., Sukhatme B.V., Sukhatme, S. and Asok, C. (1984). *Sampling theory of surveys with applications*. Iowa State University Press, Ames, Iowa, USA.

Table 1
Some members of Diana (1993) estimator

	δ	ε	η	w	$c_i \ (i=1,\ldots,k)$
$\bar{y}_{RW(st)}$	-1	-1	-1		$(-w)^i$
$\bar{y}_{Gu(st)}$	-1	-1	1		$(-1)^i\left[i(1-w)+1\right]$
$\bar{y}_{RS(st)}$	-1	1	1		$(-1)^i\, w$
$\bar{y}_{SR(st)}$	0		1	2	$\frac{\varepsilon(\varepsilon-1)\ldots(\varepsilon-i+1)}{(-1)i!}$
$\bar{y}_{TR}$	1	-1	1		$w \quad (i=1)$ $0 \quad (i=2,\ldots)$
$\bar{y}_{sr(st)}$	0	-1	1		$(-1)^i(1-w)$
$\bar{y}_{sp(st)}$	0	1	1		$1-w \quad (i=1)$ $0 \quad (i=2,\ldots)$
$\bar{y}_{rp(st)}$	-1	2	1		$1-2w \quad (i=1)$ $(-1)^i\, w \quad (i=2,\ldots)$
$\bar{y}_{MS(st)}$	1	1	-1		$w(w-1)^{i-1}$
$\bar{y}_{SM(st)}$	-1	1	-1		$(-1)^i\left[1+\sum_{j=1}^{i}(1-w)^j\right]$

Table 2:
Some members of suggested family of estimators $\bar{y}_N$

	δ	ε	η	w	c_i
$\bar{y}_{N1}$	-1	-1	-1		$c_1^* = (1-w)t_{st} - 1$
					$c_2^* = -w(1-w)t_{st}^2 - (1-w)t_{st} + 1$
					$c_3^* = \left\{-(1-w)^3 - 2*(1-w)^2 + (1-w)\right\} t_{st}^3$
					$+ w(1-w)t_{st}^2 + (1-w)t_{st} - 1$
$\bar{y}_{N2}$	-1	-1	1		$c_1^* = -(1-w)t_{st} - 1$
					$c_2^* = (1-w)t_{st}^2 + (1-w)t_{st} + 1$
					$c_3^* = -(1-w)t_{st}^3 - (1-w)t_{st}^2 - (1-w)t_{st} - 1$
$\bar{y}_{N3}$	-1	1	1		$c_1^* = (1-w)t_{st} - 1$
					$c_2^* = -(1-w)t_{st} + 1$
					$c_3^* = (1-w)t_{st} - 1$
$\bar{y}_{N4}$	0		1	2	$c_1^* = -\varepsilon t_{st}$
					$c_2^* = -\varepsilon(\varepsilon-1)t_{st}^2 / 2$
					$c_3^* = -\varepsilon(\varepsilon-1)(\varepsilon-2)t_{st}^3 / 6$
$\bar{y}_{N5}$	1	-1	1		$c_1^* = -(1-w)t_{st} + 1$
					$c_2^* = (1-w)t_{st}^2 - (1-w)t_{st}$
					$c_3^* = -(1-w)t_{st}^3 + (1-w)t_{st}^2$
$\bar{y}_{N6}$	0	-1	1		$c_1^* = -(1-w)t_{st}$
					$c_2^* = (1-w)t_{st}^2$
					$c_3^* = -(1-w)t_{st}^3$
$\bar{y}_{N7}$	0	1	1		$c_1^* = (1-w)t_{st}$ $c_2^* = 0$ $c_3^* = 0$
$\bar{y}_{N8}$	-1	2	1		$c_1^* = 2(1-w)t_{st} - 1$
					$c_2^* = (1-w)t_{st}^2 - 2(1-w)t_{st} + 1$
					$c_3^* = -(1-w)t_{st}^2 + 2(1-w)t_{st} - 1$
$\bar{y}_{N9}$	1	1	-1		$c_1^* = -(1-w)t_{st} + 1$
					$c_2^* = (1-w)^2 t_{st}^2 - (1-w)t_{st}$
					$c_3^* = -(1-w)^3 t_{st}^3 + (1-w)^2 t_{st}^2$
$\bar{y}_{N10}$	-1	1	-1		$c_1^* = -(1-w)t_{st} - 1$
					$c_2^* = (1-w)^2 t_{st}^2 + (1-w)t_{st} + 1$
					$c_3^* = -(1-w)^3 t_{st}^3 - (1-w)^2 t_{st}^2 - (1-w)t_{st} - 1$

Table 3:
Some members of suggested family of estimators $\overline{y}_K$

	δ	w	ε	η	A_h	B_h
$\overline{y}_{st(SD)}$ Kadilar and Cingi(2003)	0	0	-1	1	1	C_{xh}
$\overline{y}_{st(SK)}$ Kadilar and Cingi(2003)	0	0	-1	1	1	$\beta_{2(x)h}$
$\overline{y}_{st(US1)}$ Kadilar and Cingi(2003)	0	0	-1	1	$\beta_{2(x)h}$	C_{xh}
$\overline{y}_{st(US2)}$ Kadilar and Cingi(2003)	0	0	-1	1	C_{xh}	$\beta_{2(x)h}$
$\overline{y}^*_{st(SD)}$ Shabbir and Gupta (2005)	0	w	-1	1	1	C_{xh}
$\overline{y}^*_{st(SK)}$ Shabbir and Gupta (2005)	0	w	-1	1	1	$\beta_{2(x)h}$
$\overline{y}^*_{st(US1)}$ Shabbir and Gupta (2005)	0	w	-1	1	$\beta_{2(x)h}$	C_{xh}
$\overline{y}^*_{st(US2)}$ Shabbir and Gupta (2005)	0	w	-1	1	C_{xh}	$\beta_{2(x)h}$
$\overline{y}_{R(\alpha_{st})}$ Singh *et al.*(2008)	0	0	- ε	1	$\beta_{2(x)h}$	C_{xh}
$\overline{y}_{R(\delta_{st})}$ Singh *et al.*(2008)	0	0	- ε	1	C_{xh}	$\beta_{2(x)h}$
$\overline{y}^+_{st(SD)}$	δ	0	-1	1	1	C_{xh}
$\overline{y}^+_{st(SK)}$	δ	0	-1	1	1	$\beta_{2(x)h}$
$\overline{y}^+_{st(US1)}$	δ	0	-1	1	$\beta_{2(x)h}$	C_{xh}
$\overline{y}^+_{st(US2)}$	δ	0	-1	1	C_{xh}	$\beta_{2(x)h}$

Table 4:
Data Statistics

N_1 =127	N_2 =117	N_3 =103
N_4 =170	N_5 =205	N_6 =201
n_1 =31	n_2 =21	n_3 =29
n_4 =38	n_5 =22	n_6 =39
S_{y1} =883.835	S_{y2} =644.922	S_{y3} =1033.467
S_{y4} =810.585	S_{y5} =403.654	S_{y6} =711.723
$\bar{Y}_1$ =703.74	$\bar{Y}_2$ =413	$\bar{Y}_3$ =573.17
$\bar{Y}_4$ =424.66	$\bar{Y}_5$ =267.03	$\bar{Y}_6$ =393.84
S_{x1} =30486.751	S_{x2} =15180.769	S_{x3} =27549.697
S_{x4} =18218.931	S_{x5} =8497.776	S_{x6} =23094.141
$\bar{X}_1$ =20804.59	$\bar{X}_2$ =9211.79	$\bar{X}_3$ =14309.30
$\bar{X}_4$ =9478.85	$\bar{X}_5$ =5569.95	$\bar{X}_6$ =12997.59
S_{xy1} =25237153.52	S_{xy2} =9747942.85	S_{xy3} =28294397.04
S_{xy4} =14523885.53	S_{xy5} =3393591.75	S_{xy6} =15864573.97
ρ_1 =0.936	ρ_2 =0.996	ρ_3 =0.994
ρ_4 =0.983	ρ_5 =0.989	ρ_6 =0.965
$\beta_2(x_1)$ =4.593	$\beta_2(x_2)$ =18.543	$\beta_2(x_3)$ =15.446
$\beta_2(x_4)$ =10.162	$\beta_2(x_5)$ =21.947	$\beta_2(x_6)$ =23.114
$\beta_1(x_1)$ =2.164	$\beta_1(x_2)$ =3.867	$\beta_1(x_3)$ =3.748
$\beta_1(x_4)$ =3.121	$\beta_1(x_5)$ =4.084	$\beta_1(x_6)$ =4.411
w_1 =0.138	w_2 =0.127	w_3 =0.112
w_4 =0.184	w_5 =0.222	w_6 =0.218

Table 5:
Second Order *MSE* (in bold) of members of the family of estimators of Diana (1993) and optimum values of "*free parameter*"

	MSE II
$\overline{y}_{RW(st)}$	**190.3796**
	0.90598
$\overline{y}_{Gu(st)}$	**190.3288**
	1.09419
$\overline{y}_{RS(st)}$	**190.7910**
	0.90733
$\overline{y}_{SR(st)}$	**193.5475**
	-0.91757
$\overline{y}_{TR}$	**202.5400**
	-0.88570
$\overline{y}_{sr(st)}$	**190.7910**
	0.09267
$\overline{y}_{sp(st)}$	**202.5400**
	1.88570
$\overline{y}_{rp(st)}$	**191.0142**
	0.95436
$\overline{y}_{MS(st)}$	**211.6433**
	-0.89373
$\overline{y}_{SM(st)}$	**190.8243**
	1.09235

Table 6:
Second Order *MSE* (in bold) of members of the family of estimators of $\bar{y}_N$ and optimum values of "*free parameter*"

t_{st}	$\frac{\bar{X}\sum_{h=1}^{L} W_h C_{xh}}{\bar{X}\sum_{h=1}^{L} W_h C_{xh} + \sum_{h=1}^{L} W_h \beta_{2(x)h}}$	$\frac{\bar{X}\sum_{h=1}^{L} W_h \beta_{2(x)h}}{\bar{X}\sum_{h=1}^{L} W_h \beta_{2(x)h} + \sum_{h=1}^{L} W_h C_{xh}}$	$\frac{\bar{X}\sum_{h=1}^{L} W_h \beta_{1(x)h}}{\bar{X}\sum_{h=1}^{L} W_h \beta_{1(x)h} + \sum_{h=1}^{L} W_h \beta_{2(x)h}}$	$\frac{\bar{X}\sum_{h=1}^{L} W_h \beta_{2(x)h}}{\bar{X}\sum_{h=1}^{L} W_h \beta_{2(x)h} + \sum_{h=1}^{L} W_h \beta_{1(x)h}}$
$\bar{y}_{N1}$	**190.3846** 0.905921	**190.3842** 0.905999	**190.3844** 0.905963	**190.3842** 0.905998
$\bar{y}_{N2}$	**190.3293** 1.094264	**190.3288** 1.094186	**190.3290** 1.094222	**190.3288** 1.094187
$\bar{y}_{N3}$	**190.7910** 0.907251	**190.7910** 0.907329	**190.7910** 0.907293	**190.7910** 0.907328
$\bar{y}_{N4}$	**201.0317** 0.885785	**201.0211** 0.885059	**201.0260** 0.885394	**201.0212** 0.885068
$\bar{y}_{N5}$	**202.5875** -0.887247	**202.5405** -0.885714	**202.5622** -0.886422	**202.5411** -0.885733
$\bar{y}_{N6}$	**190.7858** 0.09191	**190.7909** 0.092662	**190.7886** 0.092315	**190.7909** 0.092653
$\bar{y}_{N7}$	**202.5400** 1.886446	**202.5400** 1.885706	**202.5400** 1.886048	**202.5400** 1.885715
$\bar{y}_{N8}$	**191.0140** 0.954316	**191.0142** 0.954355	**191.0141** 0.954337	**191.0142** 0.954354
$\bar{y}_{N9}$	**211.6433** -0.895334	**211.6433** -0.893751	**211.6433** -0.894482	**211.6433** -0.89377
$\bar{y}_{N10}$	**190.8243** 1.092443	**190.8243** 1.092366	**190.8243** 1.092402	**190.8243** 1.092367

Table 7:
First and Second Order *MSE*s (in bold) of members of the family of estimators of $\bar{y}_K$ and optimum values of "*free parameter*"

	MSE I	MSE II
$\bar{y}_{st(SD)}$	**216.349**	**217.4261**
$\bar{y}_{st(SK)}$	**215.650**	**216.7719**
$\bar{y}_{st(US1)}$	**757.561**	**1058.2967**
$\bar{y}_{st(US2)}$	**209.118**	**209.9047**
$\bar{y}^*_{st(SD)}$	**194.283** 0.0943	**186.3173** 0.0915
$\bar{y}^*_{st(SK)}$	**194.283** 0.0931	**186.3095** 0.0904
$\bar{y}^*_{st(US1)}$	**520.254** 0.2715	**637.5303** 0.3088
$\bar{y}^*_{st(US2)}$	**184.507** 0.0990	**180.7664** 0.0968
$\bar{y}_{R(\alpha_{st})}$	**520.254** 0.7285	**528.8485** 0.7032
$\bar{y}_{R(\delta_{st})}$	**184.507** 0.9010	**193.2540** 0.8784
$\bar{y}^+_{st(SD)}$	**194.283** 0.0943	**190.5817** 0.0934
$\bar{y}^+_{st(SK)}$	**194.283** 0.0930	**190.5780** 0.0921
$\bar{y}^+_{st(US1)}$	**723.930** 0.1164	**918.6896** 0.1186
$\bar{y}^+_{st(US2)}$	**186.126** 0.0964	**158.3103** 0.1337
$MSE(\bar{y}_K)_{\min}$ =**131.7842939**		
(when A_h correlation coefficient		