International
Journal
of Advanced Research in Engineering& Management (IJAREM)
ISSN: 2456-2033 || PP. 11-16

New family of estimators using two auxiliary attributes

Tolga Zaman
Department of Statistics, Çankırı Karatekin University, Çankırı, Turkey

Abstract: Utilizing the estimators in Singh and Malik (Appl. Math. Computed. 219, (2013), 10948) and Kadilar and Cingi (Appl. Math. Computed. 162 (2005), 902), It is proposed estimators using two auxiliary attributes in simple random sampling. It is obtained mean square error (MSE) equation of these estimators. Theoretically, it is compared with the MSE of proposed estimators and the MSE of traditional regression estimators using two auxiliary attributes. As a result of these comparisons, it is observed that the proposed estimators give more efficient results than the traditional regression estimators. Also, under all conditions, all of the proposed estimators more efficient than the traditional regression estimators. And, these theoretical results are supported by an application with original data.
Keywords: Ratio estimator; Regression estimator; Auxiliary attribute; Mean square error, Efficiency.

1. Introduction

Suppose that an auxiliary attribute p_i, correlated with variate of interest y_i, is obtained for each unit in the sample which in drawn by simple random sampling and that the population proportion P of the p is known. The regression estimate of $\bar{Y}$, the population mean of the y_i, is

$$t_{reg1} = \bar{y} + b(P - p) \quad (1.1)$$

where b is an estimate of the change in y when p is increased by unity. $\bar{y}$ is the sample mean of the y_i, p is the sample proportion of the p_i. MSE equation of this estimate is

$$MSE(t_{reg1}) = \frac{1-f}{n} S_y^2 (1 - \rho_{yp}^2) \quad (1.2)$$

where $f = \frac{n}{N}$, n is the sample size, N is the ppulation size, S_y^2 and S_p^2 are the population variances of the y_i and p_i respectively, $\rho_{yp} = \frac{S_{yp}}{S_y S_p}$ is the population correlation coefficient betweeny_i and p_i, S_{yp} is the population covaiance between y_i and p_i.

When there are two auxiliary attributes P_1 and P_2, the regression estimate of $\bar{Y}$ is:

$$t_{reg2} = \bar{y} + b_1(P_1 - p_1) + b_2(P_2 - p_2) \quad (1.3)$$

where $b_1 = \frac{s_{yp_1}}{s_{p_1}^2}$ and $b_2 = \frac{s_{yp_2}}{s_{p_2}^2}$. Here, $s_{p_1}^2$ and $s_{p_2}^2$ are the sample variances of the p_{1i} and p_{2i}, respectively. s_{yp_1} and s_{yp_2} are the sample covariance between y_i and p_{1i} and between y_i and p_{2i}, respectively.

MSE equation of this estimate can be found as

$$MSE(t_{reg2}) = \frac{1-f}{n}\left(S_y^2 + B_1^2 S_{p_1}^2 + B_2^2 S_{p_2}^2 - 2B_1 S_{yp_1} - 2B_2 S_{yp_2} + 2B_1 B_2 S_{p_1 p_2}\right) \quad (1.4)$$

2. The suggested estimators

Adapting Singh and Malik (2013) and Kadilar and Cingi (2005), it is proposed a multivariate ratio estimator using information on two auxiliary attributes as;

$$\bar{y}_{pri} = \bar{y}\left[\frac{m_1 P_1 + m_2 P_2}{m_1 p_1 + m_2 p_2}\right]^{\alpha} + b_1(P_1 - p_1) + b_2(P_2 - p_2) \quad (2.1)$$

where $(m_1 \neq 0)$, m_2 are either real numbers or the functions of the known parameters of the auxiliary variable such as coefficient of variation(C_p), coefficient of kurtosis $(\beta_2(\phi))$, coefficient of skewness $(\beta_1(\phi))$and coefficient of correlation (ρ_{pb}).

Table 1. Suggested Estimators

Estimators	Values of m_1	m_2
$\bar{y}_{pr1} = \bar{y}\left[\frac{\beta_2(\phi_1)P_1 + C_{p_2}P_2}{\beta_2(\phi_1)p_1 + C_{p_2}p_2}\right]^{\alpha} + b_1(P_1 - p_1) + b_2(P_2 - p_2)$	$\beta_2(\phi)$	C_p
$\bar{y}_{pr2} = \bar{y}\left[\frac{C_{p_1}P_1 + \beta_2(\phi_2)P_2}{C_{p_1}p_1 + \beta_2(\phi_2)p_2}\right]^{\alpha} + b_1(P_1 - p_1) + b_2(P_2 - p_2)$	C_p	$\beta_2(\phi)$
$\bar{y}_{pr3} = \bar{y}\left[\frac{C_{p_1}P_1 + \rho_{yp_2}P_2}{C_{p_1}p_1 + \rho_{yp_2}p_2}\right]^{\alpha} + b_1(P_1 - p_1) + b_2(P_2 - p_2)$	C_p	ρ_{pb}
$\bar{y}_{pr4} = \bar{y}\left[\frac{\rho_{yp_1}P_1 + C_{p_2}P_2}{\rho_{yp_1}p_1 + C_{p_2}p_2}\right]^{\alpha} + b_1(P_1 - p_1) + b_2(P_2 - p_2)$	ρ_{pb}	C_p
$\bar{y}_{pr5} = \bar{y}\left[\frac{\beta_2(\phi_1)P_1 + \rho_{yp_2}P_2}{\beta_2(\phi_1)p_1 + \rho_{yp_2}p_2}\right]^{\alpha} + b_1(P_1 - p_1) + b_2(P_2 - p_2)$	$\beta_2(\phi)$	ρ_{pb}
$\bar{y}_{pr6} = \bar{y}\left[\frac{\rho_{yp_1}P_1 + \beta_2(\phi_2)P_2}{\rho_{yp_1}p_1 + \beta_2(\phi_2)p_2}\right]^{\alpha} + b_1(P_1 - p_1) + b_2(P_2 - p_2)$	ρ_{pb}	$\beta_2(\phi)$
$\bar{y}_{pr7} = \bar{y}\left[\frac{\beta_1(\phi_1)P_1 + C_{p_2}P_2}{\beta_1(\phi_1)p_1 + C_{p_2}p_2}\right]^{\alpha} + b_1(P_1 - p_1) + b_2(P_2 - p_2)$	$\beta_1(\phi)$	C_p
$\bar{y}_{pr8} = \bar{y}\left[\frac{C_{p_1}P_1 + \beta_1(\phi_2)P_2}{C_{p_1}p_1 + \beta_1(\phi_2)p_2}\right]^{\alpha} + b_1(P_1 - p_1) + b_2(P_2 - p_2)$	C_p	$\beta_1(\phi)$
$\bar{y}_{pr9} = \bar{y}\left[\frac{\beta_1(\phi_1)P_1 + \rho_{yp_2}P_2}{\beta_1(\phi_1)p_1 + \rho_{yp_2}p_2}\right]^{\alpha} + b_1(P_1 - p_1) + b_2(P_2 - p_2)$	$\beta_1(\phi)$	ρ_{pb}
$\bar{y}_{pr10} = \bar{y}\left[\frac{\rho_{yp_1}P_1 + \beta_1(\phi_2)P_2}{\rho_{yp_1}p_1 + \beta_1(\phi_2)p_2}\right]^{\alpha} + b_1(P_1 - p_1) + b_2(P_2 - p_2)$	ρ_{pb}	$\beta_1(\phi)$
$\bar{y}_{pr11} = \bar{y}\left[\frac{\beta_1(\phi_1)P_1 + \beta_2(\phi_2)P_2}{\beta_1(\phi_1)p_1 + \beta_2(\phi_2)p_2}\right]^{\alpha} + b_1(P_1 - p_1) + b_2(P_2 - p_2)$	$\beta_1(\phi)$	$\beta_2(\phi)$
$\bar{y}_{pr12} = \bar{y}\left[\frac{\beta_2(\phi_1)P_1 + \beta_1(\phi_2)P_2}{\beta_2(\phi_1) + \beta_1(\phi_2)p_2}\right]^{\alpha} + b_1(P_1 - p_1) + b_2(P_2 - p_2)$	$\beta_2(\phi)$	$\beta_1(\phi)$

In Table 1, C_p, $\beta_1(\phi)$, $\beta_2(\phi)$ and ρ_{pb} are, respectively, coefficient of variation belonging to ratio of units possessing certain attributes, coefficient of population skewness and population correlation coefficient between ratio of units possessing certain attributes and study variable. $\bar{y}$ and p are sample mean belonging to study variable and sample proportionpossessing certain attributes, respectively.
MSEs of these estimators can be found using Taylor series method defined as
In general, Taylor series method for k variables can be given as;

$$h(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k) = h(\bar{X}_1, \bar{X}_2, \dots, \bar{X}_k) + \sum_{j=1}^{k} d_j\left(\bar{x}_j - \bar{X}_j\right) + R_k(\bar{X}_k, \alpha) + O_k$$

where

$$d_j = \frac{\partial h(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k)}{\partial \alpha_j}$$

and

$$R_k(\bar{X}_k, \alpha) = \sum_{j=1}^{k}\sum_{j=1}^{k} \frac{1}{2!}\frac{\partial^2 h(\bar{X}_1, \bar{X}_2, \dots, \bar{X}_k)}{\partial \bar{X}_i \bar{X}_j}\left(\bar{x}_j - \bar{X}_j\right)(\bar{x}_i - \bar{X}_i) + O_k$$

where $O_k$represents the terms in the expansion of the Taylor series of more than the second degree (Wolter, 1985). When we omit the term $R_k(\bar{X}_k, \alpha)$, we obtain Taylor series method defined as;

$$h(p_1, p_2, \bar{y}) - h(P_1, P_2, \bar{Y}) \cong \left.\frac{\partial h(p_1, p_2, \bar{y})}{\partial p_1}\right|_{P_1, P_2, \bar{Y}} (p_1 - P_1) + \left.\frac{\partial h(p_1, p_2, \bar{y})}{\partial p_2}\right|_{P_1, P_2, \bar{Y}} (p_2 - P_2) + \left.\frac{\partial h(p_1, p_2, \bar{y})}{\partial y}\right|_{P_1, P_2, \bar{Y}} (\bar{y} - \bar{Y})$$

where, $h(p_1, p_2, \bar{y}) = \bar{y}_{pr}$ and $h(P_1, P_2, \bar{Y}) = \bar{Y}$
MSE equations of the proposed estimators given in (2.1) compute as follows:

$$\bar{y}_{pr} - \bar{Y} \cong \left.\frac{\partial\left(\bar{y}\left[\frac{m_1P_1 + m_2P_2}{m_1p_1 + m_2p_2}\right]^{\alpha} + b_1(P_1 - p_1) + b_2(P_2 - p_2)\right)}{\partial p_1}\right|_{P_1, P_2, \bar{Y}} (p_1 - P_1) + \left.\frac{\partial\left(\bar{y}\left[\frac{m_1P_1 + m_2P_2}{m_1p_1 + m_2p_2}\right]^{\alpha} + b_1(P_1 - p_1) + b_2(P_2 - p_2)\right)}{\partial p_2}\right|_{P_1, P_2, \bar{Y}} (p_2 - P_2) + \left.\frac{\partial\left(\bar{y}\left[\frac{m_1P_1 + m_2P_2}{m_1p_1 + m_2p_2}\right]^{\alpha} + b_1(P_1 - p_1) + b_2(P_2 - p_2)\right)}{\partial y}\right|_{P_1, P_2, \bar{Y}} (\bar{y} - \bar{Y})$$

$$\bar{y}_{pr1} - \bar{Y} \cong \left(\frac{-\alpha\bar{Y}m_1}{m_1P_1 + m_2P_2} - b_1\right)(p_1 - P_1) + \left(\frac{-\alpha\bar{Y}m_2}{m_1P_1 + m_2P_2} - b_2\right)(p_2 - P_2) + (\bar{y} - \bar{Y})$$

$$E\left(\bar{y}_{pr} - \bar{Y}\right)^2 \cong \left(\frac{\alpha\bar{Y}m_1}{m_1P_1 + m_2P_2} + B_1\right)^2 Var(p_1) + \left(\frac{\alpha\bar{Y}m_2}{m_1P_1 + m_2P_2} + B_2\right)^2 Var(p_2) + Var(\bar{y}) + 2\left(\frac{\alpha\bar{Y}m_1}{m_1P_1 + m_2P_2} + B_1\right)\left(\frac{\alpha\bar{Y}m_2}{m_1P_1 + m_2P_2} + B_2\right)Cov(p_1, p_2) - 2\left(\frac{\alpha\bar{Y}m_1}{m_1P_1 + m_2P_2} + B_1\right)Cov(p_1, \bar{y}) - 2\left(\frac{\alpha\bar{Y}m_2}{m_1P_1 + m_2P_2} + B_2\right)Cov(p_2, \bar{y})$$

$$MSE(\bar{y}_{pr}) \cong \frac{1-f}{n}\left\{\left(\frac{\alpha\bar{Y}m_1}{m_1P_1+m_2P_2}+B_1\right)^2 S_{p_1}^2+\left(\frac{\alpha\bar{Y}m_2}{m_1P_1+m_2P_2}+B_2\right)^2 S_{p_2}^2+S_y^2\right.$$
$$+2\left(\frac{\alpha\bar{Y}m_1}{m_1P_1+m_2P_2}+B_1\right)\left(\frac{\alpha\bar{Y}m_2}{m_1P_1+m_2P_2}+B_2\right)S_{p_1p_2}-2\left(\frac{\alpha\bar{Y}m_1}{m_1P_1+m_2P_2}+B_1\right)S_{yp_1}$$
$$\left.-2\left(\frac{\alpha\bar{Y}m_2}{m_1P_1+m_2P_2}+B_2\right)S_{yp_2}\right\} \quad (2.2)$$

$$MSE(\bar{y}_{pr}) \cong \frac{1-f}{n}\left\{(\alpha R_i+B_1)^2 S_{p_1}^2+(\alpha R_j+B_2)^2 S_{p_2}^2+S_y^2+2(\alpha R_i+B_1)(\alpha R_j+B_2)S_{p_1p_2}\right.$$
$$\left.-2(\alpha R_i+B_1)S_{yp_1}-2(\alpha R_j+B_2)S_{yp_2}\right\} \quad (2.3)$$

where $B_1=\frac{S_{yp_1}}{S_{p_1}^2}$, $B_2=\frac{S_{yp_2}}{S_{p_2}^2}$

The optimum values of α to minimize the MSE of $\bar{y}_{pr}$ can easily be shown as:

$$\alpha^*=-\frac{R_iS_{yp_2}S_{p_1p_2}S_{p_1}^2+R_jS_{yp_1}S_{p_1p_2}S_{p_2}^2}{S_{p_1}^2S_{p_2}^2\left(R_i^2S_{p_1}^2+R_j^2S_{p_2}^2+2R_iR_jS_{p_1p_2}\right)} \quad (2.4)$$

The MSE expression of the estimator $\bar{y}_{pr}$ is given by:

$$MSE_{min}(\bar{y}_{pr}) \cong \frac{1-f}{n}\left\{(\alpha^*R_i+B_1)^2 S_{p_1}^2+(\alpha^*R_j+B_2)^2 S_{p_2}^2+S_y^2+2(\alpha^*R_i+B_1)(\alpha^*R_j+B_2)S_{p_1p_2}\right.$$
$$\left.-2(\alpha^*R_i+B_1)S_{yp_1}-2(\alpha^*R_j+B_2)S_{yp_2}\right\} \quad (2.5)$$

where,$R_i=\frac{\bar{Y}m_1}{m_1P_1+m_2P_2}$; $R_j=\frac{\bar{Y}m_2}{m_1P_1+m_2P_2}$ $i,j=\beta_2(\varphi), C_p, \rho_{pb}$ *and* $\beta_1(\varphi)$.

3. Efficiency Comparisons

In this section, It is compared the MSE of proposed estimators in (2.1) with the MSE of regression estimators in (1.3). As I obtain the following condition by these comparisons

$$MSE_{min}(\bar{y}_{pr})<MSE(t_{reg2})$$

$$S_{p_1}^2\left(\alpha^{*2}R_i^2+2\alpha^*R_iB_1\right)+S_{p_2}^2\left(\alpha^{*2}R_j^2+2\alpha^*R_jB_2\right)+2S_{p_1p_2}\left(\alpha^{*2}R_iR_j+\alpha^*R_iB_2+\alpha^*R_jB_1\right)-2S_{yp_1}\alpha^*R_i-2S_{yp_2}\alpha^*R_j<0 \quad (3.1)$$

When this condition is satisfied, all of the proposed estimators will be more efficient than the regression estimators in Section 1.

4. Numerical illustration

It is used the original data set in order to compare the efficiencies between the proposed estimators and other some known estimators based on MSE equations.

Population: (Source : Singh and Chaudhary, p.177)

The population consists of 34 wheat farms in 34 villages in certain region of India. The variables are defined as;

Y =area under wheat crop (in acres) during 1964,

P_1 = proportion of farms under wheat crop which have more than 750 acres land during 1961,

P_2 = proportion of farms under wheat crop which have more than 200 acres land during 1963.

For this data, we have,

Table 2. Data Statistics for Population

$N=34$	$\rho_{yp_1}=0.675$	$S_{p_2}=0.4996$	$S_{yp_1}=51.3619$
$n=10$	$\rho_{yp_2}=0.831$	$B_1=200.0972$	$S_{yp_2}=62.3280$
$\bar{Y}=199.4$	$\rho_{p_1p_2}=0.648$	$B_2=249.7571$	$S_{p_1p_2}=0.1640$
$P_1=0.4706$	$S_y=150.215$	$\beta_1(\varphi_1)=0.1234$	$C_{p_1}=1.0766$
$P_2=0.4118$	$S_{p_1}=0.5066$	$\beta_1(\varphi_2)=0.3753$	$C_{p_2}=1.2132$

By using simple random sampling assuming, we assume data $n = 10$ sample size (Cochran, 1977). Here, coefficients of correlation between study variable and auxiliary information for data set is $\rho_{yp_1} = 0.675$, $\rho_{yp_2} = 0.831$ and $\rho_{p_1p_2} = 0.648$.
Using table 3, R_i, R_j and α^* values compute as given in Table 3. Also, Table 3 denotes the efficiency condition results in (3.1).

Table 3. Efficiency condition results

m_1	m_2	R_i	R_j	α^*	**Condition Results in (3.1)**
$\beta_2(\varphi_1)$	C_{p_2}	723.3905	-489.077	-0.17416	-2365.115
C_{p_1}	$\beta_2(\varphi_2)$	-696.286	1280.113	-0.05595	-754.6878
C_{p_1}	ρ_{yp_2}	253.0147	195.1975	-0.39793	-6672.639
ρ_{yp_1}	C_{p_2}	164.7187	296.1072	-0.36713	-6044.215
$\beta_2(\varphi_1)$	ρ_{yp_2}	646.0289	-253.962	-0.26115	-4733.71
ρ_{yp_1}	$\beta_2(\varphi_2)$	-270.59	793.6025	-0.14175	-2120.332
$\beta_1(\varphi_1)$	C_{p_2}	44.1232	433.9306	-0.29843	-4788.74
C_{p_1}	$\beta_1(\varphi_2)$	324.7486	113.2159	-0.40201	-6840.723
$\beta_1(\varphi_1)$	ρ_{yp_2}	61.49918	414.0724	-0.30908	-4978.242
ρ_{yp_1}	$\beta_1(\varphi_1)$	285.0826	158.5484	-0.40228	-6790.111
$\beta_1(\varphi_1)$	$\beta_2(\varphi_2)$	-32.5026	521.5029	-0.25218	-3979.736
$\beta_2(\varphi_1)$	$\beta_1(\varphi_2)$	501.8131	-89.1435	-0.33934	-5981.984

It is calculated MSE values of traditional estimators and proposed estimators. Relative efficiency values of proposed estimators with respect to traditional estimator and the MSE values are calculated as below based on Equation (4.1) and are given in Table 4.

$$RE(\bar{y}_{pri}) = \frac{MSE(t_{reg2})}{MSE(\bar{y}_{pri})}; i = 1,2, \dots, 12 \quad (4.1)$$

Table 4. MSE values of the Ratio Estimators

Estimator	**MSE**	**Efficiency**
t_{reg1}	867.331	1.067
t_{reg2}	925.528	1
$\bar{y}_{pr1}$	758.579	1.22
$\bar{y}_{pr2}$	872.256	1.061
$\bar{y}_{pr3}$	454.519	2.036
$\bar{y}_{pr4}$	498.878	1.855
$\bar{y}_{pr5}$	591.384	1.565
$\bar{y}_{pr6}$	775.858	1.193
$\bar{y}_{pr7}$	587.5	1.575
$\bar{y}_{pr8}$	442.654	2.091
$\bar{y}_{pr9}$	574.123	1.612
$\bar{y}_{pr10}$	446.227	2.074
$\bar{y}_{pr11}$	644.606	1.436
$\bar{y}_{pr12}$	503.271	1.839

When it is examined these relative efficiency values, as we can see that efficiencies of all of the suggested estimators to traditional regression estimators are bigger than 1. That is, the suggested estimators are more efficient than traditional regression estimators. The results is not surprising because the condition (3.1) is satisfied as in Table 4. Based on these results, we noticed that the suggested estimator $\bar{y}_{pr10}$ has the highest efficiency.

5. Conclusions

It has developed new estimators, which is found more efficient than the traditional regression estimators using two auxiliary attributes for the condition (3.1). These theoretical inferences are also satisfied by the result of an application with original data. In the forthcoming studies, we hope to extent the estimator presented here for the development of a new estimators in two-stage sampling.

References

[1]. Malik S. and Singh R., An improved estimator using two auxiliary attributes. *Applied mathematics and computation*, 2013; ***219*****(23)**: 10983-10986. DOI: 10.1016/j.amc.2013.05.014.

[2]. Kadilar C.and Cingi H., A new estimator using two auxiliary variables. *Applied Mathematics and Computation*, 2005; ***162*****(2)**: 901-908.DOI.org/10.1016/j.amc.2003.12.130.

[3]. Wolter K.M.,Introduction to Variance Estimation, Springer-Verlag 1985.

[4]. Singh D. and Chaudhary F.S.,Theory and Analysis of Sample Survey Designs. John Wiley and Sons, NewYork, 1986.

[5]. Cochran W. G.,*Sampling Techniques: 3d Ed*. New York: Wiley, 1977.