On Improved Estimation of Population Mean using Qualitative Auxiliary Information

SUBHASH KUMAR YADAV¹, A.A. ADEWARA²
1. Department of Mathematics and Statistics (A Centre of Excellence), Dr. RML Avadh University, Faizabad-224001, U.P., INDIA
2. Department of Statistics, University of Ilorin, Ilorin. Kwara State, Nigeria
* Corresponding author: drskystats@gmail.com

Abstract
This paper deals with the estimation of population mean of the variable under study by improved ratio-product type exponential estimator using qualitative auxiliary information. The expression for the bias and mean squared error (MSE) of the proposed estimators has been derived to the first order of approximation. A comparative approach has been adopted to study the efficiency of proposed and previous estimators. The present estimators provide us significant improvement over previous estimators leading to the better perspective of application in various applied areas. The numerical demonstration has been presented to elucidate the novelty of paper.

Keywords: Exponential estimator, auxiliary attribute, Proportion, bias, mean squared error, efficiency.

Mathematics Subject Classification 2010: 62D05

1. Introduction
The use of supplementary (auxiliary) information has been widely discussed in sampling theory. Auxiliary variables are in use in survey sampling to obtain improved sampling designs and to achieve higher precision in the estimates of some population parameters such as the mean or the variance of the variable under study. This information may be used at both the stage of designing (leading for instance, to stratification, systematic or probability proportional to size sampling designs) and estimation stage. It is well established that when the auxiliary information is to be used at the estimation stage, the ratio, product and regression methods of estimation are widely used in many situations.

The estimation of the population mean is a burning issue in sampling theory and many efforts have been made to improve the precision of the estimates. In survey sampling literature, a great variety of techniques for using auxiliary information by means of ratio, product and regression methods has been used. Particularly, in the presence of multi-auxiliary variables, a wide variety of estimators have been proposed, following different ideas, and linking together ratio, product or regression estimators, each one exploiting the variables one at a time.

The first attempt was made by Cochran (1940) to investigate the problem of estimation of population mean when auxiliary variables are present and he proposed the usual ratio estimator of population mean. Robson (1957) and Murthy (1964) worked out independently on usual product estimator of population mean. Olkin (1958) also used auxiliary variables to estimate population mean of variable under study. He considered the linear combination of ratio estimators based on each auxiliary variable separately making use of information related to the supplementary characteristics having positive correlation with the variable under consideration.

Singh (1967a) dwelt upon a multivariate expression of Murthy’s (1964) product estimator. Further, the multi-auxiliary variables through a linear combination of single difference estimators were attempted by Raj (1965). In next bid of investigation Singh (1967b) extended the ratio-cum-product estimators to multi-supplementary variables. An innovative idea of weighted sum of single ratio and product estimators leading to multivariate was developed by Rao and Mudholkar (1967). Much versatile effort was made by John (1969) by considering a general ratio-type Estimator that, in turn, presented a unified class of estimators obtaining various particular estimators suggested by previous authors such as Olkin’s (1958) and Singh’s (1967a). Srivastava (1971) dealt with a general ratio-type estimator unifying previously developed estimators by eminent authors engaged in this area of investigation. Searls (1964) and Sisodia & Dwivedi (1981) used coefficient of variation of study and auxiliary variables respectively to estimate population mean of study variable.

et.al (2012), Onyeka (2012) etc have proposed many estimators utilizing auxiliary information. In the present study, we suggest a new estimator for estimating population mean of the variable under study.

The use of auxiliary information may increase the precision of an estimator when study variable $Y$ is highly correlated with auxiliary variable $X$. When the variable under study $Y$ is highly positively correlated with the auxiliary variable $X$, then the ratio type estimators are used to estimate the population parameter and product estimators are used when the variable under study $Y$ is highly negatively correlated with the auxiliary variable $X$ for improved estimation of parameters of variable under study. However there are situations when information on auxiliary variable is not available in quantitative form but in practice, the information regarding the population proportion possessing certain attribute $\psi$ is easily available (see Jhajj et.al. [7]), which is highly associated with the study variable $Y$. For example (i) $Y$ may be the use of drugs and $\psi$ may be the gender (ii) $Y$ may be the production of a crop and $\psi$ may be the particular variety. (iii) $Y$ may be the amount of milk produced and $\psi$ a particular breed of cow. (iv) $Y$ may be the yield of wheat crop and $\psi$ a particular variety of wheat etc. (see Shabbir and Gupta [25]).

Let there be $N$ units in the population. Let $(y_i, \psi_i), i = 1, 2, \ldots, N$ be the corresponding observation values of the $i^{th}$ unit of the population of the study variable $Y$ and the auxiliary variable $\psi$ respectively. Further we assume that $\psi_i = 1$ and $\psi_i = 0$, $i = 1, 2, \ldots, N$ if it possesses a particular characteristic or does not possess it. Let $A = \sum_{i=1}^{N} \psi_i$ and $a = \sum_{i=1}^{n} \psi_i$ denote the total number of units in the population and sample respectively possessing the attribute $\psi$. Let $p = \frac{A}{N}$ and $p = \frac{a}{n}$ denote the proportion of units in the population and sample respectively possessing the attribute $\psi$. Let a simple random sample of size $n$ from this population is taken without replacement having sample values $(y_i, \psi_i), i = 1, 2, \ldots, n$.

Naik and Gupta [15] defined the following ratio and product estimators of population mean when the prior information of population proportion of units, possessing the same attribute $\psi$ is available as

\begin{equation}
\hat{t}_1 = \frac{\bar{y}(\frac{P}{p})}{p} \tag{1.1}
\end{equation}

\begin{equation}
\hat{t}_2 = \frac{\bar{y}(\frac{P}{p})}{p} \tag{1.2}
\end{equation}

The MSE of estimators $\hat{t}_1$ and $\hat{t}_2$ up to the first order of approximation are

\begin{equation}
MSE(t_1) = f\bar{Y}^2[C_y^2 + C_{\psi}^2(1 - 2K_p)] \tag{1.3}
\end{equation}

\begin{equation}
MSE(t_2) = f\bar{Y}^2[C_y^2 + C_{\psi}^2(1 + 2K_p)] \tag{1.4}
\end{equation}

Singh et.al [27] defined the following ratio, product and ratio & product estimators respectively of population mean using qualitative auxiliary information as

\begin{equation}
\hat{t}_3 = \bar{y}\exp\left(\frac{p - P}{p + p}\right) \tag{1.5}
\end{equation}

\begin{equation}
\hat{t}_4 = \bar{y}\exp\left(\frac{p - P}{p - p}\right) \tag{1.6}
\end{equation}

\begin{equation}
\hat{t}_5 = \bar{y}\left[\alpha\exp\left(\frac{p - P}{p + p}\right) + (1 - \alpha)\exp\left(\frac{p - P}{p - p}\right)\right] \tag{1.7}
\end{equation}

Where $\alpha$ is a real constant to be determined such that the MSE of $\hat{t}_5$ is minimum. For $\alpha = 1$, $\hat{t}_5$ reduces to the
estimator $t_3$ and for $\alpha = 0$, it reduces to the estimator $t_4$. The MSE of the estimators $t_3$, $t_4$ and $t_5$ up to the first order of approximation are respectively as

$$MSE(t_3) = f \bar{Y}^2[C_y^2 + C_p^2 \left(\frac{1}{4} - K_p\right)]$$

(1.8)

$$MSE(t_4) = f \bar{Y}^2[C_y^2 + C_p^2 \left(\frac{1}{4} + K_p\right)]$$

(1.9)

$$MSE(t_5) = f \bar{Y}^2[C_y^2 + C_p^2 \left(\frac{1}{4} + \alpha^2 - \alpha + 2\rho_{pb} C_y C_p \left(\frac{1}{2} - \alpha\right)\right)]$$

(1.10)

which is minimum for optimum value of $\alpha$ as

$$\alpha = \frac{2K_p + 1}{2} = \alpha_0 \text{ (say)}$$

and the minimum MSE of $t_5$ is

$$MSE_{\text{min}}(t_5) = f \bar{Y}^2C_y^2(1 - \rho_{pb}^2) = M(t_5)_{\text{opt}}$$

which is same as that of traditional linear regression estimator, where

$$C_y^2 = \frac{S_y^2}{\bar{Y}^2}, \quad C_p^2 = \frac{S_p^2}{P^2}, \quad K_p = \frac{C_{y,p}}{C_p}, \quad \rho_{pb} = \frac{C_y}{C_p}, \quad S_y^2 = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \bar{Y})^2,$$

$$S_p^2 = \frac{1}{N}\sum_{i=1}^{N}(\psi_i - P)^2, \quad S_{xy}^2 = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \bar{Y})(\psi_i - P), \quad f = \frac{1}{n} - \frac{1}{N},$$

$$\rho_{pb} = \frac{S_{xy}}{S_y S_p}$$

is the point biserial correlation coefficient.

2. Suggested Estimators

Motivated by Prasad [19] and Gandge et al. [6], we propose

The exponential ratio type estimator as

$$\xi_1 = k\bar{Y}\exp\left(\frac{P - P}{P + P}\right)$$

(2.1)

The exponential product type estimator as

$$\xi_2 = k\bar{Y}\exp\left(\frac{P - P}{P - P}\right)$$

(2.2)

The exponential dual to ratio type estimator as

$$\xi_3 = \bar{Y}\exp\left(\frac{P - P}{P + P}\right)$$

(2.3)

$$\psi_i^* = \frac{(NP - n\psi_i)}{(N - n)}, \quad \psi_i^* = (1 + g)P - g\psi_i, \quad i = 1, 2, ..., N,$$

where $p^* = (1 + g)P - gp$ where $g = \frac{n}{N - n}$. The exponential ratio and dual to ratio type estimator as
\[ \xi_4 = \bar{Y} \left[ \alpha \exp \left( \frac{P - p}{P + p} \right) + (1 - \alpha) \exp \left( \frac{p^* - P}{p^* + P} \right) \right] \]  

(2.4)

where \( \alpha \) is a real constant to be determined such that the MSE of \( \xi_4 \) is minimum. For \( \alpha = 1 \), \( \xi_4 \) reduces to the estimator \( \xi_3 \) and for \( \alpha = 0 \), it reduces to the estimator \( \xi_3^* \).

To obtain the bias and mean squared error (MSE) of the estimators, let \( \bar{Y} = \bar{Y}(1 + e_0) \) and \( p = P(1 + e_1) \) such that \( E(e_i) = 0 \), \( i = 0,1 \) and

\[ E(e_0^2) = fC_y^2, \quad E(e_1^2) = fC_p^2 \quad \text{and} \quad E(e_0e_1) = fC_{yp} = fp_{pb}C_yC_p \]

From (2.1) by putting the values of \( \bar{Y} \) and \( p \), we have

\[ \xi_1 = k\bar{Y}(1 + e_0) \exp \left( \frac{P - P(1 + e_1)}{P + P(1 + e_1)} \right) \]

\[ = k\bar{Y}(1 + e_0) \exp \left( -\frac{e_1}{2 + e_1} \right) \]

\[ \xi_1 = k\bar{Y}(1 + e_0) \exp \left( -\frac{e_1}{2} \right) \]  

(2.5)

Expanding the right hand side of (2.5) and retaining terms up to second powers of \( e \)'s, and then subtracting \( \bar{Y} \) from both sides, we have

\[ \xi_1 - \bar{Y} = k\bar{Y} \left[ 1 + e_0 - \frac{e_1}{2} + \frac{e_1^2}{8} - \frac{e_0e_1}{2} \right] - \bar{Y} \]  

(2.6)

Taking expectations on both sides of (2.6), we get the bias of the estimator \( \xi_1 \) up to the first order of approximation, as

\[ B(\xi_1) = k\bar{Y}f \frac{C_p^2}{2} \left( \frac{1}{4} - K_p \right) + \bar{Y}(k - 1) \]  

(2.7)

Squaring both sides of equation (2.6) gives

\[ (\xi_1 - \bar{Y})^2 = k^2\bar{Y}^2 \left[ 1 + e_0 - \frac{e_1}{2} + \frac{e_1^2}{8} - \frac{e_0e_1}{2} \right]^2 + \bar{Y}^2 - 2k\bar{Y} \left[ 1 + e_0 - \frac{e_1}{2} + \frac{e_1^2}{8} - \frac{e_0e_1}{2} \right] \]

and now taking expectation, we get the MSE of the estimator \( \xi_1 \), to the first order of approximation as

\[ \text{MSE}(\xi_1) = f\bar{Y}^2[k^2fC_y^2 + (2k^2 - k)f\left( \frac{C_p^2}{4} - \rho_{pb}C_yC_p \right) + (k - 1)^2] \]  

(2.8)

The minimum of \( \text{MSE}(\xi_1) \) is obtained for the optimal value of \( k \), which is

\[ k_{opt} = \frac{A_1}{2B_1} \]  

(2.9)

where

\[ A_1 = f\left( \frac{C_p^2}{4} - \rho_{pb}C_yC_p \right) + 2 \quad \text{and} \quad B_1 = fC_y^2 + 2f\left( \frac{C_p^2}{4} - \rho_{pb}C_yC_p \right) + 1 \]

Thus the minimum MSE of the estimator \( \xi_1 \) is obtained as
\[ MSE_{\min}(\xi_1) = \overline{Y}^2 \left[ 1 - \frac{A_1^2}{4B_1} \right] \] (2.10)

Following the same procedure as above, we get the bias and MSE of the estimator \( \xi_2 \) up to the first order of approximation, as

\[ B(\xi_2) = k \overline{Y} f \left( \frac{C_p^2}{2} \left( \frac{1}{4} + K_p \right) \right) + \overline{Y} (k - 1) \] (2.11)

\[ MSE(\xi_2) = f \overline{Y}^2 \left[ k^2 fC_y^2 + (2k^2 - k) f \left( \frac{C_p^2}{4} + \rho \rho C_y C_p \right) + (k - 1)^2 \right] \] (2.12)

The minimum MSE of the estimator \( \xi_2 \) is obtained for optimal value of \( k \) as

\[ MSE_{\min}(\xi_2) = \overline{Y}^2 \left[ 1 - \frac{A_2^2}{4B_2} \right] \] (2.13)

where the optimal value of \( k \) for the estimator \( \xi_2 \) is

\[ k_{opt} = \frac{A_2}{2B_2} \]

\[ A_2 = f \left( \frac{C_y^2}{4} + \rho \rho C_y C_p \right) + 2 \quad \text{and} \quad B_2 = fC_y^2 + 2f \left( \frac{C_y^2}{4} + \rho \rho C_y C_p \right) + 1 \]

Expressing (2.3) in terms of e's, we have

\[ \xi_3 = \overline{Y} (1 + e_0) \exp \left[ \frac{g_{\epsilon_1}}{2} \right] \] (2.14)

Expanding the right hand side of (2.14) and retaining terms up to second powers of e's, and then subtracting \( \overline{Y} \) from both sides, we have

\[ \xi_3 - \overline{Y} = \overline{Y} \left[ 1 + e_0 + \frac{g_{\epsilon_1}}{2} + \frac{g_{\epsilon_1}^2}{8} + \frac{g_{\epsilon_1} e_1}{2} \right] - \overline{Y} \] (2.15)

Taking expectations on both sides of (2.15), we get the bias of the estimator \( \xi_3 \) up to the first order of approximation, as

\[ B(\xi_3) = \overline{Y} f g \left( \frac{C_p^2}{2} \left( \frac{1}{4} + K_p \right) \right) \] (2.16)

From equation (2.15), we have

\[ (\xi_3 - \overline{Y}) \cong \overline{Y} (e_0 + \frac{g_{\epsilon_1}}{2}) \] (2.17)

Squaring both sides of (2.17) and then taking expectations, we get MSE of the estimator \( \xi_3 \), up to the first order of approximation as

\[ MSE(\xi_3) = f \overline{Y}^2 \left[ C_y^2 + gC_p^2 \left( \frac{g}{4} + K_p \right) \right] \] (2.18)

Expressing (2.4) in terms of e's, we have

\[ \xi_4 = \overline{Y} (1 + e_0) \left[ \alpha \exp \left( -\frac{e_1}{2} \right) + (1 - \alpha) \exp \left( \frac{g_{\epsilon_1}}{2} \right) \right] \] (2.19)

Expanding the right hand side of (2.19) and retaining terms up to second powers of e's, and then subtracting \( \overline{Y} \) from both sides, we have
\[
\xi_4 - \bar{Y} = \bar{Y} \left[ 1 + e_0 + \alpha_1 \frac{e_1}{2} + \alpha_2 \frac{e_1^2}{8} + \alpha_1 \frac{e_0 e_1}{2} \right] - \bar{Y}
\]

(2.20)

Where \( \alpha_1 = [g - (1 + g)\alpha] \) and \( \alpha_2 = [g^2 + (1 - g^2)\alpha] \).

Taking expectations on both sides of (2.20), we get the bias of the estimator \( \xi_4 \) up to the first order of approximation, as

\[
B(\xi_4) = f\bar{Y} \left[ \alpha_2 \frac{C_y^2}{8} + \frac{\alpha_1}{2} \rho_{pb} C_y C_p \right]
\]

(2.21)

From equation (2.20), we have

\[
(\xi_4 - \bar{Y}) \cong \bar{Y} (e_0 + \alpha_1 \frac{e_1}{2})
\]

(2.22)

Squaring both sides of equation (2.22) gives

\[
(\xi_4 - \bar{Y})^2 = \bar{Y}^2 \left[ e_0^2 + \alpha_1^2 \frac{e_1^2}{4} + \alpha_1 e_0 e_1 \right]
\]

(2.23)

and now taking expectation, we get the MSE of the estimator \( \xi_4 \), to the first order of approximation as

\[
MSE(\xi_4) = f\bar{Y}^2 \left[ C_y^2 + \alpha_1^2 \frac{C_y^2}{4} + \alpha_1 \rho_{pb} C_y C_p \right]
\]

(2.24)

which is minimum for optimum value of \( \alpha \) as

\[
\alpha = \frac{2K_p + g}{1 + g} = \alpha_{opt} \text{ (say)}
\]

and the minimum MSE of \( \xi_4 \) is

\[
MSE_{\text{min}}(\xi_4) = f\bar{Y}^2 C_y^2 (1 - \rho_{pb}^2) = M(\xi_4)_{opt}
\]

(2.25)

which is same as that of traditional linear regression estimator and also equal to \( (t_2)_{opt} \).

3. Efficiency comparisons

Following are the conditions for which the proposed estimator \( \xi_4 \) is better than the \( t \) and \( \xi \) estimators.

We know that the variance of the sample mean \( \bar{Y} \) is

\[
V(\bar{Y}) = f\bar{Y}^2 C_y^2
\]

(3.1)

To compare the efficiency of the proposed estimator \( \xi_4 \) with the existing and proposed estimators, from (3.1) and (1.3), (1.4), (1.8), (1.9), (2.10), (2.13) and (2.18), we have

\[
V(\bar{Y}) - M(\xi_4)_{opt} = \rho_{pb}^2 \geq 0
\]

(3.2)

\[
MSE(t_1) - M(\xi_4)_{opt} = (C_p - \rho_{pb} C_y)^2 \geq 0
\]

(3.3)

\[
MSE(t_2) - M(\xi_4)_{opt} = (C_p + \rho_{pb} C_y)^2 \geq 0
\]

(3.4)

\[
MSE(t_1) - M(\xi_4)_{opt} = \frac{C_p}{2} - \rho_{pb} C_y \geq 0
\]

(3.5)

\[
MSE(t_2) - M(\xi_4)_{opt} = \frac{C_p}{2} + \rho_{pb} C_y \geq 0
\]

(3.6)
\[ M(\xi_4)_{opt} < \text{MSE}(\xi_1) \quad \text{if} \quad f(1 - \rho^2) < \left(1 - \frac{A_i^2}{4B_1}\right) \]  
\[ M(\xi_4)_{opt} < \text{MSE}(\xi_2) \quad \text{if} \quad f(1 - \rho^2) < \left(1 - \frac{A_2^2}{4B_2}\right) \]  
\[ M(\xi_4)_{opt} < \text{MSE}(\xi_3) \quad \text{if} \quad gC_p + 8\rho_p C_y < 0 \]

4. Empirical Study

To analyze the performance of various estimators of population mean \( \bar{Y} \) of study variable \( y \), we considered the following two data sets

**Data 1.** [Source: Sukhatme & Sukhatme [33], page 305]

\( Y = \) Area (in acres) under wheat crop in the circles and \( \psi = \) A circle consisting more than five villages.

\( N = 89, \quad n = 23, \quad \bar{Y} = 3.360, \quad P = 0.1236, \quad \rho_{pb} = 0.766, \quad C_y = 0.60400, \quad C_p = 2.19012 \)

**Data 2.** [Source: Mukhopadhyay [12], page 44]

\( Y = \) Household size and \( \psi = \) A household that availed an agricultural loan from a bank.

\( N = 25, \quad n = 7, \quad \bar{Y} = 9.44, \quad P = 0.400, \quad \rho_{pb} = -0.387, \quad C_y = 0.17028, \quad C_p = 1.27478 \)

The percent relative efficiency (PRE) of the estimators \( \bar{Y} \), proposed and mentioned existing estimators and \( (\xi_4)_{opt} \) with respect to \( \bar{Y} \) usual unbiased estimator have been computed and given in table1.

<table>
<thead>
<tr>
<th>Estimator</th>
<th>PRE of ( \bar{Y} ) with respect to ( \bar{Y} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{Y} )</td>
<td>100.00</td>
</tr>
<tr>
<td>( t_1 )</td>
<td>11.63</td>
</tr>
<tr>
<td>( t_2 )</td>
<td>5.07</td>
</tr>
<tr>
<td>( t_3 )</td>
<td>66.25</td>
</tr>
<tr>
<td>( t_4 )</td>
<td>14.16</td>
</tr>
<tr>
<td>( \xi_1 )</td>
<td>67.84</td>
</tr>
<tr>
<td>( \xi_2 )</td>
<td>16.58</td>
</tr>
<tr>
<td>( \xi_3 )</td>
<td>42.11</td>
</tr>
<tr>
<td>( (\xi_4)_{opt} )</td>
<td>241.99</td>
</tr>
<tr>
<td>( (t_5)_{opt} )</td>
<td></td>
</tr>
</tbody>
</table>

5. Conclusion

From Table1 we see that the proposed estimator \( \xi_4 \) under optimum condition performs better than the usual sample mean estimator \( \bar{Y} \), Naik and Gupta estimators \( (t_1 \) and \( t_2) \), Singh et.al estimators \( (t_3 \) and \( t_4) \), proposed estimators \( (\xi_1, \xi_2 \) and \( \xi_3) \). The \( \xi \) estimators are better than the corresponding \( t \) estimators and under optimum conditions \( \xi_4 \) and \( t_5 \) both are equally efficient.

References


Optimization sciences, Vol.XII, I, 159-163.


This academic article was published by The International Institute for Science, Technology and Education (IISTE). The IISTE is a pioneer in the Open Access Publishing service based in the U.S. and Europe. The aim of the institute is Accelerating Global Knowledge Sharing.

More information about the publisher can be found in the IISTE’s homepage: http://www.iiste.org

CALL FOR JOURNAL PAPERS

The IISTE is currently hosting more than 30 peer-reviewed academic journals and collaborating with academic institutions around the world. There’s no deadline for submission. Prospective authors of IISTE journals can find the submission instruction on the following page: http://www.iiste.org/journals/ The IISTE editorial team promises to the review and publish all the qualified submissions in a fast manner. All the journals articles are available online to the readers all over the world without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. Printed version of the journals is also available upon request of readers and authors.

MORE RESOURCES

Book publication information: http://www.iiste.org/book/
Recent conferences: http://www.iiste.org/conference/

IISTE Knowledge Sharing Partners

EBSCO, Index Copernicus, Ulrich's Periodicals Directory, JournalTOCS, PKP Open Archives Harvester, Bielefeld Academic Search Engine, Elektronische Zeitschriftenbibliothek EZB, Open J-Gate, OCLC WorldCat, Universe Digital Library, NewJour, Google Scholar