An Improvement in Regression Estimator Through Exponential Estimator Using Two Auxiliary Variables

Sachin Malik*, Kanika and Atul

Department of Mathematics, SRM University Delhi NCR, Sonipat-131029, India
E-mail: sachsen.malik@srmuniversity.ac.in; kanikasehrawat2814@gmail.com;
ak4422882@gmail.com
*Corresponding Author

Received 29 May 2023; Accepted 21 August 2023; Publication 16 March 2024

Abstract
For the case of simple random sampling, we are introducing a new regression estimator for the population mean with the supporting values of two auxiliary variables. The results for the mean square error (MSE) of the new form of regression estimator is fined. The mean square error’s results have also been checked through numerical illustration. It is observed that our introduced estimator is having less mean square error than the traditional ratio and regression estimator for two auxiliary variables.

Keywords: Mean square error (MSE), auxiliary variables, ratio estimator and regression estimator.

1 Introduction
In statistics, we are interested to know the behavior of the population based on a sample. Sample results cannot be accurate as the population results.
Every time there is a difference between the results of sample and the results of population. Throughout this paper, we are trying to minimize this mean square error (MSE). For this we have proposed an estimator using auxiliary information for two variables. Our target is to find mean square error of some estimators which is already given in literature, is always more than our introduced estimator.

The auxiliary variable’s information is effectively used to increase the efficiency of the estimators of the population mean. In the Cochran [1] and Murthy [2], ratio estimators, product estimators, and regression estimators are used in several conditions.

Abu-Dayyeh et al. [3] suggested an estimator when population means of both the auxiliary variables $X$ and $Z$ are known

$$t_1 = \bar{Y} \left( \frac{X}{\bar{X}} \right)^{\alpha_1} \left( \frac{Z}{\bar{Z}} \right)^{\alpha_2}$$  \hspace{1cm} (1)

Where $\alpha_1$ and $\alpha_2$ are constants.

Muhammad Noor-ul-Amin et al. [4] proposed estimator as

The regression estimator for two auxiliary variables is defined as

$$t_2 = \bar{Y} + b_1(X - \bar{X}) + b_2(Z - \bar{Z})$$  \hspace{1cm} (2)

Kadilar and Cingi [5] is defined a new estimator for auxiliary variables as

$$t_3 = \bar{Y} \left( \frac{X}{\bar{X}} \right)^{\alpha_1} \left( \frac{Z}{\bar{Z}} \right)^{\alpha_2} + b_1(X - \bar{X}) + b_2(Z - \bar{Z})$$  \hspace{1cm} (3)

The mean squared error of $t_1$, $t_2$, and $t_3$ after taking first order of approximation is hereby

$$\text{MSE}(t_1) = \bar{Y}^2 \left( \frac{1 - f}{n} \right) \left[ C_x^2 + C_y^2 + C_z^2 - 2\rho_{xy}C_xC_y ight. \\
- 2\rho_{yz}C_yC_z + 2\rho_{xz}C_xC_z \right]$$  \hspace{1cm} (4)

$$\text{MSE}(t_2) = \left( \frac{1 - f}{n} \right) \left[ \bar{Y}^2C_y^2 + B_1^2X^2C_x^2 + B_2^2Z^2C_z^2 + 2B_1B_2\bar{XZ}\rho_{xz}C_xC_z \right. \\
- 2\bar{Y}B_1\rho_{xy}C_xC_y - 2\bar{Y}ZB_2\rho_{yz}C_yC_z \left. \right]$$  \hspace{1cm} (5)
MSE\( (t_3) = \left(\frac{1-f}{n}\right) \left[\alpha_1^2(\bar{Y}^2C_x^2) + \alpha_2^2(\bar{Y}^2C_z^2) + 2\alpha_1\alpha_2(\bar{Y}^2\rho_{xz}C_xC_z)\right.ight.
\left. - 2\alpha_1(\bar{Y}^2\rho_{xy}C_xC_y - \bar{Y}B_1\bar{X}C_x^2 - \bar{Y}B_2\bar{Z}\rho_{xz}C_xC_z)\right.
\left. - 2\alpha_2(\bar{Y}^2\rho_{yz}C_yC_z - \bar{Y}B_2\bar{Z}C_z^2 - \bar{Y}B_1\bar{X}\rho_{xz}C_xC_z)\right.
\left. + (B_1^2\bar{X}^2C_x^2 + B_2^2\bar{Z}^2C_z^2 + 2B_1B_2\bar{X}\bar{Z}\rho_{xz}C_xC_z)\right.
\left. + \left(\bar{Y}^2C_y^2 - 2\bar{Y}B_1\bar{X}\rho_{xy}C_xC_y - 2\bar{Y}B_2\bar{Z}\rho_{yz}C_yC_z\right)\right] \quad (6)

2 The Proposed Estimator

Using (2) and (3), we proposed a new estimator defined as

\[ t_{akm} = \bar{y} \exp\left(\frac{\bar{X} - \bar{x}}{\bar{X} + (\alpha_1 - 1)\bar{x}}\right) \exp\left(\frac{\bar{Z} - \bar{z}}{\bar{Z} + (\alpha_2 - 1)\bar{z}}\right) \]
\[ + b_1(\bar{X} - \bar{x}) + b_2(\bar{Z} - \bar{z}) \quad (7) \]

To find the Mean square error (MSE) of \( t_{akm} \) up to the first order of approximation, we are using the following notations in literature [4, 6] and [7]

\[ \bar{y} = \bar{Y}(1 + e_0), \bar{x} = \bar{X}(1 + e_1), \bar{z} = \bar{Z}(1 + e_2) \]
\[ e_0 = \frac{\bar{y}}{\bar{Y}} - 1, e_1 = \frac{\bar{x}}{\bar{X}} - 1, e_2 = \frac{\bar{z}}{\bar{Z}} - 1 \]
\[ E(e_0) = 0, E(e_1) = 0, E(e_2) = 0 \]

\[ E(e_0^2) = \left(\frac{1-f}{n}\right)C_y^2, E(e_1^2) = \left(\frac{1-f}{n}\right)C_x^2, E(e_2^2) = \left(\frac{1-f}{n}\right)C_z^2, \]
\[ E(e_0e_1) = \left(\frac{1-f}{n}\right)\rho_{xy}C_xC_y, E(e_0e_2) = \left(\frac{1-f}{n}\right)\rho_{xy}C_xC_z, \]
\[ E(e_1e_2) = \left(\frac{1-f}{n}\right)\rho_{xz}C_xC_z \]

\[ b_1 = \frac{s_{yx}}{s_x^2} \text{ and } b_2 = \frac{s_{yz}}{s_z^2}, \] where \( s_{yx} \) and \( s_{yz} \) are the sample covariances between \( y \) and \( x \) and between \( z \) respectively.
Using above notations, we get
\[
t_{akm} = Y(1 + e_0) \exp \left[ \frac{X - X - Xe_1}{X(1 + (\alpha_1 - 1)(1 + e_1))} \right] \\
\times \left[ \frac{Z - Z - Ze_2}{Z(1 + (\alpha_2 - 1)(1 + e_2))} \right] \\
+ b_1(X - X - Xe_1) + b_1(X - X - Ze_2) \\
= Y(1 + e_0) \exp \left[ \frac{-e_1}{1 + (\alpha_1 - 1)(1 + e_1)} \right] \left[ \frac{-e_2}{1 + (\alpha_2 - 1)(1 + e_2)} \right] \\
- (b_1Xe_1 + b_2Ze_2)
\] (8)

After solving Equation (8), we get
\[
t_{akm} = Y \left( 1 - \frac{e_1}{\alpha_1} + k_1 e_1^2 + \frac{e_2^2}{2\alpha_1^2} - \frac{e_2}{\alpha_2} \right) + \frac{e_1e_2}{\alpha_1\alpha_2} + k_2 e_2^2 + \frac{e_2^2}{2\alpha_2^2} \\
+ e_0 - \frac{e_0e_1}{\alpha_1} - \frac{e_0e_2}{\alpha_2} \right) - (b_1Xe_1 + b_2Ze_2)
\] (9)

Let’s assume \( k_1 = \frac{\alpha_1 - 1}{\alpha_1^2}, k_2 = \frac{\alpha_2 - 1}{\alpha_2^2} \).

After avoiding the higher power of e’s in (9), we have
\[
(t_{akm} - Y) = \left[ Y(e_0 - \frac{e_1}{\alpha_1} - \frac{e_2}{\alpha_2}) - (b_1Xe_1 + b_2Ze_2) \right]
\] (10)

To find the MSE taking square of both sides of (10), we get
\[
(t_{akm} - Y)^2 = \left[ Y(e_0 - \frac{e_1}{\alpha_1} - \frac{e_2}{\alpha_2}) - (b_1Xe_1 + b_2Ze_2) \right]^2
\] (11)

Taking expectations on both the sides of (11), we have
\[
E(t_{akm} - Y)^2 = E \left[ \frac{1}{\alpha_1^2} \bar{Y}e_1^2 + \frac{1}{\alpha_2^2} \bar{Y}e_2^2 + \frac{2}{\alpha_1\alpha_2} \bar{Y}e_1e_2 \\
- \frac{2}{\alpha_1} (\bar{Y}e_0e_1 - b_1\bar{X}Ye_1^2 - b_2\bar{Z}Ye_1e_2) \\
- \frac{2}{\alpha_2} (\bar{Y}e_0e_2 - b_1\bar{X}Ye_1e_2 - b_2\bar{Z}Ye_2^2)
\]
An Improvement in Regression Estimator Through Exponential Estimator

\[ + (b_1^2x^2_e^2 + b_2^2z^2_e^2 + 2b_1b_2XZe_1e_2 + Y^2e_0^2) \]

\[ - 2b_1XYe_0e_1 - 2b_2YZe_0e_2) \]

\[ \text{MSE}(t_{akm}) = \frac{1}{\alpha_1^2}A_1 + \frac{1}{\alpha_2^2}A_2 + \frac{2}{\alpha_1\alpha_2}A_3 - \frac{2}{\alpha_1}A_4 - \frac{2}{\alpha_2}A_5 + A_6 \]

(12)

where,

\[ A_1 = E(Y^2e_1^2) = Y^2fC_x^2 \]
\[ A_2 = E(Y^2e_2^2) = Y^2fC_z^2 \]
\[ A_3 = E(Y^2e_1e_2) = Y^2f\rho_{xz}C_xC_z \]
\[ A_4 = E(Y^2e_0e_1 - Yb_1Xe_1^2 - Yb_2Ze_1e_2) \]
\[ = Y^2f\rho_{xy}C_xC_y - Yb_1Xe_1^2 - Yb_2Ze_1e_2 \]
\[ A_5 = E(Y^2e_0e_2 - Yb_2Ze_2^2 - Yb_1Xe_1e_2) \]
\[ = Y^2f\rho_{yz}C_yC_z - Yb_2Ze_2^2 - Yb_1Xe_1e_2 \]
\[ A_6 = E(b_1^2x^2_e^2 + b_2^2z^2_e^2 + 2b_1b_2XZe_1e_2 \]
\[ + Y^2e_0^2 - 2Yb_1Xe_0e_1 - 2Yb_2Ze_0e_2) \]
\[ = b_1^2X^2fC_x^2 + b_2^2Z^2fC_z^2 + 2b_1b_2XZf\rho_{xz}C_xC_z \]
\[ + Y^2fC_y^2 - 2Yb_1Xf\rho_{xy}C_xC_y - 2Yb_2Ze_2^2 \]

Differentiating (12) partially with respect to \( \alpha_1 \) and \( \alpha_2 \), we get

\[ \frac{A_1}{\alpha_1} + \frac{A_3}{\alpha_2} = A_4 \]  
(13)

\[ \frac{A_3}{\alpha_1} + \frac{A_2}{\alpha_2} = A_5 \]  
(14)

On solving Equations (13) and (14), we get

\[ \alpha_1 = \frac{A_1A_2 - A_3^2}{A_2A_4 - A_3A_5} \]
and
\[ \alpha_2 = \frac{A_3^2 - A_1A_2}{A_3A_4 - A_1A_5} \]
Table 1  Data statistics

<table>
<thead>
<tr>
<th></th>
<th>Population I</th>
<th>Population II</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>34</td>
<td>25</td>
</tr>
<tr>
<td>n</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>$\bar{x}$</td>
<td>208.88</td>
<td>14.3</td>
</tr>
<tr>
<td>$\bar{y}$</td>
<td>856.41</td>
<td>75.28</td>
</tr>
<tr>
<td>$\bar{z}$</td>
<td>199.44</td>
<td>6.82</td>
</tr>
<tr>
<td>$s_x$</td>
<td>150.22</td>
<td>3.17</td>
</tr>
<tr>
<td>$s_y$</td>
<td>733.14</td>
<td>17.27</td>
</tr>
<tr>
<td>$s_z$</td>
<td>150.22</td>
<td>1.53</td>
</tr>
<tr>
<td>$\rho_{yx}$</td>
<td>0.45</td>
<td>0.99</td>
</tr>
<tr>
<td>$\rho_{yz}$</td>
<td>0.45</td>
<td>0.89</td>
</tr>
<tr>
<td>$\rho_{xz}$</td>
<td>0.98</td>
<td>0.92</td>
</tr>
</tbody>
</table>

Table 2  Results MSE values of different estimators

<table>
<thead>
<tr>
<th>Estimators</th>
<th>MSE (Data Set I)</th>
<th>MSE (Data Set II)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_1$</td>
<td>26344.84</td>
<td>17.44</td>
</tr>
<tr>
<td>$t_2$</td>
<td>10976.42</td>
<td>15.19</td>
</tr>
<tr>
<td>$t_3$</td>
<td>8967.45</td>
<td>2.35</td>
</tr>
<tr>
<td>$t_{slcm}$</td>
<td><strong>8802.54</strong></td>
<td><strong>0.30</strong></td>
</tr>
</tbody>
</table>

3 Numerical Illustration

The performance of the proposed improve regression estimator through exponential estimator are assessed with two different data sets. From Singh and Chaudhary [8] data set I is taken and data set II is taken from the Cingi and Kadilar [9]. In the first data set area under wheat (1974) is our study variable, area under wheat (1971) is first auxiliary variable and area under wheat (1973) is the second auxiliary variable. For the second data set the population mean of the height of the fish is our study variable, the population mean of the length of the head is first auxiliary variable and the population mean of the length of the fin is the second auxiliary variable.

4 Conclusion

In the present paper, we have introduced an improvement in regression estimator through exponential estimator for finding the study variable’s
population mean using available information of two auxiliary variables. Taking results given in Table 2, we can have an idea that the introduced estimator $t_{\text{akm}}$ is performing better than other estimators in literature.

References
