A new estimator using two auxiliary variables

Cem Kadilar *, Hulya Cingi

Department of Statistics, Hacettepe University, Beytepe, Ankara 06532, Turkey

Abstract

Utilizing the estimator in Abu-Dayyeh et al. [Appl. Math. Comput. 139 (2003) 287], we suggest an estimator using two auxiliary variables in simple random sampling. We obtain mean square error (MSE) equation of the proposed estimator and theoretically compare it with the MSE of the traditional estimator using two auxiliary variables. By this comparison, we show the condition that the proposed estimator is more efficient than the traditional one. In addition, we support this theoretical result by an application with original data.

© 2004 Elsevier Inc. All rights reserved.

Keywords: Ratio estimator; Regression estimator; Auxiliary variable; Mean square error; Efficiency

1. Introduction

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

$$\bar{y}_{reg1} = \bar{y} + b(\bar{X} - \bar{x}),$$

where $b$ is an estimate of the change in $y$ when $x$ is increased by unity, $\bar{x}$ and $\bar{y}$ are the sample means of the $x_i$ and $y_i$, respectively. MSE of this estimate is

$$\text{MSE}(\bar{y}_{reg1}) = \frac{1}{n} f S_y^2 (1 - \rho_{xy}^2),$$

*Corresponding author.

E-mail addresses: kadilar@hacettepe.edu.tr (C. Kadilar), hcingi@hacettepe.edu.tr (H. Cingi).
where \( f = \frac{n}{N} \), \( n \) is the sample size, \( N \) is the population size, \( S^2_y \) and \( S^2_x \) are the population variances of the \( y_i \) and \( x_i \) respectively, \( \rho_{yx} = \frac{S_{yx}}{S_y S_x} \) is the population correlation coefficient between \( y_i \) and \( x_i \), \( S_{yx} \) is the population covariance between \( y_i \) and \( x_i \) [2].

When there are two auxiliary variates as \( X_{1i} \) and \( X_{2i} \), the regression estimate of \( Y \) will be

\[
\tilde{y}_{reg2} = \bar{y} + b_1(\overline{X}_1 - \overline{x}_1) + b_2(\overline{X}_2 - \overline{x}_2),
\]

where \( b_1 = \frac{s_{yx_1}}{s_{x_1}^2} \) and \( b_2 = \frac{s_{yx_2}}{s_{x_2}^2} \). Here, \( s_{x_1}^2 \) and \( s_{x_2}^2 \) are the sample variances of the \( x_{1i} \) and \( x_{2i} \), respectively, \( s_{yx_1} \) and \( s_{yx_2} \) are the sample covariances between \( y_i \) and \( x_{1i} \) and between \( y_i \) and \( x_{2i} \), respectively.

MSE of this estimate can be found as

\[
\text{MSE}(\tilde{y}_{reg2}) \approx \frac{1-f}{n}S^2_y\left(1 - \rho^2_{yx_1} - \rho^2_{yx_2} + 2\rho_{yx_1}\rho_{yx_2}\rho_{x_1x_2}\right).
\]

(see Appendix A).

2. The suggested estimator

Abu-Dayyeh et al. [1] proposed the following estimator of the population mean assuming that the population means \( \overline{X}_1 \) and \( \overline{X}_2 \) of the auxiliary variables were known:

\[
\tilde{y}_{\text{ratio}} = \bar{y}\left(\frac{\overline{X}_1}{\overline{x}_1}\right)^{x_1}\left(\frac{\overline{X}_2}{\overline{x}_2}\right)^{x_2},
\]

where \( x_1 \) and \( x_2 \) were real numbers.

We suggest using the ratio estimator given in (2.1) instead of \( \tilde{y} \) in (1.1). By this way, we obtain the following estimator:

\[
\tilde{y}_{pr} = \bar{y}\left(\frac{\overline{X}_1}{\overline{x}_1}\right)^{x_1}\left(\frac{\overline{X}_2}{\overline{x}_2}\right)^{x_2} + b_1(\overline{X}_1 - \overline{x}_1) + b_2(\overline{X}_2 - \overline{x}_2).
\]

MSE of this estimator can be found using Taylor series method defined as

\[
\begin{align*}
    h(\overline{x}_1, \overline{x}_2, \bar{y}) &
    \equiv
    h(\overline{X}_1, \overline{X}_2, \overline{Y}) + \frac{\partial h(c, d, e)}{\partial c}
    \bigg|_{\overline{X}_1, \overline{X}_2, \overline{Y}}(\overline{x}_1 - \overline{X}_1) \\
    &+ \frac{\partial h(c, d, e)}{\partial d}
    \bigg|_{\overline{X}_1, \overline{X}_2, \overline{Y}}(\overline{x}_2 - \overline{X}_2) + \frac{\partial h(c, d, e)}{\partial e}
    \bigg|_{\overline{X}_1, \overline{X}_2, \overline{Y}}(\bar{y} - \overline{Y})
\end{align*}
\]

(see [3]).
\[ \bar{y}_{pr} - \bar{y}_{pr} \cong \left( - z_1 \bar{Y}_{x_1} \left( \frac{X_1}{x_1^{x_1+1}} \right)^{x_2} - b_1 \right) \left( \bar{x}_1 - \bar{X}_1 \right) \\
\cong \left( - z_2 \bar{Y}_{x_2} \left( \frac{X_2}{x_2^{x_2+1}} \right)^{x_2} - b_2 \right) \left( \bar{x}_2 - \bar{X}_2 \right) \\
\cong \left( - z_1 \bar{Y}_{x_1} - b_1 \right) \left( \bar{x}_1 - \bar{X}_1 \right) \\
+ \left( - z_2 \bar{Y}_{x_2} - b_2 \right) \left( \bar{x}_2 - \bar{X}_2 \right) + (\bar{y} - \bar{Y}), \]

\[ E(\bar{y}_{pr} - \bar{y}_{pr})^2 \cong E \left[ (\bar{y} - \bar{Y}) - (z_1 R_1 + B_1)(\bar{x}_1 - \bar{X}_1) - (z_2 R_2 + B_2)(\bar{x}_2 - \bar{X}_2) \right]^2. \]

Note that we omit the differences of \( E(h_1 - B_1) \) and \( E(b_2 - B_2) \) [2].

\[ \text{MSE}(\bar{y}_{pr}) \cong \frac{1 - f}{n} \left\{ S_y^2 + (z_1 R_1 + B_1)^2 S_{x_1}^2 + (z_2 R_2 + B_2)^2 S_{x_2}^2 - 2(z_1 R_1 + B_1) S_y S_{y,x_1} - 2(z_2 R_2 + B_2) S_y S_{y,x_2} + 2(z_1 R_1 + B_1) \times (z_2 R_2 + B_2) S_{y,x_1,x_2} \right\}, \quad (2.3) \]

where \( R_1 = \frac{\bar{Y}}{\bar{X}_1} \), \( R_2 = \frac{\bar{Y}}{\bar{X}_2} \), \( B_1 = \frac{S_y}{S_{y,x_1}} \) and \( B_2 = \frac{S_y}{S_{y,x_2}} \).

We can have the optimal values of \( z_1 \) and \( z_2 \) in (2.3) by the following equations:

\[ z_1^* = \frac{S_y}{R_1 S_{x_1}} \rho_1^* \quad \text{and} \quad z_2^* = \frac{S_y}{R_2 S_{x_2}} \rho_2^*, \]

where

\[ \rho_1^* = \frac{\rho_{x_1 x_2} (\rho_{y,x_1} - \rho_{y,x_2})}{1 - \rho_{x_1 x_2}^2} \quad \text{and} \quad \rho_2^* = \frac{\rho_{x_1 x_2} (\rho_{y,x_2} - \rho_{y,x_1})}{1 - \rho_{x_1 x_2}^2} \]

(see Appendix B).

We can obtain minimum MSE of the suggested estimate using the optimal equations of \( z_1 \) and \( z_2 \) in (2.3) as follows:

\[ \text{MSE}_{\text{min}}(\bar{y}_{pr}) = \frac{1 - f}{n} S_y^2 \left\{ 1 + C_1^2 + C_2^2 + 2C_1 C_2 \rho_{x_1 x_2} - 2C_1 \rho_{y,x_1} - 2C_2 \rho_{y,x_2} \right\}, \quad (2.4) \]

where \( C_1 = \rho_1^* + \rho_{y,x_1} \) and \( C_2 = \rho_2^* + \rho_{y,x_2} \) (see Appendix C).
3. Efficiency comparison

We compare the MSE of the proposed estimator given in (2.4) with the MSE of the traditional estimator given in (1.2). We will have the condition as follows:

\[
\text{MSE}(\bar{y}_p) < \text{MSE}(\bar{y}_{\text{reg}2}) \\
\times \frac{1-f}{n} S_y^2 \left\{ 1 + (\rho_1 + \rho_{yx1})^2 + (\rho_2 + \rho_{yx2})^2 + 2(\rho_1 + \rho_{yx1})(\rho_2 + \rho_{yx2})\rho_{x1x2} - 2(\rho_1 + \rho_{yx1})\rho_{yx1} - 2(\rho_2 + \rho_{yx2})\rho_{yx2} \right\} \\
< \frac{1-f}{n} S_y^2 \left\{ 1 - \rho_{yx1}^2 - \rho_{yx2}^2 + 2\rho_{yx1}\rho_{yx2}\rho_{x1x2} \right\} \\
\times \rho_1^2 + \rho_2^2 + 2\rho_{x1x2}(\rho_1^*\rho_2^* + \rho_1^*\rho_{yx2} + \rho_2^*\rho_{yx1}) < 0. \quad (3.1)
\]

When this condition is satisfied, the proposed estimator will be more efficient than the traditional estimator.

4. Application

We have applied the traditional and proposed estimator on the data of apple production amount in 1999 (as interest of variate) and number of apple trees in 1999 (as first auxiliary variate), apple production amount in 1998 (as second auxiliary variate) of 204 villages in Black Sea Region of Turkey (Source: Institute of Statistics, Republic of Turkey). The statistics of the data of first auxiliary variate and interest of variate for all regions of Turkey can be found in Kadilar and Cingi [4].

In Table 1, we observe the sample size and the statistics about the population. However, we know that sample size has no effect on efficiency comparison as shown in Section 3. From Table 2, we see that the proposed estimator has smaller MSE than the traditional estimator has. Therefore we can say that the proposed estimator is more efficient than the traditional estimator.

Table 1

<table>
<thead>
<tr>
<th>N = 204</th>
<th>( S_{x1y2} = 94636084 )</th>
<th>( R_1 = 0.04 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 50</td>
<td>( S_{x1} = 77372777 )</td>
<td>( R_2 = 0.95 )</td>
</tr>
<tr>
<td>( \bar{X}_1 = 26441 )</td>
<td>( S_{x2} = 5684276 )</td>
<td>( z_1 = -1.33 )</td>
</tr>
<tr>
<td>( X_2 = 1014 )</td>
<td>( \rho_{x1x2} = 0.83 )</td>
<td>( z_2 = 0.17 )</td>
</tr>
<tr>
<td>( Y = 966 )</td>
<td>( \rho_{yx1} = 0.71 )</td>
<td>( \rho_1^* = -0.89 )</td>
</tr>
<tr>
<td>( S_{x1} = 45402.78 )</td>
<td>( \rho_{yx2} = 0.94 )</td>
<td>( \rho_2^* = 0.17 )</td>
</tr>
<tr>
<td>( S_{x2} = 2521.40 )</td>
<td>( B_1 = 0.04 )</td>
<td>( C_1 = -0.21 )</td>
</tr>
<tr>
<td>( S_y = 2389.76 )</td>
<td>( B_2 = 0.89 )</td>
<td>( C_2 = 1.12 )</td>
</tr>
</tbody>
</table>
estimator for these data. This result is not surprising because the condition (3.1) is satisfied as follows:

\[ \rho_1^2 + \rho_2^2 + 2 \rho_{x_1 x_2} (\rho_1^* \rho_2^* + \rho_1^* \rho_{y x_2} + \rho_2^* \rho_{y x_1}) = -0.62. \]

5. Conclusion

We have developed a new estimator, which is found more efficient than the traditional estimator using two auxiliary variables for the condition (3.1). This theoretical inference is also satisfied by the result of an application with original data. In future, we hope to extend the estimator presented here for the development of a new estimator in two-stage sampling.

Appendix A

\[
\begin{align*}
\bar{y}_{\text{reg}2} &= \bar{y} + b_1(\bar{X} - \bar{x}_1) + b_2(\bar{X}_2 - \bar{x}_2), \\
V(\bar{y}_{\text{reg}2}) &= \frac{1 - f}{n} S_y^2 = \frac{1 - f}{N - 1} \left( \frac{1}{N - 1} \sum_{i=1}^{N} (y_i - \bar{Y})^2 \right) \\
&= \frac{1 - f}{n} \left( \frac{1}{N - 1} \sum_{i=1}^{N} (y_i - \bar{Y} - B_1(x_{1i} - \bar{X}_1) - B_2(x_{2i} - \bar{X}_2))^2 \right) \\
&= \frac{1 - f}{n} \left\{ \frac{1}{N - 1} \sum_{i=1}^{N} (y_i - \bar{Y})^2 + B_1^2(x_{1i} - \bar{X}_1)^2 + B_2^2(x_{2i} - \bar{X}_2)^2 \\
&\quad - 2B_1(y_i - \bar{Y})(x_{1i} - \bar{X}_1) - 2B_2(y_i - \bar{Y})(x_{2i} - \bar{X}_2) \\
&\quad + 2B_1B_2(x_{1i} - \bar{X}_1)(x_{2i} - \bar{X}_2) \right\} \\
&= \frac{1 - f}{n} (S_y^2 + B_1^2 S_{x_1}^2 + B_2^2 S_{x_2}^2 - 2B_1 S_{yx_1} - 2B_2 S_{yx_2} + 2B_1B_2 S_{x_1x_2}) \\
&= \frac{1 - f}{n} S_y^2(1 + \rho_{x_1}^2 + \rho_{x_2}^2 - 2 \rho_{x_1} \rho_{x_2} - 2 \rho_{x_2}^2 + 2 \rho_{x_1} \rho_{x_2}^2 \rho_{x_1x_2}) \\
&= \frac{1 - f}{n} S_y^2(1 - \rho_{x_1}^2 - \rho_{x_2}^2 + 2 \rho_{x_1} \rho_{x_2} \rho_{x_1x_2}).
\end{align*}
\]
When first degree approximation is used, it is known that MSE is equal to the variance.

**Appendix B**

\[
\frac{\partial \text{MSE}(\bar{y}_{pr})}{\partial x_1} = \frac{1-f}{n} \left[(2x_1R_1^2 + 2R_1B_1)S_{x_1}^2 - 2R_1S_{x_{1x_2}} + 2R_1(x_2R_2 + B_2)S_{x_1x_2}\right] \\
= 0,
\]

\[
(x_1R_1 + B_1)S_{x_1}^2 - S_{x_{1x_1}} + (x_2R_2 + B_2)S_{x_1x_2} = 0,
\]

\[
x_1 = \frac{-B_1S_{x_1}^2 + S_{x_{1x_1}} - x_2R_2S_{x_1x_2} - B_2S_{x_1x_2}}{R_1S_{x_1}^2} = \frac{-x_2R_2S_{x_1x_2} - B_2S_{x_1x_2}}{R_1S_{x_1}^2} \\
= \frac{-S_{x_1x_2}(x_2R_2 + B_2)}{R_1S_{x_1}^2}. \quad (B.1)
\]

\[
\frac{\partial \text{MSE}(\bar{y}_{pr})}{\partial x_2} = \frac{1-f}{n} \left[(2x_2R_2^2 + 2R_2B_2)S_{x_2}^2 - 2R_2S_{x_{1x_2}} + 2R_2(x_1R_1 + B_1)S_{x_1x_2}\right] \\
= 0,
\]

\[
x_2 = \frac{-x_1R_1S_{x_1x_2} - B_1S_{x_1x_2}}{R_2S_{x_2}^2} \\
= \frac{-S_{x_1x_2}(x_1R_1 + B_1)}{R_2S_{x_2}^2} \quad (B.2)
\]

Using (B.2) in (B.1), we have

\[
x_1 = \frac{(x_1R_1S_{x_1x_2} + B_1S_{x_1x_2})S_{x_1x_2} - B_2S_{x_1x_2}S_{x_2}^2}{R_1S_{x_1}^2S_{x_2}^2},
\]

\[
x_1R_1S_{x_1x_2}^2 - x_1R_1S_{x_1x_2}^2 = B_1S_{x_1x_2}^2 - B_2S_{x_1x_2}S_{x_2}^2,
\]

\[
x_1' = \frac{B_1S_{x_1x_2}^2S_{x_2}^2 \rho_{x_1x_2} - B_2\rho_{x_1x_2}S_{x_1x_2}^3}{R_1S_{x_1}^2S_{x_2}^2(1 - \rho_{x_1x_2}^2)} \\
= \frac{B_1\rho_{x_1x_2}^2S_{x_1} - B_2\rho_{x_1x_2}S_{x_2}}{R_1S_{x_1}(1 - \rho_{x_1x_2}^2)} = \frac{\rho_{x_1x_2}(B_1S_{x_1}^2\rho_{x_1x_2} - B_2S_{x_2}^2)}{R_1S_{x_1}(1 - \rho_{x_1x_2}^2)} \quad (B.4)
\]
Note that $B_1S_{x_1} = \rho_{x_1} S_y$ and $B_2S_{x_2} = \rho_{x_2} S_y$.

\[
\alpha_1^* = \frac{\rho_{x_1} S_y (\rho_{x_1} - \rho_{x_2})}{R_1 S_{x_1} (1 - \rho_{x_2}^2)} = \frac{S_y}{R_1 S_{x_1}} \frac{\rho_{x_1} (\rho_{x_1} - \rho_{x_2})}{1 - \rho_{x_2}^2}.
\]

Using (B.4) in (B.3), we have

\[
\alpha_2^* = -\frac{S_{x_1} (B_1 \rho_{x_1}^2 S_{x_1} - B_2 \rho_{x_2} S_{x_2})}{R_2 S_{x_2}^2 (1 - \rho_{x_2}^2)} + \frac{B_1}{R_2 S_{x_2}^2}
\]

\[
= -\frac{S_{x_1} (B_1 \rho_{x_1}^2 S_{x_1} - B_2 \rho_{x_2} S_{x_2})}{R_2 S_{x_2}^2 (1 - \rho_{x_2}^2)} - \frac{\rho_{x_1} S_y (\rho_{x_1} - \rho_{x_2})}{R_2 S_{x_2} (1 - \rho_{x_2}^2)}
\]

\[
= \frac{S_y}{R_2 S_{x_2}} \frac{\rho_{x_1} (\rho_{x_1} - \rho_{x_2})}{1 - \rho_{x_2}^2}.
\]

Appendix C

\[
MSE_{\text{min}}(\bar{y}_{pr}) \approx \frac{1 - f}{n} \left\{ S_y^2 + \left( \frac{S_y}{R_1 S_{x_1}} \rho_1^* R_1 + B_1 \right)^2 S_{x_1}^2 \right. \\
+ \left( \frac{S_y}{R_2 S_{x_2}} \rho_2^* R_2 + B_2 \right)^2 S_{x_2}^2 - 2 \left( \frac{S_y}{R_1 S_{x_1}} \rho_1^* R_1 + B_1 \right) S_{x_1} \\
- \left. 2 \left( \frac{S_y}{R_2 S_{x_2}} \rho_2^* R_2 + B_2 \right) S_{x_2} + 2 \left( \frac{S_y}{R_1 S_{x_1}} \rho_1^* R_1 + B_1 \right) \right\}
\]

\[
\approx \frac{1 - f}{n} \left\{ 1 + \rho_1^{*2} + 2 \frac{B_1 S_{x_1}}{S_y} \rho_1^* + B_1^2 \frac{S_{x_1}^2}{S_y} + \rho_2^{*2} + 2 \frac{B_2 S_{x_2}}{S_y} \rho_2^* \right.
\]

\[
+ B_1^2 \frac{S_{x_2}^2}{S_y} - 2 \frac{S_{x_1} S_{x_2}}{S_y} \rho_1^* - 2 B_1 \frac{S_{x_1}}{S_y^2} - 2 \frac{S_{x_2}}{S_y} \rho_2^* \\
- 2 B_2 \frac{S_{x_1}}{S_y} + 2 \rho_1^* \rho_2^* \frac{S_{x_1} S_{x_2}}{S_y} + 2 \rho_1^* B_2 \frac{S_{x_2}}{S_y} \\
+ 2 \rho_2^* B_1 \frac{S_{x_1}}{S_y} + 2 B_1 B_2 \frac{S_{x_1} S_{x_2}}{S_y} \right\}
\]
\[
\begin{align*}
\Rightarrow & \quad \frac{1-f}{n} S^2 \left\{ 1 + \rho_1^2 + 2 \rho_{3y1} \rho_1^* + \rho_{3x1}^2 + \rho_2^2 + 2 \rho_{3y2} \rho_2^* + \rho_{3x2}^2 \\
&\quad - 2 \rho_{3y1} \rho_1^* - 2 \rho_{3x1}^2 - 2 \rho_{3y2} \rho_2^* - 2 \rho_{3x2}^2 + 2 \rho_1^* \rho_2^* \rho_{x1x2} \\
&\quad + 2 \rho_1^* \rho_{3y2} \rho_{x1x2} + 2 \rho_2^* \rho_{3y1} \rho_{x1x2} + 2 \rho_{3y1} \rho_{3y2} \rho_{3x1x2} \right\} \\
\Rightarrow & \quad \frac{1-f}{n} S^2 \left\{ 1 + (\rho_1^* + \rho_{3y1})^2 + (\rho_2^* + \rho_{3y2})^2 \\
&\quad + 2(\rho_1^* + \rho_{3y1})(\rho_2^* + \rho_{3y2}) \rho_{x1x2} - 2(\rho_1^* + \rho_{3y1}) \rho_{3x1} \\
&\quad - 2(\rho_2^* + \rho_{3y2}) \rho_{3x2} \right\}.
\end{align*}
\]

References