Abstract

Chao [1] considered the ratio estimator for the population total of the variable of interest in adaptive cluster sampling by using only one auxiliary variable. In this paper we focus on the use of transformed auxiliary variables for estimating the population total for the variable of interest by using two auxiliary variables (based on Gupta and Shabbir [2]) . Simulations showed the proposed estimator had the smallest estimated mean square error when compared the ratio estimators using only one auxiliary variable.

Keywords: adaptive cluster sampling, auxiliary variable, ratio estimator.

1. INTRODUCTION

Adaptive cluster sampling, proposed by Thompson [4], is an efficient method for sampling rare and hidden clustered populations. In adaptive cluster sampling, an initial sample of units is selected by simple random sampling. If the value of the variable of interest from a sampled unit satisfies a pre-specified condition $C$, that is $\{i, y_i \geq c\}$, then the unit’s neighborhood will also be added to the sample. If any other units that are “adaptively” added also satisfy the condition $C$, then their neighborhoods are also added to the sample. This process is continued until no more units that satisfy the condition are found. The set of all units selected and all neighboring units that satisfy the condition is called a network. The adaptive sample units, which do not satisfy the
condition are called edge units. A network and its associated edge units are called a cluster. If a unit is selected in the initial sample and does not satisfy the condition $C^*$, then there is only one unit in the network. A neighborhood must be defined such that if unit $i$ is in the neighborhood of unit $j$ then unit $j$ is in the neighborhood of unit $i$. In this paper, a neighborhood of a unit is defined as the four spatially adjacent units, that is to the left, right, top and bottom of that unit as shown in figure 1.

<table>
<thead>
<tr>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 1 The example of network where a unit neighborhood is defined as four spatially adjacent units

Figure 1 illustrates the example of a network. The unit with a star is the initial unit selected. The condition to adaptively added units is a value greater than or equal to 1. Units that are to the left, right, top, and bottom of one another make up a neighborhood. The units in the gray shading form a single network. The units in bold numbers are edge units of the network. The network and its edge units make up a cluster.

In some situations, the researcher obtains observations from more than one variable, including the variable of interest and the auxiliary variable. The use of auxiliary variable is a common method to improve the precision of the estimation of the population parameter. Chao [1] proposed a ratio estimator by using one auxiliary variable in adaptive cluster sampling. In this paper, we will study the ratio estimator by using two auxiliary variables in adaptive cluster sampling. Section 2 and 3 are a review of important concepts of adaptive cluster sampling and the ratio estimation in adaptive cluster sampling, respectively. The proposed estimator using two auxiliary variables is provided in section 4. Simulation studies are in section 5 and the conclusions of this study are in the last section.
2. Adaptive Cluster Sampling

Let \( y \) be the variable of interest defined on the finite population, and the population consists of a set of \( N \) units \( \{u_1, u_2, \ldots, u_N\} \) indexed by their labels \( S = \{1, 2, \ldots, N\} \). With unit \( i \) is associated the variable of interest \( y_i \). The population total of \( y \) is \( \tau_y = \sum_{i=1}^{N} y_i \). Let \( \hat{\tau}_{y,A} \) be the estimator of the population total in adaptive cluster sampling.

Let \( n \) denote the initial sample size and \( v \) denote the final sample size. Let \( \psi_i \) denote the network that includes unit \( i \) and \( m_i \) be the number of units in that network. The initial sample of units is selected by simple random sampling without replacement.

The Hansen-Hurwitz estimator of the population total for the variable of interest can be written as (Thompson, [5]; Thompson and Seber, [6])

\[
\hat{\tau}_{y,A} = \frac{N}{n} \sum_{i=1}^{N} \frac{f_i}{m_i} \frac{y_i}{E(f_i)},
\]

where \( f_i \) is the number of units in the initial sample which fall in the network \( \psi_i \) that includes units \( i \). \( f_i = 0 \) if no units in the initial sample fall in the network \( \psi_i \). Since \( f_i \) units are selected from the \( m_i \) units in the network \( \psi_i \), \( f_i \) has the hypergeometric distribution \( \binom{N, m_i, n}{N, m_i, n} \), so \( E(f_i) = nm_i/N \), we have

\[
\hat{\tau}_{y,A} = \frac{N}{n} \sum_{i=1}^{N} \frac{y_i}{m_i} \frac{f_i}{m_i},
\]

\[
= \frac{N}{n} \sum_{i=1}^{N} \frac{1}{m_i} \sum_{j \in \psi_i} y_j,
\]

\[
= \frac{N}{n} \sum_{i=1}^{N} \left( w_y \right)_i, \tag{1}
\]

where \( \left( w_y \right)_i \) is the average of the variable of interest in the network that includes unit \( i \) of the initial sample, that is, \( \left( w_y \right)_i = \frac{1}{m_i} \sum_{j \in \psi_i} y_j \).
The variance of \( \hat{\tau}_{y \_A} \) is

\[
V(\hat{\tau}_{y \_A}) = \frac{N(N-n)}{(N-1)} \sum_{i=1}^{N} \left( w_i - \frac{\tau_y}{N} \right)^2.
\] (2)

An unbiased estimator of \( V(\hat{\tau}_{y \_A}) \) is

\[
\hat{V}(\hat{\tau}_{y \_A}) = \frac{N(N-n)}{n(n-1)} \sum_{i=1}^{n} \left( w_i - \frac{\hat{\tau}_{y \_A}}{N} \right)^2.
\] (3)

3. The Ratio Estimator in Adaptive Cluster Sampling

For adaptive cluster sampling, the initial sample of units is selected by simple random sampling without replacement. Let \( x \) be the auxiliary variable. The population total of \( x \) is \( \tau_x = \sum_{i=1}^{N} x_i \)

The modified Hansen-Hurwitz estimator of the population total of the variable of interest can be written as (Thompson, [5]; Thompson and Seber, [6])

\[
\hat{\tau}_{y \_A} = \frac{N}{n} \sum_{i=1}^{n} (w_y)_i,
\]

The modified Hansen-Hurwitz estimator of the population total of the auxiliary variable (\( x \)) can be written as

\[
\hat{\tau}_{x \_A} = \frac{N}{n} \sum_{i=1}^{n} (w_x)_i,
\] (4)

where \( (w_x)_i \) is the average of the auxiliary variable in the network that includes unit \( i \) of the initial sample, that is,

\[
(w_x)_i = \frac{1}{m_i} \sum_{j \in \psi_i} x_j.
\]

Let \( R \) be the population ratio between \( y \) and \( x \), \( R = \frac{\tau_y}{\tau_x} \). The ratio estimator of the population total is

\[
\hat{\tau}_{y \_A} = \left( \frac{\hat{\tau}_{y \_A}}{\hat{\tau}_{x \_A}} \right) \tau_x
\]
\[ \hat{\tau}_{y-A} \left( \frac{\tau_x}{\hat{\tau}_{x-A}} \right) \]

\[ \hat{\tau}_{y-A} + \hat{R}_A \left( \tau_x - \hat{\tau}_{x-A} \right), \quad (5) \]

where \( \hat{R}_A = \frac{\hat{y}_{y-A}}{\hat{\tau}_{x-A}} \) is the estimator of the population ratio.

The first term in the Taylor’s expansion about the point \((\tau_x, \tau_y)\) gives the approximation

\[ \hat{\tau}_{R_x-A} \approx \hat{\tau}_{y-A} + R \left( \tau_x - \hat{\tau}_{x-A} \right) \]

\[ \hat{\tau}_{R_x-A} - \tau_y \approx \hat{\tau}_{y-A} + R \left( \tau_x - \hat{\tau}_{x-A} \right) - \tau_y = \hat{\tau}_{y-A} - \hat{R}_x \hat{\tau}_{x-A} . \quad (6) \]

So the variance of \( \hat{\tau}_{R_x-A} \) is approximated by

\[ V \left( \hat{\tau}_{R_x-A} \right) \approx E \left( \left( \hat{\tau}_{y-A} - R \hat{\tau}_{x-A} \right)^2 \right) = V \left( \hat{\tau}_{y-A} - R \hat{\tau}_{x-A} \right) . \quad (7) \]

4. The Proposed Estimator

Let \( y \) be the variable of interest and \( x \) and \( z \) be two auxiliary variables defined on the finite population. Let \( \delta_y = \left( \hat{\tau}_{y-A} - \tau_y \right) / \tau_y \), \( \delta_x = \left( \hat{\tau}_{x-A} - \tau_x \right) / \tau_x \), and \( \delta_z = \left( \hat{\tau}_{z-A} - \tau_z \right) / \tau_z \) then we have \( E(\delta_y) = 0 \), \( E(\delta_x) = 0 \), \( E(\delta_z) = 0 \) and

\[ E \left( \delta_y^2 \right) = \lambda \lambda y_{y-A}, \quad E \left( \delta_x^2 \right) = \lambda \lambda x_{x-A} \quad \text{and} \quad E \left( \delta_z^2 \right) = \lambda \lambda z_{z-A}, \quad \text{where} \]

\[ \hat{\tau}_{j-A} \quad \text{is the estimator of the population total in adaptive cluster sampling,} \quad (j = y, x, z), \]

\[ \lambda = N^2 \left( N - n \right) / Nn , \]

\[ A_j^2 = S_j^2 / \tau_j^2 , \]

\[ S_j^2 = \sum_{i=1}^{N} \left( w_j - \frac{\tau_j}{N} \right)^2 / (N - 1) \quad (j = y, x, z) . \]
Let \( \rho_{ij,A} = s_{ij,A}/s_{i,A} s_{j,A} \) (\( i, j = x, y, z \) and \( i \neq j \)) be the correlation coefficient between \( i \) and \( j \) and \( s_{i,A}, s_{j,A}, s_{ij,A} \) be the standard deviations and the covariance term.

The modified Hansen-Hurwitz estimator of the population total of the auxiliary variable \( z \) can be written as
\[
\hat{\tau}_{z,A} = \frac{N}{n} \sum_{i=1}^{n} (w_z)_i ,
\]  
where \( (w_z)_i \) is the average of the auxiliary variable in the network that includes unit \( i \) of the initial sample, that is,
\[
(w_z)_i = \frac{1}{m_i} \sum_{j \in \psi_i} z_j .
\]

Let \( u_i = x_i \) and \( v_i = z_i \) \( (i = 1, 2, \ldots, N) \). The population total of these transformed variable are \( U = \sum_{i=1}^{N} x_i = \tau_x \) and \( V = \sum_{i=1}^{N} z_i = \tau_z \). The estimator of \( U \) and \( V \) are
\[
u = \frac{N}{n} \sum_{i=1}^{n} (w_x)_i = \hat{\tau}_{x,A} \quad \text{and} \quad \nu = \frac{N}{n} \sum_{i=1}^{n} (w_z)_i = \hat{\tau}_{z,A} .
\]

We now introduce the class of ratio estimators. The form of multiplicative generalization for the estimator of a population total can be written as (based on Gupta and Shabbir [2])
\[
\hat{\tau}_{R_{xz}} = \hat{\tau}_y \left( f(x) \right)^{J_1} \left( f(z) \right)^{J_2} .
\]

So the multiplicative generalization of \( \hat{\tau}_{R_{xz,A}} \) is
\[
\hat{\tau}_{R_{xz,A}} = \hat{\tau}_y (\frac{\tau_x}{\hat{\tau}_{x,A}})^{J_1}\left( \frac{\tau_z}{\hat{\tau}_{z,A}} \right)^{J_2} = \hat{\tau}_y (\frac{U}{\nu})^{J_1} \left( \frac{V}{\nu} \right)^{J_2} .
\]  
(9)

So
\[
\hat{\tau}_{R_{xz,A}} = \hat{\tau}_y (1 + \delta_y) (1 + \delta_x)^{-J_1} (1 + \delta_z)^{-J_2} ,
\]  
(10)

where \( \delta_y = (\hat{\tau}_{y,A} - \tau_y)/\tau_y , \delta_x = (\hat{\tau}_{x,A} - \tau_x)/\tau_x \) and \( \delta_z = (\hat{\tau}_{z,A} - \tau_z)/\tau_z \).

Note 1) If \( J_1 = 0 \) and \( J_2 = 0 \) then \( \hat{\tau}_{R_{xz,A}} = \hat{\tau}_{y,A} \)
2) If \((J_1, J_2) = (1,1), (-1,-1), (1,-1)\) and \((-1,1)\) then \(\hat{\tau}_{R_{xz}}\) are called ratio-type, product-type, ratio-cum-product-type and product-cum-ratio-type estimator.

A Taylor series expansion is possible in (10). This gives (based on Gupta and Shabbir [2])

\[
\hat{\tau}_{R_{xz}} = \tau_y \left[ 1 + \delta_y - J_1 \delta_x - J_2 \delta_z - J_2 \delta_y \delta_x - J_1 J_2 \delta_x \delta_z + \frac{J_1 (J_1 + 1)}{2} \delta_x^2 + \frac{J_2 (J_2 + 1)}{2} \delta_z^2 \right] + f(\delta_i),
\]

where \(f(\delta_i)\) represents the third or higher order terms in \(\delta_i\)’s.

Therefore, the estimator of the population total can be approximated by

\[
\hat{\tau}_{R_{xz}} \approx \tau_y \left[ 1 + \delta_y - J_1 \delta_x - J_2 \delta_z - J_2 \delta_y \delta_x - J_1 J_2 \delta_x \delta_z + \frac{J_1 (J_1 + 1)}{2} \delta_x^2 + \frac{J_2 (J_2 + 1)}{2} \delta_z^2 \right].
\] (11)

The expected of the approximation of \(\hat{\tau}_{R_{xz}}\) is

\[
E(\hat{\tau}_{R_{xz}}) \approx \tau_y - \tau_y J_1 \rho_{yx} A_y A_x - \tau_y J_2 \rho_{yz} A_y A_z + \tau_y J_2 \rho_{yx} A_x A_y + \tau_y J_1 \rho_{yz} A_x A_z.
\]

So the approximate bias is given by

\[
B(\hat{\tau}_{R_{xz}}) \approx \tau_y .
\]

and the approximate MSE is given by
The value of $J_1$ and $J_2$ are derived by minimizing \( \text{MSE}(\hat{\tau}_{R_{x_{z}-A}}) \) in (12) with respect to $J_1$ and $J_2$. That is, \( \frac{\partial}{\partial J_1} \text{MSE}(\hat{\tau}_{R_{x_{z}-A}}) = 0 \) and \( \frac{\partial}{\partial J_2} \text{MSE}(\hat{\tau}_{R_{x_{z}-A}}) = 0 \).

So the optimum values of $J_1$ and $J_2$ are

\[
J_1 = \frac{A_{y_{-}A}}{A_{x_{-}A}(1-\rho_{xz_{-}A}^2)} \left( \rho_{yx_{-}A} - \rho_{yz_{-}A} \rho_{zx_{-}A} \right) \quad \text{and} \quad J_2 = \frac{A_{y_{-}A}}{A_{z_{-}A}(1-\rho_{xz_{-}A}^2)} \left( \rho_{yz_{-}A} - \rho_{yx_{-}A} \rho_{zx_{-}A} \right). \tag{14}
\]

Substituting (14) in (13), we get the minimum MSE of $\hat{\tau}_{R_{x_{z}-A}}$ as

\[
\text{MSE}(\hat{\tau}_{R_{x_{z}-A}}) \approx \lambda \tau_y^2 A_{y_{-}A}^2 \left[ 1 - \frac{\left( \rho_{yx_{-}A}^2 + \rho_{yz_{-}A}^2 - 2 \rho_{yx_{-}A} \rho_{yz_{-}A} \rho_{zx_{-}A} \right)}{1 - \rho_{xz_{-}A}^2} \right].
\]

Because \( \frac{\partial^2}{\partial J_1^2} \text{MSE}(\hat{\tau}_{R_{x_{z}-A}}) \geq 0 \) and \( \frac{\partial^2}{\partial J_2^2} \text{MSE}(\hat{\tau}_{R_{x_{z}-A}}) \geq 0 \).

Substituting (14) in (12), we get the bias of $\hat{\tau}_{R_{x_{z}-A}}$ as

\[
B(\hat{\tau}_{R_{x_{z}-A}}) \approx \frac{\lambda \tau_y A_{y_{-}A}}{2(1-\rho_{xz_{-}A}^2)} \left[ A_{x_{-}A} \left( \rho_{yx_{-}A} - \rho_{yz_{-}A} \rho_{zx_{-}A} \right) + A_{z_{-}A} \left( \rho_{yz_{-}A} - \rho_{yx_{-}A} \rho_{zx_{-}A} \right) \right. \\
- \left. A_{y_{-}A} \left( \rho_{yx_{-}A}^2 + \rho_{yz_{-}A}^2 - 2 \rho_{yx_{-}A} \rho_{yz_{-}A} \rho_{zx_{-}A} \right) \right].
\]

The $J_1$ and $J_2$ were estimated by

\[
\hat{J}_1 = \frac{A_{y_{-}A}}{A_{x_{-}A}(1-\rho_{xz_{-}A}^2)} \left( r_{yx_{-}A} - r_{yz_{-}A} r_{zx_{-}A} \right) \quad \text{and} \quad \hat{J}_2 = \frac{A_{y_{-}A}}{A_{z_{-}A}(1-\rho_{xz_{-}A}^2)} \left( r_{yx_{-}A} - r_{yx_{-}A} r_{zx_{-}A} \right),
\]
where \( \hat{A}_{j-A}^2 = \frac{s_{j-A}^2}{\hat{\tau}_j^2} \), \( s_{j-A}^2 = \sum_{i=1}^{n} \left( \frac{w_{ij}}{n} - \frac{\hat{\tau}_{y-A}}{N} \right)^2 \), \( (n-1) \), \( (j = y, x, z) \).

Let \( r_{ij-A} = s_{ij-A}/s_{i-A}s_{j-A} \) \( (i, j = x, y, z \text{ and } i \neq j) \) be the sample correlation coefficient between \( i \) and \( j \) and \( s_{i-A}, s_{j-A}, s_{ij-A} \) be the standard deviation and the covariance term of the sample.

The \( MSE(\hat{\tau}_{R_{xz-A}}) \) was estimated by

\[
MSE(\hat{\tau}_{R_{xz-A}}) \approx \lambda \hat{\tau}_{y-A}^2 \hat{A}_{y-A}^2 \left[ 1 - \frac{r_{yx-A}^2 + r_{yz-A}^2 - 2r_{yx-A}r_{yz-A}r_{xz-A}}{1 - r_{xz-A}^2} \right].
\]

5. Simulation Study

The step of the simulation data \( X, Y \) and \( Z \) are as follows:

1. To generate the number of network and random the position of the start unit for the network. To generate the value of \( X, Y \) and \( Z \) from the normal distribution, where \( \mu_y = 20, \sigma_y = 5, \mu_x = 10, \sigma_x = 2 \) and \( \mu_z = 50, \sigma_z = 5 \).

2. For each the start unit, to consider the unit’s neighborhoods and for each unit’s neighborhoods, to generate \( 0 \leq P \leq 1 \) if \( P \leq 0.3 \) then generating \( X, Y \) from the bivariate normal distribution with \( \rho_{yx} = 0.80, \) where \( \mu_y = 5, \sigma_y = 2, \mu_x = 2, \sigma_x = 1 \) and generating \( Y, Z \) from the bivariate normal distribution with \( \rho_{yz} = 0.80, \) where \( \mu_z = 10, \sigma_x = 3, \) otherwise \( X, Y \) and \( Z \) are equal 0. The values of \( X, Y \) and \( Z \) are truncated to be the integer. If the values of \( X, Y \) and \( Z \) are negative values then \( X, Y \) and \( Z \) are set to be zero.
The populations were shown in Figure 2 – 4.

<p>| | | | | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>24</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>22</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>33</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>21</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>5</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>7</td>
<td>0</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>5</td>
<td>8</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>7</td>
<td>65</td>
<td>0</td>
<td>4</td>
<td>5</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>0</td>
<td>7</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>27</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>29</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 2. $Y$ values
<table>
<thead>
<tr>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>11</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>11</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>3</td>
<td>18</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>27</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 3. $X$ values
\begin{figure}
\centering
\begin{tabular}{cccccccccccccccccccc}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 12 & 77 & 8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 57 & 10 & 8 & 14 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 10 & 0 & 10 & 8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 55 & 1 & 0 & 0 & 97 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 74 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 11 & 0 & 10 & 11 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 12 & 12 & 0 & 10 & 10 & 12 & 12 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 6 & 8 & 12 & 0 & 12 & 8 & 12 & 10 & 0 & 0 & 10 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 14 & 10 & 8 & 12 & 14 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 12 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 59 & 0 & 0 & 54 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 63 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\end{tabular}
\caption{Z values}
\end{figure}
The values of the population totals are $\tau_y = 489$, $\tau_x = 222$, $\tau_z = 1012$. The correlation between $w_y$ and $w_x$ is 0.926 and. The correlation between $w_y$ and $w_z$ is 0.949. The values of $S_{y,y_A}^2$, $S_{x,y_A}^2$, $S_{z,y_A}^2$ are 12.686, 3.794 and 59.834, respectively. The values of $A_{y,y_A}^2$, $A_{x,y_A}^2$, $A_{z,y_A}^2$ are $5.305 \times 10^{-5}$, $7.698 \times 10^{-5}$ and $5.842 \times 10^{-5}$, respectively.

For each iteration, an initial sample of units is selected by simple random sampling without replacement. The $y$-values are obtained for keeping the sample network. In each sample network the $x$-values and $z$-values are obtained. We can compute the estimator of the population total of the variable of interest, 

$$\hat{\tau}_{y,y_A} = \frac{N}{n} \sum_{i=1}^{n} (w_y)_i$$

the estimator of the population total of the auxiliary variable ($x$),

$$\hat{\tau}_{x,y_A} = \frac{N}{n} \sum_{i=1}^{n} (w_x)_i$$

and the estimator of the population total of the auxiliary variable ($z$),

$$\hat{\tau}_{z,y_A} = \frac{N}{n} \sum_{i=1}^{n} (w_z)_i.$$ So the multiplicative generalization of $\hat{\tau}_{y,y_A}$ is

$$\hat{\tau}_{R_{xz,y_A}} = \frac{\hat{\tau}_{y,y_A}}{\hat{\tau}_{x,y_A}} \left( \frac{\tau_x}{\hat{\tau}_{x,y_A}} \right)^{J_1} \left( \frac{\tau_z}{\hat{\tau}_{z,y_A}} \right)^{J_2}$$

For each estimator 5,000 iterations were performed to obtain an accuracy estimate. The condition for added units in the sample is defined by $C = \{ y : y > 0 \}$. Initial SRS sizes were varied $n = 2, 5, 10, 15, 20, 25$ and 30 were used.

The estimated mean square error of the estimate total is

$$MSE(\hat{\tau}) = \frac{1}{5,000} \sum_{i=1}^{5,000} \left( \hat{\tau}_i - \tau_y \right)^2,$$

where $\hat{\tau}_i$ is the value for the relevant estimator for sample $i$. Let $v$ denote the final sample size.

The results of the adaptive cluster sampling $\left( \hat{\tau}_{y,y_A} \right)$, using the ratio estimator in adaptive cluster sampling with one auxiliary variable $\left( \hat{\tau}_{R_{x,y_A}}, \hat{\tau}_{R_{z,y_A}} \right)$ and the ratio
estimator in adaptive cluster sampling with two auxiliary variables \( \hat{\tau}_{Rz \cdot A} \) are as follows:

Table 1. The estimated bias of the estimators for the population total of the variable of interest.

<table>
<thead>
<tr>
<th>( n )</th>
<th>( E(\nu) )</th>
<th>( \hat{B}(\hat{\tau}_{Rz \cdot A}) )</th>
<th>( \hat{B}(\hat{\tau}_{Rz \cdot A}) )</th>
<th>( \hat{B}(\hat{\tau}_{Rz \cdot A}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>13.17</td>
<td>-416.58</td>
<td>-349.49</td>
<td>-355.94</td>
</tr>
<tr>
<td>5</td>
<td>29.52</td>
<td>-347.02</td>
<td>-217.78</td>
<td>-251.17</td>
</tr>
<tr>
<td>10</td>
<td>54.25</td>
<td>-281.17</td>
<td>-96.59</td>
<td>-158.90</td>
</tr>
<tr>
<td>15</td>
<td>70.83</td>
<td>-254.57</td>
<td>-46.15</td>
<td>-121.83</td>
</tr>
<tr>
<td>20</td>
<td>87.84</td>
<td>-227.81</td>
<td>-1.08</td>
<td>-85.17</td>
</tr>
<tr>
<td>25</td>
<td>99.41</td>
<td>-225.80</td>
<td>1.43</td>
<td>-82.70</td>
</tr>
<tr>
<td>30</td>
<td>110.01</td>
<td>-220.50</td>
<td>7.24</td>
<td>-76.81</td>
</tr>
</tbody>
</table>

Table 2. The estimated \( MSE \) of the estimators for the population total of the variable of interest.

<table>
<thead>
<tr>
<th>( n )</th>
<th>( M\hat{SE}(\hat{\tau}_{y \cdot A}) )</th>
<th>( M\hat{SE}(\hat{\tau}_{Rz \cdot A}) )</th>
<th>( M\hat{SE}(\hat{\tau}_{Rz \cdot A}) )</th>
<th>( M\hat{SE}(\hat{\tau}_{Rz \cdot A}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>982,826.38</td>
<td>191,711.49</td>
<td>184,808.59</td>
<td>182,529.52</td>
</tr>
<tr>
<td>5</td>
<td>428,116.04</td>
<td>143,203.84</td>
<td>130,289.15</td>
<td>124,493.11</td>
</tr>
<tr>
<td>10</td>
<td>225,191.10</td>
<td>94,235.93</td>
<td>70,377.19</td>
<td>64,720.12</td>
</tr>
<tr>
<td>15</td>
<td>145,934.65</td>
<td>75,642.93</td>
<td>47,530.53</td>
<td>42,521.95</td>
</tr>
<tr>
<td>20</td>
<td>105,545.92</td>
<td>62,212.98</td>
<td>28,899.31</td>
<td>25,656.58</td>
</tr>
<tr>
<td>25</td>
<td>88,082.78</td>
<td>56,850.15</td>
<td>21,342.17</td>
<td>19,290.18</td>
</tr>
<tr>
<td>30</td>
<td>70,263.95</td>
<td>53,483.88</td>
<td>15,320.62</td>
<td>14,994.71</td>
</tr>
</tbody>
</table>

6. Conclusions

Adaptive cluster sampling is an efficient method for sampling rare and hidden clustered populations. An estimator of the population total in adaptive cluster sampling by using the auxiliary variables can reduce the mean square error of the estimator when the auxiliary variables are correlated with the response of interest. The numerical study showed that the estimated mean square error for the estimators of population total using
the auxiliary variable are less than the estimated mean square error for the \( \hat{\tau}_{y-A} \).
Moreover the estimated mean square error of the estimator of population total from two auxiliary variables, \( \hat{\tau}_{R_{yx}} \), is less than the estimated mean square error of the estimator of population total from only one auxiliary variable, \( \hat{\tau}_{R_{x}} \) and \( \hat{\tau}_{R_{z}} \).

7. Acknowledgements

This research was supported by Research Institute of Mahasarakham University. We would also like to profoundly thanks Mr. Paveen Chutiman for his programming advice.

Appendix A

A.1. Let \( \delta_y = \left( \frac{\hat{\tau}_{y-A} - \tau_y}{\tau_y} \right) \). The expected value is shown as

\[
E(\delta_y) = E\left( \frac{\hat{\tau}_{y-A}}{\tau_y} \right) - E\left( \frac{\tau_y}{\tau_y} \right)
= \frac{1}{\tau_y} E\left( \frac{\hat{\tau}_{y-A}}{\tau_y} \right) - 1
\]

\[
E\left( \frac{\hat{\tau}_{y-A}}{\tau_y} \right) = \tau_y \quad (\hat{\tau}_{y-A} \text{ is unbiased estimator of } \tau_y)
E(\delta_y) = 0.
\]

Similarly, \( E(\delta_x) = 0 \) and \( E(\delta_z) = 0 \).

A.2. \( E\left( \delta_y^2 \right) \) is shown as

\[
E\left( \delta_y^2 \right) = V(\delta_y) + \left[ E(\delta_y) \right]^2
= V\left( \frac{\hat{\tau}_{y-A} - \tau_y}{\tau_y} \right)
= \frac{1}{\tau_y^2} V\left( \hat{\tau}_{y-A} \right)
\]
\[
\frac{1}{\tau_y^2} \frac{N^2 (N-n)}{Nn} \sum_{i=1}^{N} \left( \frac{w_y}{N} - \frac{\tau_y}{N} \right)^2 \\
= \frac{1}{\tau_y^2} \frac{N^2 (N-n)}{Nn} S_y^{2-A}
\]

\[
E\left( \delta^2_y \right) = \lambda A_y^{2-A},
\]

where \( \lambda = \frac{N^2 (N-n)}{Nn} \) and \( A_y^{2-A} = \frac{S_y^{2-A}}{\tau_y^2} \).

Similarly, \( E\left( \delta^2_x \right) = \lambda A_x^{2-A} \) and \( E\left( \delta^2_z \right) = \lambda A_z^{2-A} \).

References


