Article

Not peer-reviewed version

Enhanced Ratio-Type Estimators in Adaptive Cluster Sampling Using Jackknife Method

Supawadee Wichitchan , Athipakon Nathomthong , Pannarat Guayjarernpanishk , Nipaporn Chutiman *

Posted Date: 8 May 2025

doi: 10.20944/preprints202505.0517.v1

Keywords: Adaptive cluster sampling; auxiliary information; ratio estimator; jackknife method

doi:10.20944/preprints202505.0517.v1

Article

Enhanced Ratio-Type Estimators in Adaptive Cluster Sampling Using Jackknife Method

Supawadee Wichitchan [1], Athipakon Nathomthong [1], Pannarat Guayjarernpanishk [2] and Nipaporn Chutiman [1,*]

[1] Department of Mathematics, Faculty of Science, Mahasarakham University, Maha Sarakham, Thailand
[2] Faculty of Interdisciplinary Studies, Nong Khai Campus, Khon Kaen University, Nong Khai, Thailand
* Correspondence: nipaporn.c@msu.ac.th

Abstract: Adaptive cluster sampling is a methodology designed for data collection in contexts where the population is rare and spatially clustered. This approach has been effectively applied in various disciplines, including epidemiology and resource management. The present study introduces novel estimators that incorporate auxiliary variable information to improve estimation efficiency. These estimators are developed using the Jackknife resampling technique, which is employed to enhance the performance of ratio-type estimators. Theoretical properties, including bias and mean square error (MSE), are derived, and a simulation study is conducted to validate the theoretical findings. Results demonstrate that the proposed estimators consistently outperform conventional estimators that do not utilize auxiliary variables across all network sample sizes. Furthermore, in several scenarios, the proposed estimators also exhibit superior efficiency compared to existing ratio estimators that do incorporate auxiliary information.

Keywords: Adaptive cluster sampling; auxiliary information; ratio estimator; jackknife method

MSC: 62D05; 62F40

1. Introduction

Sampling methods are employed when it is impractical to collect data from an entire population. For statistical inferences to be valid, the selected sample must be obtained through probability-based sampling techniques. One of the most commonly used probability sampling methods is simple random sampling. The population mean for a variable of interest, denoted by μ_y, is typically estimated using the sample mean $\bar{y} = \sum_{i=1}^{n} y_i / n$, where n is the sample size. This estimator is known to be unbiased. A primary objective in sampling theory is to improve the efficiency of such estimators.

In many cases, the estimation efficiency can be enhanced by utilizing an auxiliary variable that is correlated with the variable of interest. When both variables exhibit a high positive correlation, the ratio estimator, which incorporates the auxiliary variable's mean, is widely adopted. Numerous studies have proposed improvements to the ratio estimator in the context of simple random sampling. For example, Sisodia and Dwivedi [1] introduced a modified ratio estimator based on the coefficient of variation of the auxiliary variable. Singh and Tailor [2] proposed an estimator that incorporates the population correlation coefficient between the target and auxiliary variables. Yadav et al. [3] developed ratio-cum-product estimators, while Jerajuddin and Kishun [4] enhanced ratio estimators by considering sample size. Soponviwatkul and Lawson [5] proposed further refinements by incorporating the coefficient of variation, correlation coefficient, and regression coefficient.

However, in situations where the population is both rare and clustered, simple random sampling may be suboptimal. To address this, Thompson [6] introduced adaptive cluster sampling (ACS) in 1990. In ACS, an initial sample is selected using simple random sampling without

 doi:10.20944/preprints202505.0517.v1

replacement. If a unit in this initial sample satisfies a pre-specified condition for the variable of interest, its neighboring units are added to the sample. This expansion continues iteratively until no additional units meet the condition. The collection of initial and subsequently added units forms a network. Units that do not meet the condition are referred to as edge units. The union of a network and its edge units constitutes a cluster. If the initial unit fails to satisfy the condition, it remains a singleton network. For this study, neighborhoods are defined as the four orthogonally adjacent units (up, down, left, and right), with mutual neighborhood relationships assumed.

Thompson also proposed an estimator and demonstrated that ACS yields improved efficiency in clustered populations. Analogous to simple random sampling, incorporating auxiliary variable information in ACS can further improve estimator performance. Chao [7] introduced a ratio estimator for ACS, while Dryver and Chao [8] introduced modified ratio estimators. Chutiman and Kum-phon [9] presented a ratio estimator using two auxiliary variables. Chutiman[10] and Yadav et al. [11] proposed ratio estimators based on population parameters, including the coefficient of variation, kurtosis, skewness, and correlations with auxiliary variables. Chaudhry and Hanif [12,13] proposed generalized exponential-type estimators, while Bhat et al. [14] developed a generalized class of ratio-type estimators.

Increasing the auxiliary variable information is the primary focus of the development of the parameter estimators discussed above. This study presents the development of Chao's ratio-type estimator in adaptive cluster sampling, which uses the Jackknife method to leverage data from a single auxiliary variable, specifically the auxiliary variable's mean. Section 2 outlines relevant estimators in ACS, Section 3 introduces the proposed Jackknife-based estimators, Section 4 presents simulation results, and Section 5 concludes the study.

2. Adaptive Cluster Sampling

In adaptive cluster sampling, an initial sample of units is selected using simple random sampling without replacement.

Let n represent the initial sample size and ν be the final sample size. Let ψ_i denote the network that includes unit i, and let m_i represent the number of units in that network.

Let $\left(w_y\right)_i$ be the average of the y-value in the network that includes the initial sample unit i, that is, $\left(w_y\right)_i = \frac{1}{m_i}\sum_{j\in\psi_i} y_j$.

The Hansen-Hurwitz estimator of the population mean for the variable of interest is [15]:

$$\bar{w}_y = \frac{1}{n}\sum_{i=1}^{n}\left(w_y\right)_i . \tag{1}$$

The mean square error (MSE) of $\bar{w}_y$ is:

$$MSE\left(\bar{w}_y\right) = \left(1-f_n\right)\frac{S_{wy}^2}{n}, \tag{2}$$

where $f_n = \frac{n}{N}$ and $S_{wy}^2 = \frac{1}{(N-1)}\sum_{i=1}^{n}\left[\left(w_y\right)_i - \mu_y\right]^2$.

Let x be the auxiliary variable. The population mean of x is μ_x and $\left(w_x\right)_i$ is the average of the auxiliary variable in the network that includes the initial sample unit i, that is, $\left(w_x\right)_i = \frac{1}{m_i}\sum_{j\in\psi_i} x_j$. The modified Hansen-Hurwitz estimator of the population mean of the auxiliary variable is:

$$\bar{w}_x = \frac{1}{n}\sum_{i=1}^{n}\left(w_x\right)_i \tag{3}$$

Let R be the population ratio between y and x, $R = \frac{\mu_y}{\mu_x}$. Chao [7] introduced the ratio estimator of the population mean, which is

$$\overline{w}_R = \left(\frac{\overline{w}_y}{\overline{w}_x}\right)\mu_x = \hat{R}\mu_x, \quad (4)$$

where $\overline{w}_R$ is a biased estimator of μ_y. The bias of $\overline{w}_R$ is:

$$Bias(\overline{w}_R) = \frac{\mu_y}{n}(1-f_n)C_{wx}\left(C_{wx} - \rho_{wx \cdot wy}C_{wy}\right), \quad (5)$$

where $\rho_{wx \cdot wy} = \frac{S_{wx \cdot wy}}{S_{wx}S_{wy}}$, $C_{wx} = \frac{S_{wx}}{\mu_x}$, $S_{wy}^2 = \frac{1}{(N-1)}\sum_{i=1}^{n}\left[(w_y)_i - \mu_y\right]^2$, $S_{wx}^2 = \frac{1}{(N-1)}\sum_{i=1}^{n}\left[(w_x)_i - \mu_x\right]^2$ and $S_{wx \cdot wy} = \frac{1}{(N-1)}\sum_{i=1}^{n}\left[(w_x)_i - \mu_x\right]\left[(w_y)_i - \mu_y\right]$.

The mean square error (MSE) of $\overline{w}_R$ is:

$$MSE(\overline{w}_R) = \frac{\mu_y^2}{n}(1-f_n)\left(C_{wx}^2 + C_{wy}^2 - 2\rho_{wx \cdot wy}C_{wx}C_{wy}\right) \quad (6)$$

3. Proposed Estimators in Adaptive Cluster Sampling Using the Jackknife Method

Motivated by Banerjee and Tiwari [16], Quenouille's Jackknife method [17] was applied to propose the estimators. The sample network of size n is randomly partitioned into two groups, each of size m = n/2.

The proposed estimators are:

1) $\overline{w}'_R = \frac{\overline{w}_R^{(1)} + \overline{w}_R^{(2)}}{2}$, (7)

where $\overline{w}_R^{(i)} = \frac{\overline{w}_y^{(i)}}{\overline{w}_x^{(i)}}\mu_x$ and $i = 1,2$, where $\overline{w}_y^{(i)}$ and $\overline{w}_x^{(i)}$ are the sample means based on group i of size m, for the y-variable and x-variable, respectively.

2) $\overline{w}''_R = \frac{\overline{w}_R - K\overline{w}'_R}{1-K}$, where $K = \frac{Bias(\overline{w}_R)}{Bias(\overline{w}'_R)}$. (8)

3) $\overline{w}_{JK} = \frac{1}{n}\sum_{i=1}^{n}\overline{w}_{R(i)}$, (9)

where $\overline{w}_{R(i)} = \frac{\overline{w}_{y(i)}}{\overline{w}_{x(i)}}\mu_x$ is the ratio estimator of the population mean for the y-variable in the delete network i, and $\overline{w}_{y(i)}$, $\overline{w}_{x(i)}$ are the modified Hansen-Hurwitz estimators of the population mean in the delete network i for the y-variable and x-variable, respectively.

The bias and MSE of each estimator are as follows:

The first estimator: $\overline{w}'_R = \frac{\overline{w}_R^{(1)} + \overline{w}_R^{(2)}}{2}$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 May 2025 doi:10.20944/preprints202505.0517.v1

Let $\varepsilon_0^{(i)} = \dfrac{\overline{w}_y^{(i)} - \mu_y}{\mu_y}$ and $\varepsilon_1^{(i)} = \dfrac{\overline{w}_x^{(i)} - \mu_x}{\mu_x}$.

Where $i = 1$:

$$\overline{w}_R^{(1)} = \frac{\overline{w}_y^{(1)}}{\overline{w}_x^{(1)}}\mu_x = \left(1+\varepsilon_0^{(1)}\right)\left(1+\varepsilon_1^{(1)}\right)^{-1}\mu_y$$

Assuming $\left|\varepsilon_1^{(1)}\right| < 1$, the term $\left(1+\varepsilon_1^{(1)}\right)^{-1}$ can be expanded as an infinite series.

$$\overline{w}_R^{(1)} = \mu_y\left(1+\varepsilon_0^{(1)}\right)\left[1-\varepsilon_1^{(1)}+\left(\varepsilon_1^{(1)}\right)^2-\ldots\right]$$

$$= \mu_y\left[1+\varepsilon_0^{(1)}-\varepsilon_1^{(1)}+\left(\varepsilon_1^{(1)}\right)^2-\varepsilon_0^{(1)}\varepsilon_1^{(1)}\ldots\right]$$

Therefore, $\overline{w}_R^{(1)} - \mu_y \approx \mu_y\left[\varepsilon_0^{(1)}-\varepsilon_1^{(1)}+\left(\varepsilon_1^{(1)}\right)^2-\varepsilon_0^{(1)}\varepsilon_1^{(1)}\right]$.

The bias of $\overline{w}_R^{(1)}$ is given by:

$$E\left[\overline{w}_R^{(1)}-\mu_y\right] = \mu_y E\left[\varepsilon_0^{(1)}-\varepsilon_1^{(1)}+\left(\varepsilon_1^{(1)}\right)^2-\varepsilon_0^{(1)}\varepsilon_1^{(1)}\right],$$

where $E\left(\varepsilon_0^{(1)}\right) = E\left(\varepsilon_1^{(1)}\right) = 0$, $E\left[\left(\varepsilon_1^{(1)}\right)^2\right] = \dfrac{(1-f_m)}{m}C_{wx}^2$

$$E\left(\varepsilon_0^{(1)}\varepsilon_1^{(1)}\right) = \frac{(1-f_m)}{m}\rho_{wx\cdot wy}C_{wx}C_{wy} \text{ and } f_m = \frac{m}{N}$$

Therefore, $Bias\left(\overline{w}_R^{(1)}\right) = E\left[\overline{w}_R^{(1)}-\mu_y\right] = \dfrac{(1-f_m)}{m}\mu_y C_{wx}\left(C_{wx}-\rho_{wx\cdot wy}C_{wy}\right)$.

The bias of $\overline{w}_R^{(2)}$ is derived in the same way as that of $\overline{w}_R^{(1)}$, and the $Bias\left(\overline{w}_R^{(2)}\right)$ is equal to the $Bias\left(\overline{w}_R^{(1)}\right)$.

$$E\left(\overline{w}_R' - \mu_y\right) = E\left[\frac{\overline{w}_R^{(1)}+\overline{w}_R^{(2)}}{2}-\mu_y\right]$$

$$= \frac{1}{2}\left[E\left(\overline{w}_R^{(1)}-\mu_y\right)+E\left(\overline{w}_R^{(2)}-\mu_y\right)\right]$$

Therefore, the bias of $\overline{w}_R'$ is:

$$Bias\left(\overline{w}_R'\right) = E\left(\overline{w}_R'-\mu_y\right) = \frac{(1-f_m)}{m}\mu_y C_{wx}\left(C_{wx}-\rho_{wx\cdot wy}C_{wy}\right) \quad (10)$$

For the MSE of $\overline{w}_R'$,

$$MSE\left(\overline{w}_R'\right) = E\left(\overline{w}_R'-\mu_y\right)^2 = E\left[\frac{\overline{w}_R^{(1)}+\overline{w}_R^{(2)}}{2}-\mu_y\right]^2$$

$$= \frac{1}{4}\left\{E\left(\overline{w}_R^{(1)}-\mu_y\right)^2+E\left(\overline{w}_R^{(2)}-\mu_y\right)^2+2E\left[\left(\overline{w}_R^{(1)}-\mu_y\right)\left(\overline{w}_R^{(2)}-\mu_y\right)\right]\right\}$$

where $E\left(\overline{w}_R^{(1)}-\mu_y\right)^2 = MSE\left(\overline{w}_R^{(1)}\right) = \dfrac{\mu_y^2}{m}(1-f_m)\left(C_{wx}^2+C_{wy}^2-2\rho_{wx\cdot wy}C_{wx}C_{wy}\right)$

$$E\left(\overline{w}_R^{(2)}-\mu_y\right)^2 = MSE\left(\overline{w}_R^{(2)}\right) = \frac{\mu_y^2}{m}(1-f_m)\left(C_{wx}^2+C_{wy}^2-2\rho_{wx\cdot wy}C_{wx}C_{wy}\right)$$

$$E\left[\left(\overline{w}_R^{(1)}-\mu_y\right)\left(\overline{w}_R^{(2)}-\mu_y\right)\right] = COV\left(\overline{w}_R^{(1)},\overline{w}_R^{(2)}\right).$$

 doi:10.20944/preprints202505.0517.v1

The MSE of $\overline{w}_R'$ is:

$$MSE\left(\overline{w}_R'\right)=\frac{1}{2}\left\{\frac{\mu_y^2}{m}\left(1-f_m\right)\left(C_{wx}^2+C_{wy}^2-2\rho_{wx\cdot wy}C_{wx}C_{wy}\right)+COV\left(\overline{w}_R^{(1)},\overline{w}_R^{(2)}\right)\right\} \tag{11}$$

The second estimator: $\overline{w}_R''=\dfrac{\overline{w}_R-K\overline{w}_R'}{1-K}$, where $K=\dfrac{Bias\left(\overline{w}_R\right)}{Bias\left(\overline{w}_R'\right)}$

From $Bias\left(\overline{w}_R\right)=\dfrac{\mu_y}{n}\left(1-f_n\right)C_{wx}\left(C_{wx}-\rho_{wx\cdot wy}C_{wy}\right)$ and

$Bias\left(\overline{w}_R'\right)=\dfrac{\left(1-f_m\right)}{m}\mu_yC_{wx}\left(C_{wx}-\rho_{wx\cdot wy}C_{wy}\right)$, it follows that $K=\dfrac{\left(1-f_n\right)/n}{\left(1-f_m\right)/m}=\dfrac{N-n}{2N-n}$, and $\overline{w}_R''$ is an unbiased estimator of μ_y to the first order of approximation (based on Banerjee and Tiwari [15]).

For the MSE of $\overline{w}_R''$,

$$MSE\left(\overline{w}_R''\right)=E\left(\overline{w}_R''-\mu_y\right)^2=E\left(\frac{\overline{w}_R-K\overline{w}_R'}{1-K}-\mu_y\right)^2$$

$$=\frac{1}{\left(1-K\right)^2}\left\{E\left(\overline{w}_R-\mu_y\right)^2+K^2E\left(\overline{w}_R'-\mu_y\right)^2-2KE\left[\left(\overline{w}_R-\mu_y\right)\left(\overline{w}_R'-\mu_y\right)\right]\right\},$$

where $E\left(\overline{w}_R-\mu_y\right)^2=MSE\left(\overline{w}_R\right)=\dfrac{\mu_y^2}{n}\left(1-f_n\right)\left(C_{wx}^2+C_{wy}^2-2\rho_{wx\cdot wy}C_{wx}C_{wy}\right)$,

$$E\left(\overline{w}_R'-\mu_y\right)^2=MSE\left(\overline{w}_R'\right)=\frac{1}{2}\left\{\frac{\mu_y^2}{m}\left(1-f_m\right)\left(C_{wx}^2+C_{wy}^2-2\rho_{wx\cdot wy}C_{wx}C_{wy}\right)+COV\left(\overline{y}_R^{(1)},\overline{y}_R^{(2)}\right)\right\},$$

and

$$E\left[\left(\overline{w}_R-\mu_y\right)\left(\overline{w}_R'-\mu_y\right)\right]=\frac{1}{2}\left\{COV\left(\overline{w}_R,\overline{w}_R^{(1)}\right)+COV\left(\overline{w}_R,\overline{w}_R^{(2)}\right)\right\}.$$

Therefore, The MSE of $\overline{w}_R''$ is

$$MSE\left(\overline{w}_R''\right)=\frac{1}{\left(1-K\right)^2}\left\{MSE\left(\overline{w}_R\right)+K^2MSE\left(\overline{w}_R'\right)-K\left[COV\left(\overline{w}_R,\overline{w}_R^{(1)}\right)+COV\left(\overline{w}_R,\overline{w}_R^{(2)}\right)\right]\right\}. \tag{12}$$

The Third estimator: $\overline{w}_{JK}=\dfrac{1}{n}\sum_{i=1}^{n}\overline{w}_{R(i)}$

Let $\varepsilon_{0(i)}=\dfrac{\overline{w}_{y(i)}-\mu_y}{\mu_y}$ and $\varepsilon_{1(i)}=\dfrac{\overline{w}_{x(i)}-\mu_x}{\mu_x}$.

Where i =1:

$$\overline{w}_{R(1)}=\frac{\overline{w}_{x(1)}}{\overline{w}_{y(1)}}\mu_x=\left(1+\varepsilon_{0(1)}\right)\left(1+\varepsilon_{1(1)}\right)^{-1}\mu_y$$

Assuming $\left|\varepsilon_{1(1)}\right|<1$, the term $\left(1+\varepsilon_{1(1)}\right)^{-1}$ can be expanded as an infinite series.

$$\overline{w}_{R(1)}=\mu_y\left(1+\varepsilon_{0(1)}\right)\left[1-\varepsilon_{1(1)}+\left(\varepsilon_{1(1)}\right)^2-\ldots\right]$$

$$=\mu_y\left[1+\varepsilon_{0(1)}-\varepsilon_{1(1)}+\left(\varepsilon_{1(1)}\right)^2-\varepsilon_{0(1)}\varepsilon_{1(1)}\cdots\right]$$

$$\overline{w}_{R(1)}-\mu_y\approx\mu_y\left[\varepsilon_{0(1)}-\varepsilon_{1(1)}+\left(\varepsilon_{1(1)}\right)^2-\varepsilon_{0(1)}\varepsilon_{1(1)}\right].$$

The bias of $\overline{w}_{R(1)}$ is given by:

$$E\left[\bar{w}_{R(1)}-\mu_y\right]=\mu_y E\left[\varepsilon_{0(1)}-\varepsilon_{1(1)}+\left(\varepsilon_{1(1)}\right)^2-\varepsilon_{0(1)}\varepsilon_{1(1)}\right],$$

where $E\left(\varepsilon_{0(1)}\right)=E\left(\varepsilon_{1(1)}\right)=0$, $E\left[\left(\varepsilon_{1(1)}\right)^2\right]=\frac{\left(1-f_{(n-1)}\right)}{n-1}C_{wx}^2$, $f_{(n-1)}=\frac{n-1}{N}$,

and $E\left(\varepsilon_{0(1)}\varepsilon_{1(1)}\right)=\frac{\left(1-f_{(n-1)}\right)}{n-1}\rho_{wx\cdot wy}C_{wx}C_{wy}$.

$$Bias\left(\bar{w}_{R(1)}\right)=E\left[\bar{w}_{R(1)}-\mu_y\right]=\frac{\left(1-f_{(n-1)}\right)}{n-1}\mu_y C_{wx}\left(C_{wx}-\rho_{wx\cdot wy}C_{wy}\right).$$

The bias of $\bar{w}_{R(2)}$ is derived in the same way as that of $\bar{w}_{R(1)}$, and the $Bias\left(\bar{w}_{R(2)}\right)$ is equal to the $Bias\left(\bar{w}_{R(1)}\right)$. Therefore, the bias of $\bar{w}_{JK}$ is:

$$Bias\left(\bar{w}_{JK}\right)=E\left[\bar{w}_{JK}-\mu_y\right]=\frac{1}{n}\left[\sum_{i=1}^{n}E\left(\bar{w}_{R(i)}-\mu_y\right)\right]$$

$$=\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\left(1-f_{(n-1)}\right)}{n-1}\mu_y C_{wx}\left(C_{wx}-\rho_{wx\cdot wy}C_{wy}\right)\right\}$$

$$Bias\left(\bar{w}_{JK}\right)=\frac{\left(1-f_{(n-1)}\right)}{n-1}\mu_y C_{wx}\left(C_{wx}-\rho_{wx\cdot wy}C_{wy}\right). \qquad (13)$$

For the MSE of $\bar{w}_{JK}$,

$$MSE\left(\bar{w}_{JK}\right)=E\left(\bar{w}_{JK}-\mu_y\right)^2=E\left[\frac{1}{n}\sum_{i=1}^{n}\bar{w}_{R(i)}-\mu_y\right]^2$$

$$=\frac{1}{n^2}E\left[\sum_{i=1}^{n}\left(\bar{w}_{R(i)}-\mu_y\right)\right]^2$$

$$=\frac{1}{n^2}\left\{\sum_{i=1}^{n}E\left(\bar{w}_{R(i)}-\mu_y\right)^2+\sum_{i\neq j=1}^{n}\sum_{j=1}^{n}E\left[\left(\bar{w}_{R(i)}-\mu_y\right)\left(\bar{w}_{R(j)}-\mu_y\right)\right]\right\},$$

where $E\left(\bar{w}_{R(i)}-\mu_y\right)^2=MSE\left(\bar{w}_{R(i)}\right)=\frac{\mu_y^2}{n-1}\left(1-f_{(n-1)}\right)\left(C_{wx}^2+C_{wy}^2-2\rho_{wx\cdot wy}C_{wx}C_{wy}\right)$

and $E\left[\left(\bar{w}_{R(i)}-\mu_y\right)\left(\bar{w}_{R(j)}-\mu_y\right)\right]=COV\left(\bar{w}_{R(i)},\bar{w}_{R(j)}\right)$.

Therefore, the MSE of $\bar{w}_{JK}$ is :

$$MSE\left(\bar{w}_{JK}\right)=\frac{1}{n^2}\left\{\frac{\mu_y^2}{n-1}\left(1-f_{(n-1)}\right)\left(C_{wx}^2+C_{wy}^2-2\rho_{wx\cdot wy}C_{wx}C_{wy}\right)+\sum_{i\neq j=1}^{n}\sum_{j=1}^{n}COV\left(\bar{w}_{R(i)},\bar{w}_{R(j)}\right)\right\}.$$

(14)

4. Simulation Study and Discussion

The populations for both the auxiliary variable and the variable of interest, as used in Chao [7], were generated using a linked-pairs process in conjunction with a bivariate Poisson cluster process. The resulting population comprised a 20 × 20 grid, yielding a total of 400 units. The mean of the variable of interest (y) in the population was 0.635, and the Pearson correlation coefficient between the auxiliary variable (x) and y was 0.707035. For each simulation iteration, initial sample units were selected via simple random sampling without replacement. The expansion criterion for adaptive cluster sampling was defined by the condition $C=\{y:y>0\}$. A total of 10,000 iterations were conducted for each estimator under investigation. The number of initial networks n was varied across

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 May 2025 doi:10.20944/preprints202505.0517.v1

the values 4, 8, 10, 16, 20, 26, 30, 40, 50, 100, and 200. The expected final sample size $E(\nu)$ was computed as follows: $E(\nu)=\frac{1}{10,000}\sum_{i=1}^{10,000}\nu_i$.

The estimated absolute relative bias was defined as:

$$ARB(\overline{w})=\frac{1}{10,000}\sum_{i=1}^{10,000}\frac{\left|\overline{w}_i-\mu_y\right|}{\mu_y}$$

The estimated mean square error of the estimator was defined as:

$$M\hat{S}E(\overline{w})=\frac{1}{10,000}\sum_{i=1}^{10,000}\left(\overline{w}_i-\mu_y\right)^2$$

The percentage relative efficiency of the proposed estimator, compared with $\overline{w}_y$, was defined as: $PRE(\overline{w})_{pro}=\frac{M\hat{S}E(\overline{w}_y)}{M\hat{S}E(\overline{w}_{pro})}\times 100$.

The estimated absolute relative bias, estimated mean square error (MSE), and percentage relative efficiency of the estimators are presented in Tables 1–3.

Table 1. The estimated absolute relative bias of the estimators for the population mean of the variable of interest.

n	$E(\nu)$	$ARB(\overline{w}_R)$	$ARB(\overline{w}'_R)$	$ARB(\overline{w}_{JK})$
4	6.8705	0.5291	0.7104	0.6129
8	12.8042	0.3045	0.5526	0.3550
10	15.8498	0.2046	0.4560	0.2390
16	24.8431	0.0651	0.2804	0.0790
20	30.9154	0.0254	0.1957	0.0332
26	38.9001	0.0337	0.1386	0.0354
30	43.8715	0.0263	0.0997	0.0278
40	56.7309	0.0077	0.0389	0.0084
50	68.5383	0.0100	0.0274	0.0103
100	120.1685	0.0018	0.0083	0.0019
200	215.1886	0.0031	0.0004	0.0003

Note: $\overline{w}_y$ and $\overline{w}''_R$ are unbiased estimators. Therefore, the absolute relative bias is not presented.

Table 2. The estimated MSE of the estimators for the population mean of the variable of interest.

n	$E(\nu)$	Estimators that do not use auxiliary variable information		Estimators use auxiliary variable information			
		$\overline{y}$	$\overline{w}_y$	$\overline{w}_R$	$\overline{w}'_R$	$\overline{w}''_R$	$\overline{w}_{JK}$
4	6.8705	10.8216	1.1305	0.2777	0.2730	0.3777	0.2625
8	12.8042	10.8059	0.5291	0.2115	0.1986	0.3679	0.1929
10	15.8498	10.7229	0.4623	0.1792	0.1705	0.3361	0.1649
16	24.8431	9.6113	0.2606	0.1158	0.1149	0.2407	0.1102
20	30.9154	9.2723	0.2122	0.0990	0.0962	0.2040	0.0938
26	38.9001	7.7315	0.1790	0.0873	0.0863	0.1665	0.0856
30	43.8715	6.9521	0.1315	0.05270	0.0521	0.0906	0.0517
40	56.7309	5.9159	0.1004	0.04060	0.0495	0.0651	0.0401
50	68.5383	4.7074	0.0717	0.02520	0.0322	0.0292	0.0215
100	120.1685	1.5727	0.0238	0.0078	0.0091	0.0080	0.0078
200	215.1886	0.2266	0.0076	0.0023	0.0027	0.0024	0.0023

Table 3. The percentage relative efficiency of the estimators compared with $\overline{w}_y$

n	$E(v)$	The PRE of the estimators compared with $\overline{w}_y$				
		$\overline{w}_y$	$\overline{w}_R$	$\overline{w}'_R$	$\overline{w}''_R$	$\overline{w}_{JK}$
4	6.8705	100	407.1231	414.1172	299.2905	430.7640
8	12.8042	100	250.1418	266.3881	143.8358	274.2445
10	15.8498	100	257.9432	271.0884	137.5487	279.4969
16	24.8431	100	224.9978	226.8193	108.2551	236.5163
20	30.9154	100	214.2381	220.6093	103.9796	226.1834
26	38.9001	100	205.0630	207.3911	107.5389	209.2577
30	43.8715	100	249.4971	252.5163	145.1634	254.5208
40	56.7309	100	247.4008	203.0738	154.3736	250.6114
50	68.5383	100	284.5574	222.8856	245.8162	333.5505
100	120.1685	100	306.0567	260.4167	296.5044	306.0567
200	215.1886	100	324.7863	284.6442	323.4043	324.7863

Discussion

Based on the data studied, the variable of interest is positively correlated with the auxiliary variable. The results from the simulation data are presented as follows:

Table 1 presents the estimated absolute relative bias of the biased estimators, namely $\overline{w}_R$, $\overline{w}'_R$, and $\overline{w}_{JK}$. It can be observed that as the sample size increases, the estimated absolute relative bias for all estimators decreases and approaches zero.

Table 2 shows that the estimators incorporating auxiliary variable information—given the positive correlation with the variable of interest—consistently yield lower MSEs compared to estimators that do not use such information. Notably, $\overline{w}'_R$ achieves a lower MSE than the traditional ratio estimator $\overline{w}_R$ when the network sample size is small. Among all estimators, $\overline{w}_{JK}$ provides the lowest MSE across all network sample sizes. Although the estimator $\overline{w}''_R$ is unbiased, its MSE is higher than that of $\overline{w}_R$ despite being lower than estimators that do not use auxiliary information.

Table 3 presents the percentage relative efficiency of each estimator compared to $\overline{w}_y$. The estimator $\overline{w}_{JK}$ consistently exhibits the highest percentage relative efficiency. Moreover, as the sample size increases, the efficiency of $\overline{w}_{JK}$ converges with that of the traditional ratio estimator $\overline{w}_R$.

5. Conclusions

Adaptive cluster sampling (ACS) is particularly effective for studying rare and spatially clustered populations. This research proposed three enhanced ratio-type estimators for ACS, building on Chao's [7] original ratio estimator and employing the Jackknife method to reduce bias and improve efficiency. Analytical derivations of bias and MSE were provided for each estimator, and their performance was evaluated through extensive simulation. The simulation results demonstrated that all three proposed estimators outperformed conventional estimators that do not utilize auxiliary variable information. Specifically, $\overline{w}'_R$ proved to be more efficient than Chao's estimator for small network sample sizes, while $\overline{w}_{JK}$ exhibited superior efficiency for both small and moderate sample sizes. In large-sample settings, the efficiency of $\overline{w}_{JK}$ became comparable to that of the traditional ratio estimator. Although $\overline{w}''_R$ is an unbiased estimator, its efficiency was the lowest among the estimators that incorporate auxiliary variable information.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 May 2025 doi:10.20944/preprints202505.0517.v1

Author Contributions: Conceptualization, S.W and N.C.; methodology, S.W.; software, N.C. and P.G; investigation, A.N.; writing—original draft preparation, N.C. and A.N.; writing—review and editing, P.G and S.W..; funding acquisition, N.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research project was financially supported by Mahasarakham University.

Data Availability Statement: Data are contained within the article.

Acknowledgments: The authors would like to thank the editor and the referees for their valuable feedback and insightful suggestions.

Conflicts of Interest: The authors have no conflicts of interest to declare that are relevant to the content of this article.

Appendix A

Figures A1 and A2 display the population distributions for the variable of interest and the auxiliary variable, respectively, as generated according to Chao [7].

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
0	0	0	0	0	0	0	0	0	1	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
1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1
2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0
1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	6	0	1
0	0	1	0	0	0	0	0	6	18	3	0	1	0	0	0	9	27	9	0
1	0	0	0	0	0	0	1	14	46	11	0	0	0	0	1	8	25	14	1
0	0	0	0	0	0	0	0	4	7	2	0	0	0	0	0	1	2	1	0

Figure A.1. The population of the variable of interest (y)

0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	1	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	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	1	0	0	0	0	0	1	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	1	0	0	2
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	1	0	0	1	0	0	0	1	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	1	0	0	0
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	1	2	0	1	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	4	6	0	1	0	0	0	4	3	4	2	3
0	0	1	0	0	0	0	4	6	11	4	3	1	0	0	2	6	8	3	7
0	0	0	0	0	0	1	4	4	9	11	4	0	0	2	1	8	10	0	2
0	0	0	0	0	0	1	2	7	4	3	1	1	0	0	0	0	2	2	0

Figure A.2. The population of the auxiliary variable (x)

References

1. Sisodia, B. V. S. and Dwivedi, V. K. A. Modified ratio Estimator Using Coefficient of Variation of Auxiliary Variable. *Journal of Indian Society Agricultural Statistics*,**1981,** 33, 13-18.
2. Singh, H. P. & Tailor, R. (2003). Use of Known Correlation Coefficient in Estimating the Finite Population Mean. *Statistics in Transition*, **2003,** 6, 555-560.
3. Yadav, S. K., Subramani, J., Mishra, S. S. and Shukla, A. K. Improved Ratio-Cum- Product Estimators of Population Mean Using Known Population Parameters of Auxiliary Variables. *American Journal of Operational Research*, **2016,** 6(2), 48-54. https://doi:10.5923/j.ajor.20160602.03
4. Jerajuddin, M. and Kishun, J. Modified Ratio Estimators for Population Mean Using Size of the Sample, Selected from Population. International *Journal of Scientific Research in Science, Engineering and Technology*, **2016,** 2, 10-16.
5. Soponviwatkul, K. and Lawson, N. New Ratio Estimators for Estimating Population Mean in Simple Random Sampling using a Coefficient of Variation, Correlation Coefficient and a Regression Coefficient. *Gazi University Journal of Science*, **2017**, 30, 610-621.
6. Thompson, S.K. Adaptive cluster sampling. *J. Am. Statist. Assoc.* **1990**, 85(412), 1050-1059. https://doi.org/10.2307/2289601.
7. Chao, C.T. Ratio estimation on adaptive cluster sampling. *Journal of Chinese Statistical Association* **2004,** 42, 307–27. https://doi.org/10.29973/JCSA.200409.0006.
8. Dryver, A.L. and Chao, C.T. Ratio estimators in adaptive cluster sampling. *Environmetric* **2007**, 18, 607-620. https://doi.org/10.1002/env.838.
9. Chutiman, N. and Kumphon, B. Ratio estimator using two auxiliary variables for adaptive cluster sampling. *Thailand Statistician* **2008**, 6(2), 241–256. https://ph02.tci-thaijo.org/index.php/thaistat/article/view/34339.
10. Chutiman, N. Adaptive cluster sampling using auxiliary variable. *Journal of Mathematics and Statistics***2013**, 9(3), 249-255. https://doi.org/10.3844/jmssp.2013.249.255.
11. Yadav, S.K., Misra, S. and Mishra, S. Efficient estimator for population variance using auxiliary variable. *American Journal of Operational Research* **2016**, 6, 9-15. https://doi.org/10.5923/j.ajor.20160601.02.
12. Chaudhry, M. S., and M. Hanif. Generalized exponential-cum-exponential estimator in adaptive cluster sampling. *Pakistan Journal of Statistics and Operation Research* **2015,** 11(4), 553–574. https://doi.org/10.18187/pjsor.v11i4.1009.
13. Chaudhry, M. S. and M. Hanif. Generalized difference-cum-exponential estimator in adaptive cluster sampling. *Pakistan Journal of Statistics*, **2017**, 33 (5):335–367.
14. Bhat, A.A , Sharma, M. , Shah, M. and Bhat, M. Generalized ratio type estimator under adaptive cluster sampling. *Journal of Scientific Research* **2023**, 67, 46-51. https://doi.org/10.37398/JSR.2023.670307.
15. Thompson, S.K. Sampling. 3rd ed., John Wiley & Sons, Inc., Hoboken, New Jersey, **2012**; 319-337.
16. Banerjie, J. and Tiwari, N. Improved ratio type estimator using jack-knife method of estimation. *Journal of Reliability and Statistical Studies*, **2011**, 4(1), 53-63.
17. Quenouille, M. H. Notes on Bias in Estimation. *Biometrika*, **1956**, 43(3) 353–360. https://doi.org/10.2307/2332914.