Sains Malaysiana 55(6)(2026): 1077-1087
http://doi.org/10.17576/jsm-2026-5506-12

Ratio Estimators under Stratified Inverse Random Sampling

(Penganggar Nisbah di bawah Pensampelan Rawak Songsang Berstrata)

PRAYAD SANGNGAM[1], WIPAWAN LAOARUN[1,*], PANPHARISA KHONGTHIP[1] & SUREEPORN SUNGSUWAN[2]

[1]*Department of Statistics, Faculty of Science, Silpakorn University, Nakhon Pathom, 73000, Thailand*
[2]*Department of Mathematics, Faculty of Science, Mahanakorn University of Technology, Nong Chok, Bangkok, 10530, Thailand*

Received: 20 January 2026/Accepted: 5 June 2026

ABSTRACT

Stratified inverse random sampling (SIRS) is an effective technique for studying rare populations, as it guarantees a pre-specified number of interested units within each stratum. Although ratio estimation has been extensively studied under various sampling designs, its application under SIRS remains largely unexplored. This study addresses this gap by developing two novel ratio estimators for the population mean under SIRS: the separate ratio estimator and the combined ratio estimator. The contributions of this work are threefold. First, first-order approximations for the bias and mean squared error (MSE) of each estimator are rigorously derived using Taylor series expansions under large-sample assumptions, and consistent sample-based estimators of the MSE are developed. Second, analytical conditions, expressed in terms of population parameters, are established under which the proposed estimators outperform the conventional unbiased estimator. Third, an extensive Monte Carlo simulation study with 50,000 replications is conducted to assess finite-sample performance. The simulation results demonstrate that the separate ratio estimator consistently attains the smallest MSE across all scenarios, achieving efficiency gains of up to 108.9% relative to the unbiased estimator. Moreover, all proposed estimators exhibit a decrease in absolute bias, MSE, skewness, and kurtosis as the pre-specified number of interested units in the sample increases.

Keywords: Bias; combined ratio estimator; mean square error; rare population; separate ratio estimator

ABSTRAK

Pensampelan rawak songsang berstrata (SIRS) merupakan teknik yang berkesan untuk mengkaji populasi yang jarang ditemui kerana ia menjamin bilangan unit yang berminat yang telah ditentukan terlebih dahulu dalam setiap stratum. Walaupun anggaran nisbah telah dikaji secara meluas di bawah pelbagai reka bentuk pensampelan, aplikasinya di bawah SIRS masih belum diterokai sepenuhnya. Kajian ini menangani jurang ini dengan membangunkan dua penganggar nisbah baharu untuk min populasi di bawah SIRS: penganggar nisbah berasingan dan penganggar nisbah gabungan. Sumbangan kerja ini adalah tiga kali ganda. Pertama, anggaran tertib pertama untuk bias dan ralat kuasa dua min (MSE) bagi setiap penganggar diterbitkan secara teliti menggunakan pengembangan siri Taylor di bawah andaian sampel besar dan penganggar berasaskan sampel yang tekal bagi MSE dibangunkan. Kedua, keadaan analitikal yang dinyatakan dalam bentuk parameter populasi diwujudkan dengan penganggar yang dicadangkan mengatasi penganggar tidak berat sebelah konvensional. Ketiga, kajian simulasi Monte Carlo yang meluas dengan 50,000 ulangan dijalankan untuk menilai prestasi sampel terhingga. Keputusan simulasi menunjukkan bahawa penganggar nisbah berasingan secara tekal mencapai MSE terkecil merentasi semua senario, mencapai peningkatan kecekapan sehingga 108.9% berbanding dengan penganggar tidak berat sebelah. Selain itu, semua penganggar yang dicadangkan menunjukkan penurunan dalam bias mutlak, MSE, kecondongan dan kurtosis apabila bilangan unit yang berminat yang telah ditentukan dalam sampel meningkat.

Kata kunci: Bias; penganggar nisbah berasingan; penganggar nisbah gabungan; populasi yang jarang berlaku; ralat min kuasa dua

INTRODUCTION

A primary concern in sample survey design is the difficulty of obtaining accurate parameter estimates for rare populations, particularly when the number of units of interest is limited. Under fixed-size sampling designs, low-prevalence phenomena - such as rare disease cases, endangered species sightings, or off-season crop failures - often result in samples that contain no units of interest, yielding parameter estimates of zero. Inverse sampling, which involves sequential sampling until a pre-specified number of interested units is observed, was introduced to overcome this limitation by guaranteeing the inclusion of a fixed number of units of interest in the sample.

Haldane (1945) was the first to derive an unbiased estimator of the population proportion under simple inverse random sampling with replacement. Finney (1949) subsequently proposed an unbiased estimator for the variance of Haldane's estimator. Salehi and Seber (2001) derived unbiased estimators for both the population mean and its variance under inverse sampling without replacement, a design referred to as inverse random sampling.

Christman and Lan (2001) integrated inverse sampling with adaptive cluster sampling by proposing three stopping rules based on the count of interested units, yielding an unbiased estimator of the population total; however, variance estimation was not addressed. Salehi and Seber (2001) extended inverse sampling to general adaptive cluster sampling, proposing unbiased estimators for the population total and its variance under sampling without replacement. A common feature of these methods is their reliance on equal-probability sampling. However, it is well established that estimator efficiency can be substantially enhanced through the incorporation of auxiliary information. Greco and Naddeo (2007) extended inverse sampling theory to designs with unequal selection probabilities and replacement, deriving unbiased estimators for the population total and its variance.

The precision of population parameter estimates can be substantially improved by incorporating an auxiliary variable that is strongly correlated with the study variable. In such cases, the ratio estimator is a widely used tool for improving efficiency relative to the standard unbiased estimator. Murthy (1967) demonstrated that under simple random sampling, the ratio estimator outperforms the unbiased estimator when a common condition is satisfied. Building on this principle, Sisodia and Dwivedi (1981) introduced a modified ratio estimator for settings where the coefficient of variation of the auxiliary variable is known. Subsequent developments in ratio estimation under simple random sampling include Kadilar and Cingi (2004), Subramani (2013), and Singh and Nigam (2020), each proposing modifications to improve estimator efficiency under varying conditions.

Within the inverse sampling framework, Sungsuwan and Suwattee (2014) introduced a model-assisted estimator for the population total under inverse random sampling, demonstrating substantial efficiency gains over the unbiased estimator when the auxiliary and study variables are strongly correlated. More recently, Sangngam and Laoarun (2025) proposed design-based ratio estimators for the population mean under inverse random sampling and found that these ratio estimators can be more efficient than the conventional unbiased estimator.

An alternative and complementary strategy for variance reduction is stratification. Stratified sampling is among the most effective design-based methods for variance reduction in surveys (Cochran 1977), achieved by partitioning the population into internally homogeneous strata. Sangngam and Suwattee (2012) combined stratification with inverse sampling to develop stratified inverse random sampling (SIRS), in which sampling proceeds independently within each stratum until a predetermined number of interested units is obtained. They derived an unbiased estimator for the population mean using Murthy's estimator, applicable to both equal-probability and probability-proportional-to-size (PPS) schemes. This design guarantees adequate representation of rare units and reduces between-stratum variance. However, a key limitation of the existing SIRS framework is its inability to incorporate continuous auxiliary information at the estimation stage - a gap that motivates the present work.

This paper addresses this methodological gap by proposing and evaluating two novel ratio estimators for the population mean under SIRS: A separate ratio estimator and a combined ratio estimator. We derive first-order approximations to the bias and mean squared error (MSE) of each estimator under the assumption that sample sizes within strata are sufficiently large. Consistent, sample-based MSE estimators are also constructed. Furthermore, we establish explicit analytical conditions under which each proposed estimator achieves a lower MSE than the conventional unbiased estimator, providing theoretically grounded guidance for practical application. Finally, we conduct a Monte Carlo simulation study with 50,000 replications to assess finite-sample performance, providing quantitative comparisons of biases and MSEs across various estimators and sample size configurations.

MATERIALS AND METHODS

ESTIMATORS UNDER INVERSE RANDOM SAMPLING

Consider a finite population of N distinct units, indexed by $i = 1, 2, ..., N$. For each unit i, let y_i denoted the value of study variable and x_i be the value of an auxiliary variable. The parameter of interest is the population mean, defined as:

$$\overline{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i \, .$$

The population is partitioned into two disjoint classes, C and C', based on whether the study variable, y_i, satisfies a given condition. A typical condition takes the form $y_i > c$, where c is a predetermined constant. The class C consists of all units for which the condition holds, while C' comprises the remaining units. The sizes of classes C and C' are denoted by M and $N - M$, respectively.

Under an inverse random sampling design, units are drawn sequentially with equal probability and without replacement until a predetermined number, m, of units from class C are included in the sample. The final sample size, denoted by n, is therefore a random variable. The resulting sample, s, consists of two disjoint subsets: S_C, the set of m

units from C, and $S_{C'}$, the set of $n - m$ units from C'. Salehi and Seber (2001) showed that an unbiased estimator for the population mean, $\bar{Y}$, under this design is given by

$$\bar{y}_U = \hat{P}\bar{y}_C + \left(1-\hat{P}\right)\bar{y}_{C'}, \quad (1)$$

where $\hat{P} = \frac{m-1}{n-1}$, $\bar{y}_C = \frac{1}{m}\sum_{i\in s_C} y_i$, and $\bar{y}_{C'} = \frac{1}{n-m}\sum_{i\in s_{C'}} y_i$.

The variance of $\bar{y}_U$ is given by

$$V(\bar{y}_U) = \left(\bar{Y}_C - \bar{Y}_{C'}\right)^2 V\left(\hat{P}\right) + \left(1-\frac{m}{M}\right)\frac{S^2_{YC}}{m}E\left(\hat{P}^2\right) + \frac{S^2_{YC'}}{m-1}E\left[\hat{P}\left(1-\hat{P}\right)\left(1-\frac{n-m}{N-M}\right)\right], \quad (2)$$

where $V\left(\hat{P}\right)$ is the variance of estimator $\hat{P}$ for estimating the population proportion P, $S^2_{YC} = \frac{1}{M-1}\sum_{i\in C} y_i$, $S^2_{YC'} = \frac{1}{N-M-1}\sum_{i\in C'} y_i$, $\bar{Y}_C = \frac{1}{M}\sum_{i\in C} y_i$ and $\bar{Y}_{C'} = \frac{1}{N-M}\sum_{i\in C'} y_i$.

An unbiased estimator of the variance (2) is given by

$$\hat{V}(\bar{y}_U) = \left(\bar{y}_C - \bar{y}_{C'}\right)^2 \hat{V}\left(\hat{P}\right) + \frac{\hat{S}^2_{YC}}{m}\left(\hat{P}^* - \frac{m}{N}\hat{P}\right) + \hat{S}^2_{YC'}\left[\frac{\hat{P}\left(1-\hat{P}\right)}{m-1} - \frac{\left(1-\hat{P}\right)}{N} - \frac{\hat{V}\left(\hat{P}\right)}{n-m}\right]$$

where $\hat{V}\left(\hat{P}\right) = \frac{\hat{P}\left(1-\hat{P}\right)}{n-2}$, $\hat{P}^* = \frac{(m-1)(m-2)}{(n-1)(n-2)}$, $\hat{S}^2_{YC} = \frac{1}{m-1}\sum_{i\in s_C}\left(y_i - \bar{y}_C\right)^2$, and $\hat{S}^2_{YC'} = \frac{1}{n-m-1}\sum_{i\in s_{C'}}\left(y_i - \bar{y}_{C'}\right)^2$.

By replacing the study variable y with the auxiliary variable x in the preceding formulas, analogous notations for auxiliary variable can be derived. The ratio estimator for the population mean, $\bar{Y}$, under the inverse random sampling design, as proposed by Sangngam and Laoarun (2025), is given by:

$$\bar{y}_R = \frac{\bar{y}_U}{\bar{x}_U}\bar{X}, \quad (3)$$

where $\bar{x}_U$ is the unbiased estimator of the population mean $\bar{X}$.

The bias and MSE approximations that follow are derived using a first-order Taylor series expansion. This approach is valid under the assumption that the sample sizes are sufficiently large so that higher-order terms are negligible. The bias of ratio estimator is approximated as follows:

$$B\left(\bar{y}_R\right) = \frac{1}{\bar{X}}\left[RV\left(\bar{x}_U\right) - Cov\left(\bar{y}_U, \bar{x}_U\right)\right]. \quad (4)$$

The approximated mean squared error of the estimator $\bar{y}_R$ is expresses as:

$$MSE\left(\bar{y}_R\right) = V\left(\bar{y}_U\right) + R^2 V\left(\bar{x}_U\right) - 2RCov\left(\bar{y}_U, \bar{x}_U\right), \quad (5)$$

where $R = \frac{\bar{Y}}{\bar{X}}$, $Cov\left(\bar{y}_U, \bar{x}_U\right) = \left(\bar{Y}_C - \bar{Y}_{C'}\right)\left(\bar{X}_C - \bar{X}_{C'}\right)V\left(\hat{P}\right) + \left(1-\frac{m}{M}\right)\frac{S_{XYC}}{m}E\left(\hat{P}^2\right)$ $+ \frac{S_{XYC'}}{m-1}E\left[\hat{P}\left(1-\hat{P}\right)\left(1-\frac{n-m}{N-M}\right)\right]$, $S_{XYC} = \frac{1}{M-1}\sum_{i\in C}\left(y_i - \bar{Y}_C\right)\left(x_i - \bar{X}_C\right)$ and $S_{XYC'} = \frac{1}{N-M-1}\sum_{i\in C'}\left(y_i - \bar{Y}_{C'}\right)\left(x_i - \bar{X}_{C'}\right)$.

A consistent estimator for the MSE in (5), denoted by $M\hat{S}E(\bar{y}_R)$, was introduced by Sangngam and Laoarun (2025) as follows:

$$M\hat{S}E\left(\bar{y}_R\right) = \hat{V}\left(\bar{y}_U\right) + \hat{R}^2\hat{V}\left(\bar{x}_U\right) - 2\hat{R}C\hat{o}v\left(\bar{y}_U, \bar{x}_U\right),$$

where $\hat{R} = \frac{\bar{y}_U}{\bar{x}_U}$, $C\hat{o}v\left(\bar{y}_U, \bar{x}_U\right) = \left(\bar{y}_C - \bar{y}_{C'}\right)\left(\bar{x}_C - \bar{x}_{C'}\right)\hat{V}\left(\hat{P}\right) + \frac{\hat{S}_{XYC}}{m}\left(\hat{P}^* - \frac{m}{N}\hat{P}\right)$ $+ \hat{S}_{XYC'}\left[\frac{\hat{P}\left(1-\hat{P}\right)}{m-1} - \frac{\left(1-\hat{P}\right)}{N} - \frac{\hat{V}\left(\hat{P}\right)}{n-m}\right]$, $\hat{S}_{XYC} = \frac{1}{m-1}\sum_{i\in s_C}\left(x_i - \bar{x}_C\right)\left(y_i - \bar{y}_C\right)$ and $\hat{S}_{XYC'} = \frac{1}{n-m-1}\sum_{i\in s_{C'}}\left(x_i - \bar{x}_{C'}\right)\left(y_i - \bar{y}_{C'}\right)$. The estimator $C\hat{o}v\left(\bar{y}_U, \bar{x}_U\right)$ is an unbiased estimator of $Cov\left(\bar{y}_U, \bar{x}_U\right)$.

AN UNBIASED ESTIMATOR UNDER SISR

Let the population be stratified into L non-overlapping strata. The *h-th* stratum, for $h = 1,2,..., L$ contains N_h units where $N = \sum_{h=1}^{L} N_h$. Let y_{hi} be the value of study variable for the *i-th* unit in the *h-th* stratum. The population mean of the study variable in the *h-th* stratum is represented by $\bar{Y}_h = \frac{1}{N_h}\sum_{j=1}^{N_h} y_{hj}$. Within each stratum, the units are portioned into two mutually exclusive subsets, C_h and C'_h, according to a criterion defined on the study variable y_{hi}. This criterion often takes the form of a threshold comparison, such as $y_{hi} > c$ for a predetermined value of c. Subsequently, C_h is composed of all units that meet the criterion, whereas C'_h contains all units that do not. The size of subsets C_h and C'_h are denoted by M_h and $N_h - M_h$, respectively. The population mean can be written as, $\bar{Y} = \sum_{i=1}^{N} W_h\bar{Y}_h$ where $W_h = \frac{N_h}{N}$.

The sampling design involves applying inverse random sampling independently within each stratum h. Specifically, units are drawn sequentially without replacement until exactly m_h units from class C_h are obtained, where m_h is a pre-specified number. This overall procedure is known as stratified inverse random sampling.

For each stratum, the resulting sample, denoted by $s_h = \left\{i_{h1}, i_{h2}, \ldots, i_{hn_h}\right\}$, has a random sample size n_h. This sample can be partitioned into two disjoint subsets: S_{Ch}, which consists of the m_h units drawn from class C_h, and $S_{C'h}$, which consists of $n_h - m_h$ units drawn from class C'_h. Consequently, the sample is the union of these two subsets, $s_h = s_{Ch} \cup s_{C'h}$ where $s_{Ch} \cap s_{C'h} = \varnothing$. The total sample size $n = \sum_{h=1}^{L} n_h$, is also a random variable.

Under stratified inverse random sampling, Sangngam and Suwatte (2012) showed that an unbiased estimator of $\bar{Y}$ is given by

$$\bar{y}_{st} = \sum_{h=1}^{L} W_h\left[\hat{P}_h\,\bar{y}_{Ch} + (1-\hat{P}_h)\,\bar{y}_{C'h}\right] = \sum_{h=1}^{L} W_h\bar{y}_{Uh}, \quad (6)$$

where $\bar{y}_{Uh} = \hat{P}_h\,\bar{y}_{Ch} + (1-\hat{P}_h)\,\bar{y}_{C'h}$, $\hat{P}_h = \frac{m_h - 1}{n_h - 1}$, $\bar{y}_{Ch} = \frac{1}{m_h}\sum_{i\in s_{Ch}} y_{hi}$ and $\bar{y}_{C'h} = \frac{1}{n_h - m_h}\sum_{i\in s_{C'h}} y_{hi}$. The variance of $\bar{y}_{st}$ is

$$V\left(\bar{y}_{st}\right)=\sum_{h=1}^{L}W_h^2\left[\left(\bar{Y}_{Ch}-\bar{Y}_{C'h}\right)^2V\left(\hat{P}_h\right)+\left(1-\frac{m_h}{M_h}\right)\frac{S_{Ch}^2}{m_h}E\left(\hat{P}_h^2\right)+\frac{S_{C'h}^2}{m_h-1}E\left[\hat{P}_h\left(1-\hat{P}_h\right)\left(1-\frac{n_h-m_h}{N_h-M_h}\right)\right]\right],$$

where $\bar{Y}_{Ch}=\frac{1}{M_h}\sum_{i\in Ch}y_{hi}$, $\bar{Y}_{C'h}=\frac{1}{N_h-M_h}\sum_{i\in\bar{C}h}y_{hi}$, $S_{Ch}^2=\frac{1}{M_h-1}\sum_{i\in Ch}\left(y_{hi}-\bar{Y}_{Ch}\right)^2$ and $S_{C'h}^2=\frac{1}{N_h-M_h-1}\sum_{i\in\bar{C}h}\left(y_{hi}-\bar{Y}_{C'h}\right)^2$. An unbiased estimator of the variance is

$$\hat{V}\left(\bar{y}_{st}\right)=\sum_{h=1}^{L}W_h^2\left[\left(\bar{y}_{Ch}-\bar{y}_{C'h}\right)^2\hat{V}\left(\hat{P}_h\right)+\frac{\hat{S}_{Ch}^2}{m_h}\left(\hat{P}_h^*-\frac{m_h}{N_h}\hat{P}_h\right)+\hat{S}_{C'h}^2\left(\frac{\hat{P}_h\left(1-\hat{P}_h\right)}{m_h-1}-\frac{1-\hat{P}_h}{N_h}-\frac{\hat{V}\left(\hat{P}_h\right)}{n_h-m_h}\right)\right].$$

where $\hat{V}\left(\hat{P}_h\right)=\frac{\hat{P}_h\left(1-\hat{P}_h\right)}{n_h-2}$, $\hat{P}_h^*=\frac{\left(m_h-1\right)\left(m_h-2\right)}{\left(n_h-1\right)\left(n_h-2\right)}$, $\hat{S}_{Ch}^2=\frac{1}{m_h-1}\sum_{i\in s_{Ch}}\left(y_{hi}-\bar{y}_{Ch}\right)^2$ and $\hat{S}_{C'h}^2=\frac{1}{n_h-m_h-1}\sum_{i\in s_{\bar{C}h}}\left(y_{hi}-\bar{y}_{C'h}\right)^2$.

This paper employs the SIRS framework to construct two proposed ratio estimators and to derive their biases and MSEs.

RESULTS

PROPOSED ESTIMATORS

Under SIRS, two ratio estimators for the population mean, $\bar{Y}$ are proposed. Both estimators leverage the strong positive correlation between the study variable and the auxiliary variable. However, they require different levels of auxiliary information: the first estimator assumes that the stratum-specific population means $\bar{X}_h$ to be known for all strata, whereas the second estimator requires only the overall population mean $\bar{X}$ is known.

The first estimator, known as the separate ratio estimator, involves calculating a distinct ratio estimate for the population mean within each stratum. Subsequently, an overall estimate for the population mean is obtained by taking a weighted sum of these individual stratum estimates. A stratum h contains N_h units with observations (x_{hi}, y_{hi}) where h = 1,2,..., L and i = 1,2,..., N_h. By replacing the study value y_{hi} with the auxiliary value x_{hi}, we can derive the analogous formula of auxiliary variable required for the proposed ratio estimators. If $\bar{y}_{Uh}$, $\bar{x}_{Uh}$, are the unbiased estimates of the stratum means in the h-th stratum, the $\bar{X}_h=\frac{1}{N_h}\sum_{i=1}^{N_h}x_{hi}$ is the stratum mean of the x_{hi}, the proposed separate ratio estimate $\bar{y}_{RS}$ is

$$\bar{y}_{RS}=\sum_{h=1}^{L}W_h\frac{\bar{y}_{Uh}}{\bar{x}_{Uh}}\bar{X}_h, \tag{7}$$

where $W_h=\frac{N_h}{N}$, $\bar{y}_{Uh}=\hat{P}_h\bar{y}_{Ch}+\left(1-\hat{P}_h\right)\bar{y}_{C'h}$, $\hat{P}_h=\frac{m_h-1}{n_h-1}$, $\bar{y}_{Ch}=\frac{1}{m_h}\sum_{i\in s_{Ch}}y_{hi}$ and $\bar{y}_{C'h}=\frac{1}{n_h-m_h}\sum_{i\in s_{C'}}y_{hi}$. The symbols associated with the auxiliary variable x are defined in the same manner as those corresponding to the study variable y.

This estimator does not presuppose the homogeneity of true ratios across strata. Nevertheless, the computation of the estimate depends on the availability of the stratum means for the auxiliary variable.

To derive the MSE and bias of the proposed separated ratio estimator, we apply the expressions for the MSE and bias of the estimator $\bar{y}_{Rh}=\frac{\bar{y}_{Uh}}{\bar{x}_{Uh}}\bar{X}_h$ under inverse random sampling, where units are drawn independently within each stratum. The resulting expressions are given as follows:

$$\begin{aligned}B\left(\bar{y}_{RS}\right)&=E\left(\bar{y}_{RS}\right)-\bar{Y}\\&=E\left(\sum_{h=1}^{L}W_h\frac{\bar{y}_{Uh}}{\bar{x}_{Uh}}\bar{X}_h\right)-\sum_{h=1}^{L}W_h\bar{Y}_h\\&=\sum_{h=1}^{L}W_hE\left(\frac{\bar{y}_{Uh}}{\bar{x}_{Uh}}\bar{X}_h-\bar{Y}_h\right)\\&=\sum_{h=1}^{L}W_hB\left(\bar{y}_{Rh}\right)\end{aligned}$$

$$=\sum_{h=1}^{L}W_h\frac{1}{\bar{X}_h}\left[R_hV\left(\bar{x}_{Uh}\right)-Cov\left(\bar{y}_{Uh},\bar{x}_{Uh}\right)\right], \tag{8}$$

$$MSE\left(\bar{y}_{RS}\right)=\sum_{h=1}^{L}W_h^2MSE\left(\bar{y}_{Rh}\right)$$

$$=\sum_{h=1}^{L}W_h^2\left[V\left(\bar{y}_{Uh}\right)+R_h^2V\left(\bar{x}_{Uh}\right)-2R_hCov\left(\bar{y}_{Uh},\bar{x}_{Uh}\right)\right], \tag{9}$$

where $Cov\left(\bar{y}_{Uh},\bar{x}_{Uh}\right)=\left(\bar{Y}_{Ch}-\bar{Y}_{C'h}\right)\left(\bar{X}_{Ch}-\bar{X}_{C'h}\right)V\left(\hat{P}_h\right)+\left(1-\frac{m_h}{M_h}\right)\frac{S_{XYCh}}{m_h}E\left(\hat{P}_h^2\right)+\frac{S_{XYC'h}}{m_h-1}E\left[\hat{P}_h\left(1-\hat{P}_h\right)\left(1-\frac{n_h-m_h}{N_h-M_h}\right)\right]$, $S_{XYCh}=\frac{1}{M_h-1}\sum_{i\in Ch}\left(y_{hi}-\bar{Y}_{Ch}\right)\left(x_{hi}-\bar{X}_{Ch}\right)$, $S_{XYC'h}=\frac{1}{N_h-M_h-1}\sum_{i\in C'h}\left(y_{hi}-\bar{Y}_{C'h}\right)\left(x_{hi}-\bar{X}_{C'h}\right)$, $R_h=\frac{\bar{Y}_h}{\bar{X}_h}$, $V\left(\bar{y}_{Uh}\right)=\left(\bar{Y}_{Ch}-\bar{Y}_{C'h}\right)^2V\left(\hat{P}_h\right)+\left(1-\frac{m_h}{M_h}\right)\frac{S_{YCh}^2}{m_h}E\left(\hat{P}_h^2\right)+\frac{S_{YC'h}^2}{m_h-1}E\left[\hat{P}_h\left(1-\hat{P}_h\right)\left(1-\frac{n_h-m_h}{N_h-M_h}\right)\right]$, $S_{YCh}^2=\frac{1}{M_h-1}\sum_{i\in Ch}y_{hi}$, $S_{YC'h}^2=\frac{1}{N_h-M_h-1}\sum_{i\in C'h}y_{hi}$, $\bar{Y}_{Ch}=\frac{1}{M_h}\sum_{i\in Ch}y_{hi}$, $\bar{Y}_{C'h}=\frac{1}{N_h-M_h}\sum_{i\in C'h}y_{hi}$

and $V\left(\hat{P}_h\right)$ is the variance of estimator $\hat{P}_h$ for estimating the parameter $P_h=\frac{M_h}{N_h}$.

The validity of the formula (8) and (9) are conditional upon a sufficiently large sample size within each stratum, which ensures the reliability of the approximate bias and MSE. This limitation is a critical consideration for practical applications.

For estimating $MSE(\bar{y}_{RS})$, we substitute the sample-based estimates in to (9) as

$$M\hat{S}E\left(\bar{y}_{RS}\right)=\sum_{h=1}^{L}W_h^2\left[\hat{V}\left(\bar{y}_{Uh}\right)+\hat{R}_h^2\hat{V}\left(\bar{x}_{Uh}\right)-2\hat{R}_hC\hat{o}v\left(\bar{y}_{Uh},\bar{x}_{Uh}\right)\right], \tag{10}$$

where $\hat{R}_h=\frac{\bar{y}_{Uh}}{\bar{x}_{Uh}}$, $\hat{V}\left(\hat{P}_h\right)=\frac{\hat{P}_h\left(1-\hat{P}_h\right)}{n_h-2}$, $\hat{P}_h^*=\frac{\left(m_h-1\right)\left(m_h-2\right)}{\left(n_h-1\right)\left(n_h-2\right)}$, $\hat{S}_{YCh}^2=\frac{1}{m_h-1}\sum_{i\in s_{Ch}}\left(y_{hi}-\bar{y}_{Ch}\right)^2$, $\hat{S}_{YC'h}^2=\frac{1}{n_h-m_h-1}\sum_{i\in s_{C'h}}\left(y_{hi}-\bar{y}_{C'h}\right)^2$, $C\hat{o}v\left(\bar{y}_{Uh},\bar{x}_{Uh}\right)=\left(\bar{y}_{Ch}-\bar{y}_{C'h}\right)\left(\bar{x}_{Ch}-\bar{x}_{C'h}\right)\hat{V}\left(\hat{P}_h\right)+\frac{\hat{S}_{XYCh}}{m_h}\left(\hat{P}_h^*-\frac{m_h}{N_h}\hat{P}_h\right)+\hat{S}_{XYC'h}\left[\frac{\hat{P}_h\left(1-\hat{P}_h\right)}{m_h-1}-\frac{\left(1-\hat{P}_h\right)}{N_h}-\frac{\hat{V}\left(\hat{P}_h\right)}{n_h-m_h}\right]$, $\hat{S}_{XYCh}=\frac{1}{m_h-1}\sum_{i\in s_{Ch}}\left(x_{hi}-\bar{x}_{Ch}\right)\left(y_{hi}-\bar{y}_{Ch}\right)$ and $\hat{S}_{XYC'h}=\frac{1}{n_h-m_h-1}\sum_{i\in s_{C'h}}\left(x_{hi}-\bar{x}_{C'h}\right)\left(y_{hi}-\bar{y}_{C'h}\right)$.

An alternative estimator is derived using a single combined ratio. Based on the sample data, the proposed combined ratio estimator is given by

$$\bar{y}_{RC}=\frac{\bar{y}_{st}}{\bar{x}_{st}}\bar{X}. \tag{11}$$

The computation of this estimate requires knowledge of the overall population mean of auxiliary variable ($\bar{X}$), rather than the individual stratum means ($\bar{X}_h$).

To obtain the MSE and bias of proposed combined ratio estimator, let $e_1 = \frac{\bar{y}_{st} - \bar{Y}}{\bar{Y}}$ and $e_2 = \frac{\bar{x}_{st} - \bar{X}}{\bar{X}}$. It may be noted that $E(e_1) = 0$, $E(e_2) = 0$, $E(e_1^2) = \frac{V(\bar{y}_{st})}{\bar{Y}^2}$, $E(e_2^2) = \frac{V(\bar{x}_{st})}{\bar{X}^2}$ and $E(e_1 e_2) = \frac{Cov(\bar{y}_{st}, \bar{x}_{st})}{\bar{Y}\,\bar{X}}$. The estimator $\bar{y}_{RC}$ can be written as $\bar{y}_{RC} = \bar{Y}(1+e_1)(1+e_2)^{-1}$. Using Taylor series expansion,

$$(1+e_2)^{-1} = (1 - e_2 + e_2^2 - ....)$$

We assume that the sample size is large enough to ensure that the relative errors e_1 and e_2 sufficiently small. The first-order Taylor series expansion is $(1+e_2)^{-1} \approx (1 - e_2 + e_2^2)$. We obtain that

$$\begin{aligned} \bar{y}_{RC} &\approx \bar{Y}\left(1+e_1\right)\left(1-e_2+e_2^2\right) \\ &\approx \bar{Y}\left(1+e_1-e_2+e_2^2-e_1e_2+e_1e_2^2\right). \end{aligned}$$

When the terms with degree greater than two are ignored, we get

$$\begin{aligned} B\left(\bar{y}_{RC}\right) = E\left(\bar{y}_{RC}-\bar{Y}\right) &\approx \bar{Y}E\left(e_2^2-e_1e_2\right) \\ &= \bar{Y}\left[\frac{V\left(\bar{x}_{st}\right)}{\bar{X}^2} - \frac{Cov\left(\bar{x}_{st},\bar{y}_{st}\right)}{\bar{X}\,\bar{Y}}\right] \\ &= \frac{1}{\bar{X}}\left[RV\left(\bar{x}_{st}\right) - Cov\left(\bar{y}_{st},\bar{x}_{st}\right)\right], \end{aligned} \qquad (12)$$

where $R = \frac{\bar{Y}}{\bar{X}}$. Since the units are drawn independently within each stratum, the covariance of $\bar{x}_{st}$ and $\bar{y}_{st}$ is given by

$$\begin{aligned} Cov\left(\bar{y}_{st},\bar{x}_{st}\right) &= E\left(\bar{y}_{st}\bar{x}_{st}\right) - E\left(\bar{y}_{st}\right)E\left(\bar{x}_{st}\right) \\ &= E\left(\bar{y}_{st}\bar{x}_{st}\right) - \bar{Y}\,\bar{X} \\ &= E\left[\left(\sum_{h=1}^{L} W_h \bar{y}_{Uh}\right)\left(\sum_{h=1}^{L} W_h \bar{x}_{Uh}\right)\right] - \left(\sum_{h=1}^{L} W_h \bar{Y}_h\right)\left(\sum_{h=1}^{L} W_h \bar{X}_h\right) \\ &= \sum_{h=1}^{L} W_h^2 Cov\left(\bar{y}_{Uh},\bar{x}_{Uh}\right) + \sum_{h=1}^{L}\sum_{k\neq h}^{L} W_h W_k Cov\left(\bar{y}_{Uh},\bar{x}_{Uk}\right) \\ &= \sum_{h=1}^{L} W_h^2 Cov\left(\bar{y}_{Uh},\bar{x}_{Uh}\right) \\ &= \sum_{h=1}^{L} W_h^2 \left[\left(\bar{Y}_{Ch}-\bar{Y}_{C'h}\right)\left(\bar{X}_{Ch}-\bar{X}_{C'h}\right)V\left(\hat{P}_h\right) + \left(1-\frac{m_h}{M_h}\right)\frac{S_{Ch}^2(x,y)}{m_h}E\left(\hat{P}_h^2\right)\right. \\ &\qquad \left. + \frac{S_{C'h}^2(x,y)}{m_h-1}E\left[\hat{P}_h\left(1-\hat{P}_h\right)\left(1-\frac{n_h-m_h}{N_h-M_h}\right)\right]\right]. \end{aligned}$$

By ignoring terms of degree greater than two, the approximate mean squared error (MSE) of the proposed combined ratio estimator is obtained as

$$\begin{aligned} MSE\left(\bar{y}_{RC}\right) &= E\left(\bar{y}_{RC} - \bar{Y}\right)^2 \\ &\approx \bar{Y}^2 E\left(e_1^2 + e_2^2 - e_1e_2\right) \\ &= \bar{Y}^2\left[\frac{V\left(\bar{y}_{st}\right)}{\bar{Y}^2} + \frac{V\left(\bar{x}_{st}\right)}{\bar{X}^2} - 2\frac{Cov\left(\bar{y}_{st},\bar{x}_{st}\right)}{\bar{Y}\,\bar{X}}\right]. \\ &= V\left(\bar{y}_{st}\right) + R^2V\left(\bar{x}_{st}\right) - 2RCov\left(\bar{y}_{st},\bar{x}_{st}\right) \end{aligned} \qquad (13)$$

For sample estimate of the $MSE(\bar{y}_{RC})$, we substitute the sample-based estimates of R, $V(\bar{y}_{st})$, $V(\bar{x}_{st})$ and $Cov(\bar{y}_{st}, \bar{x}_{st})$ into (13). The estimate of the $MSE(\bar{y}_{RC})$ is given by

$$M\hat{S}E\left(\bar{y}_{RC}\right) = \hat{V}\left(\bar{y}_{st}\right) + \hat{R}^2\hat{V}\left(\bar{x}_{st}\right) - 2\hat{R}C\hat{o}v\left(\bar{y}_{st},\bar{x}_{st}\right) \qquad (14)$$

where $\hat{R} = \frac{\bar{y}_{st}}{\bar{x}_{st}}$ and $C\hat{o}v\left(\bar{y}_{st},\bar{x}_{st}\right) = \sum_{h=1}^{L} W_h^2 \left\{\left(\bar{y}_{Ch}-\bar{y}_{C'h}\right)\left(\bar{x}_{Ch}-\bar{x}_{C'h}\right)\hat{V}\left(\hat{P}_h\right) + \frac{\hat{S}_{XYCh}}{m_h}\left(\hat{P}_h^* - \frac{m_h}{N_h}\hat{P}_h\right)\right.$ $\left. + \hat{S}_{XYC'h}\left[\frac{\hat{P}_h\left(1-\hat{P}_h\right)}{m_h-1} - \frac{\left(1-\hat{P}_h\right)}{N_h} - \frac{\hat{V}\left(\hat{P}_h\right)}{n_h-m_h}\right]\right\}$.

COMPARISON OF THE EFFICIENCY

The performance of the proposed ratio estimators is investigated in this section, benchmarked against established unbiased estimators in the context of stratified inverse random sampling. Initially, the proposed separate ratio estimator is compared with the conventional unbiased estimator. The condition under which the proposed estimator is more efficient is given as follows:

$$\begin{aligned} & MSE\ (\bar{y}_{RS}) \leq V\ (\bar{y}_{st}), \\ & \sum_{h=1}^{L} W_h^2\left[V\left(\bar{y}_{Uh}\right) + R_h^2V\left(\bar{x}_{Uh}\right) - 2R_hCov\left(\bar{y}_{Uh},\bar{x}_{Uh}\right)\right] \leq \sum_{h=1}^{L} W_h^2 V\left(\bar{y}_{Uh}\right), \\ & \sum_{h=1}^{L}\frac{1}{2}W_h^2R_hV\left(\bar{x}_{Uh}\right) \leq \sum_{h=1}^{L} W_h^2Cov\left(\bar{y}_{Uh},\bar{x}_{Uh}\right). \end{aligned}$$

The proposed separate ratio estimator achieves higher efficiency than the unbiased estimator provided that all three conditions listed below are met:

$$\frac{1}{2}R_h\left(\bar{X}_{Ch}-\bar{X}_{C'h}\right)^2 \leq \left(\bar{Y}_{Ch}-\bar{Y}_{C'h}\right)\left(\bar{X}_{Ch}-\bar{X}_{C'h}\right), \qquad (14)$$

$$\frac{1}{2}R_hS_{XCh}^2 \leq S_{XYCh}, \qquad (15)$$

and $\frac{1}{2}R_hS_{XC'h}^2 \leq S_{XYC'h}$ for all stratum h. (16)

Next, the proposed combined ratio estimator is evaluated against the conventional unbiased estimator. The proposed estimator is more efficient under the following condition:

$$\begin{aligned} & MSE\ (\bar{y}_{RC}) \leq V\ (\bar{y}_{st}), \\ & \sum_{h=1}^{L} W_h^2\left[V\left(\bar{y}_{Uh}\right) + R^2V\left(\bar{x}_{Uh}\right) - 2RCov\left(\bar{y}_{Uh},\bar{x}_{Uh}\right)\right] \leq \sum_{h=1}^{L} W_h^2V\left(\bar{y}_{Uh}\right), \end{aligned}$$

and $\sum_{h=1}^{L}\frac{1}{2}W_h^2RV\left(\bar{x}_{Uh}\right) \leq \sum_{h=1}^{L}W_h^2Cov\left(\bar{y}_{Uh},\bar{x}_{Uh}\right)$ for all stratum h.

The proposed combined ratio estimator outperforms the unbiased estimator in terms of efficiency, provided that the following three conditions are satisfied:

$$\frac{1}{2}R\left(\bar{X}_{Ch}-\bar{X}_{C'h}\right)^2 \leq \left(\bar{Y}_{Ch}-\bar{Y}_{C'h}\right)\left(\bar{X}_{Ch}-\bar{X}_{C'h}\right), \qquad (17)$$

$$\frac{1}{2}RS_{XCh}^2 \leq S_{XYCh}, \qquad (18)$$

and $\frac{1}{2}RS_{XC'h}^2 \leq S_{XYC'h}$, for all stratum h. (19)

Finally, the proposed separate ratio estimator is compared with the combined ratio estimator, proving to be more efficient when the following condition is satisfied:

$$MSE(\bar{y}_{RC}) \leq MSE(\bar{y}_{RC}),$$
$$\sum_{h=1}^{L} W_h^2 \left[R_h^2 V(\bar{x}_{Uh}) - 2R_h Cov(\bar{y}_{Uh}, \bar{x}_{Uh}) \right] \leq \sum_{h=1}^{L} W_h^2 \left[R^2 V(\bar{x}_{Uh}) - 2R Cov(\bar{y}_{Uh}, \bar{x}_{Uh}) \right],$$
$$\sum_{h=1}^{L} W_h^2 \left[(R_h^2 - R^2) V(\bar{x}_{Uh}) - 2(R_h - R) Cov(\bar{y}_{Uh}, \bar{x}_{Uh}) \right] \leq 0,$$
$$\sum_{h=1}^{L} W_h^2 \left\{ (R_h - R)^2 V(\bar{x}_{Uh}) + 2(R - R_h) \left[Cov(\bar{y}_{Uh}, \bar{x}_{Uh}) - R_h V(\bar{x}_{Uh}) \right] \right\} \leq 0. \quad (20)$$

Unless the ratio R_h remains constant across strata, the use of a proposed separate ratio estimator in each stratum is likely to be more precise, provided the sample size per stratum is sufficiently large such that the $MSE(\bar{y}_{RS})$ approximation is valid and the cumulative bias $B(\bar{y}_{RS})$ is negligible.

SIMULATION STUDY

In this section, the simulated x-values and y-values from Chao (2004) were examined. To avoid a zero sample mean for the auxiliary variable in class C', each auxiliary value was increased by one. The simulation dataset, consisting of 400 units (20x20), was divided into four strata of equal size. Each stratum comprised 100 units (20×5), formed from 20 rows and 5 columns. The population mean of the y-values was $\bar{Y}$ = 0.635, the population mean of the x-values was $\bar{X}$ = 0.490, and the population ratio was R = 0.412. The population is visually represented in Figures 1 and 2, respectively.

A simulation study was conducted to compare the properties of the proposed ratio estimators against the conventional unbiased estimator. The condition $y_{hi} > 0$was used to classify units into class C_h or C'_{hx} where $h = 1,2,3,4$ and $i = 1,2,...,100$. A stratified inverse random sampling scheme was employed for sample selection, where the target number of units of interest in stratum h was set to 3, 4, 5, 6, 7 and 8. The simulation study was conducted using 50,000 replications. In each replication, the population mean ($\bar{Y}$) was estimated. The resulting estimates $\bar{y}$'s and the final sample sizes n's was then averaged across all replications. These averages served as approximations for their respective expected values, namely $\tilde{E}(\bar{y}) = \frac{1}{50,000}\sum_{i=1}^{50,000}\bar{y}_i$ and $\tilde{E}(n_h) = \frac{1}{50,000}\sum_{i=1}^{50,000} n_{hi}$. The performance of the estimators was then evaluated by comparing their estimated Bias and MSE.

The estimated Bias is given by: $\tilde{B}(\bar{y}) = \frac{1}{50,000}\sum_{i=1}^{50,000}\bar{y}_i - \bar{Y}$.
The estimated MSE is given by: $M\tilde{S}E(\bar{y}) = \frac{1}{50,000}\sum_{i=1}^{50,000}(\bar{y}_i - \bar{Y})^2$.

The relative efficiency (RE) of the proposed estimators is evaluated relative to the conventional unbiased estimator, computed as the ratio of their mean squared errors (MSEs) and defined as $\tilde{R}E(\bar{y}) = \frac{M\tilde{S}E(\bar{y}_{st})}{M\tilde{S}E(\bar{y})}$.

A relative efficiency greater than one, $\frac{M\tilde{S}E(\bar{y}_{st})}{M\tilde{S}E(\bar{y})} > 1$, indicates that $\bar{y}$ is more efficient than $\bar{y}_{st}$. To provide a visual representation of the sampling distributions, the case where $m_h = 4$ and $m_h = 6$ were selected for detailed analysis. Figures 3 to 8 illustrate the histograms generated from the estimates of the three respective estimators. Furthermore, the skewness and kurtosis coefficients for each distribution are reported accompanied by its histogram. These summary measures provide quantitative evidence on the shape characteristics of the distributions, supporting the assessment of the estimators' behavior under repeated sampling.

Table 1 presents the average sample sizes per stratum across all configurations. Stratum 3 consistently yields the largest average sample size, followed by strata 1, 4, and 2, respectively, while stratum 2 exhibits the smallest average sample size. Given equal stratum sizes N_h, this pattern reflects the inverse relationship between the average sample size and the number of interested units M_h in each stratum. The strata with a lower prevalence of units of interest require a larger expected number of draws to achieve the target number of interested units m_h. As expected, the average sample size increases monotonically with m_h for all strata, confirming the theoretical property of the SIRS design.

Table 2 summarizes the average estimates and approximate bias for all three estimators across all configurations. The average estimates from the unbiased estimator $\bar{y}_{st}$ are consistently closest to the true population mea ($\bar{Y}$ = 0.635), thereby verifying its property of unbiasedness. In each situation, the bias size of the proposed separate ratio estimator $\bar{y}_{RS}$ is higher than that of the proposed combined ratio estimator. When the number of interested units in sample m_h increases, the bias sizes of the two ratio estimators decrease, confirming that bias is controlled by increasing the number of interested units in the sample.

Table 3 presents the MSE and relative efficiency of the three estimators. For all three estimators, the MSE values decrease monotonically as the pre-specified number of interested units m_h increases, demonstrating that estimation precision improves with a larger sample size - a result consistent with the theoretical MSE expressions derived in equations (9) and (13). Across all configurations, a consistent hierarchy is observed: $\bar{y}_{RS}$ achieves the lowest MSE, followed by $\bar{y}_{RC}$, with $\bar{y}_{st}$ exhibiting the highest MSE. In terms of relative efficiency, the results show that $\tilde{R}E(\bar{y}_{RC})$ ranges from 1.353 to 1.563, representing a 35.3% to 56.3% gain in statistical efficiency. More substantially, $\tilde{R}E(\bar{y}_{RS})$ ranges from 1.586 to 2.089, representing a 58.6% to 108.9% gain in statistical efficiency. These quantitative comparisons demonstrate the considerable practical advantage of the proposed separate ratio estimator. Furthermore, the relative efficiency of both proposed estimators decreases as m_h increases. This pattern indicates that while larger sample sizes improve the absolute precision of all estimators, the relative advantage of the ratio estimators over the unbiased estimator diminishes with increasing m_h.

Figures 3–8 display the histograms of the three estimators for $m_h = 3$ and $m_h = 8$. The skewness and kurtosis coefficients accompanying each histogram provide quantitative evidence on the distributional shape of the estimators under repeated sampling. The distributions of

1	1	1	1	1
1	1	1	1	1
2	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	2	1
1	1	1	1	1
2	1	1	1	1
1	1	2	1	2
1	1	1	2	1
2	1	2	1	1
1	1	1	1	1
1	1	1	2	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	2	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1

1	1	1	1	2
1	1	1	1	1
2	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	2
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	2	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	2	3	1
1	1	1	5	7
1	1	5	7	12
1	2	5	5	10
1	2	3	8	5
1	1	1	1	1

1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
3	1	1	1	1
1	2	1	1	1
2	1	1	1	1
1	1	1	1	1
2	1	1	1	1
2	1	1	1	1
1	2	1	1	1
5	4	2	1	1
12	5	1	1	3
4	2	2	1	1
1	1	1	1	1

1	1	2	2	1
1	1	1	1	1
1	1	1	1	1
1	1	1	2	1
1	1	1	1	2
1	1	1	1	1
1	2	1	1	3
1	1	1	1	1
1	1	2	2	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	2	1	1	1
1	1	1	1	2
1	1	1	1	2
5	4	5	3	4
3	7	9	4	8
2	9	11	1	3
1	1	3	3	1
1	1	1	1	1

FIGURE 1. The values of auxiliary variable *x*

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	0
2	0	1	0	0
0	0	0	0	0
1	0	0	0	1
0	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	1	0	0
1	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
1	0	0	0	0
0	0	0	0	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	0
0	0	0	1	0
0	0	0	0	0
1	0	0	0	0
0	0	0	0	0
0	0	1	1	0
0	0	0	0	1
0	0	0	6	18
0	0	1	14	46
0	0	0	4	7

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
1	0	0	0	0
1	1	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
3	0	1	0	0
11	0	0	0	0
2	0	0	0	0

0	0	1	0	1
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	1	0	1	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	1	0	0	0
0	0	0	0	0
0	0	0	0	1
0	0	0	0	0
0	1	6	0	1
0	9	27	9	0
1	8	25	14	1
0	1	2	1	0

FIGURE 2. The values of study variable *y*

TABLE 1. Average sample sizes

m_1, m_2, m_3, m_4	$\tilde{E}(n_1)$	$\tilde{E}(n_2)$	$\tilde{E}(n_3)$	$\tilde{E}(n_4)$
3,3,3,3	27.520	17.843	33.609	11.648
4,4,4,4	36.748	23.725	44.959	15.527
5,5,5,5	45.878	29.644	56.127	19.372
6,6,6,6	55.055	35.607	67.321	23.300
7,7,7,7	64.296	41.562	78.488	27.220
8,8,8,8	73.419	47.529	89.800	31.075

TABLE 2. Average estimates, and approximate bias of proposed ratio estimators

m_1, m_2, m_3, m_4	$\tilde{E}(\bar{y}_{st})$	$\tilde{E}(\bar{y}_{RC})$	$\tilde{E}(\bar{y}_{RS})$	$\tilde{B}(\bar{y}_{RC})$	$\tilde{B}(\bar{y}_{RS})$
3,3,3,3	0.620	0.594	0.554	-0.041	-0.081
4,4,4,4	0.632	0.612	0.581	-0.023	-0.054
5,5,5,5	0.634	0.619	0.595	-0.016	-0.040
6,6,6,6	0.635	0.623	0.604	-0.012	-0.031
7,7,7,7	0.636	0.627	0.612	-0.008	-0.023
8,8,8,8	0.635	0.628	0.617	-0.007	-0.018

TABLE 3. Comparison of approximate mean squared errors (MSEs) for the estimators

m_1, m_2, m_3, m_4	$\tilde{MSE}(\bar{y}_{st})$	$\tilde{MSE}(\bar{y}_{RC})$	$\tilde{MSE}(\bar{y}_{RS})$	$\tilde{RE}(\bar{y}_{st})$	$\tilde{RE}(\bar{y}_{RC})$	$\tilde{RE}(\bar{y}_{RS})$
3,3,3,3	0.211	0.135	0.101	1.000	1.563	2.089
4,4,4,4	0.145	0.096	0.075	1.000	1.510	1.933
5,5,5,5	0.104	0.071	0.057	1.000	1.465	1.825
6,6,6,6	0.078	0.055	0.045	1.000	1.418	1.733
7,7,7,7	0.060	0.043	0.036	1.000	1.395	1.667
8,8,8,8	0.046	0.034	0.029	1.000	1.353	1.586

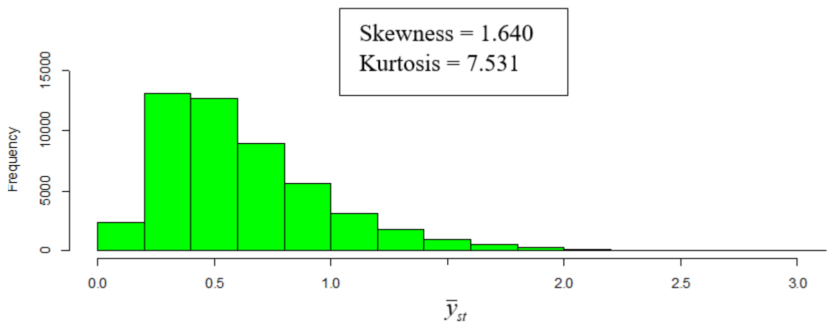

FIGURE 3. The histogram of $\bar{y}_{st}$ with $m_1 = m_2 = m_3 = m_4 = 4$

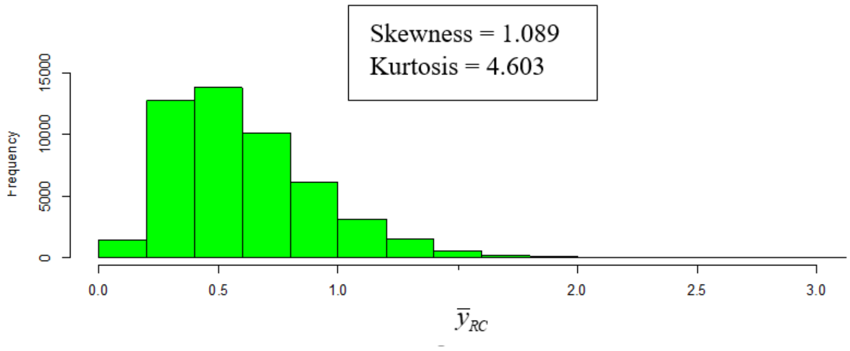

FIGURE 4. The histogram of $\bar{y}_{RC}$ with $m_1 = m_2 = m_3 = m_4 = 4$

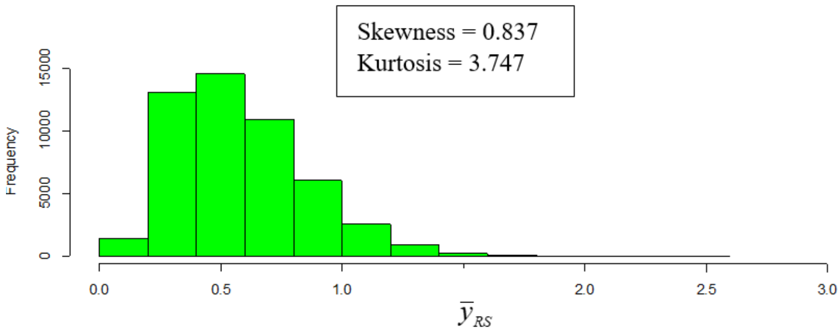

FIGURE 5. The histogram of $\bar{y}_{RS}$ with $m_1 = m_2 = m_3 = m_4 = 4$

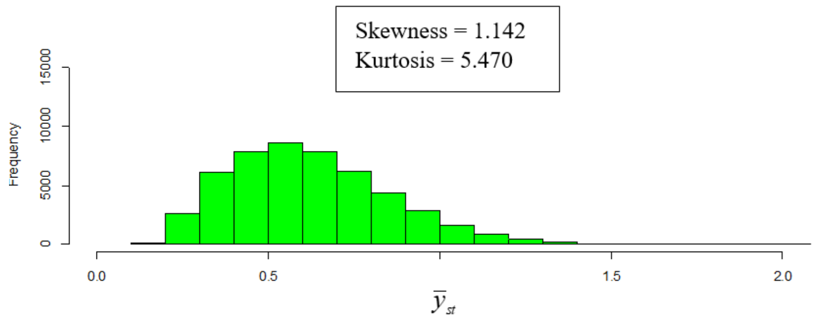

FIGURE 6. The histrogram of $\bar{y}_{st}$ with $m_1 = m_2 = m_3 = m_4 = 6$

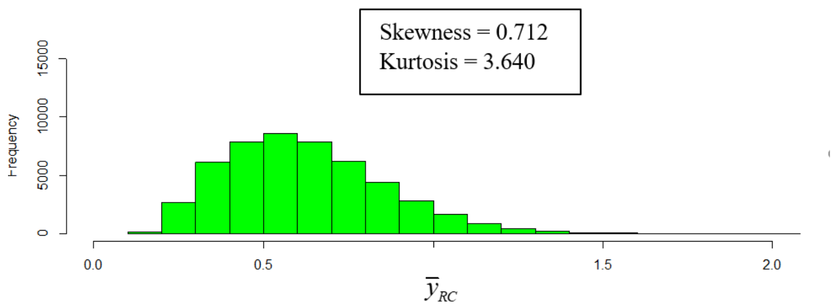

FIGURE 7. The histrogram of $\bar{y}_{RC}$ with $m_1 = m_2 = m_3 = m_4 = 6$

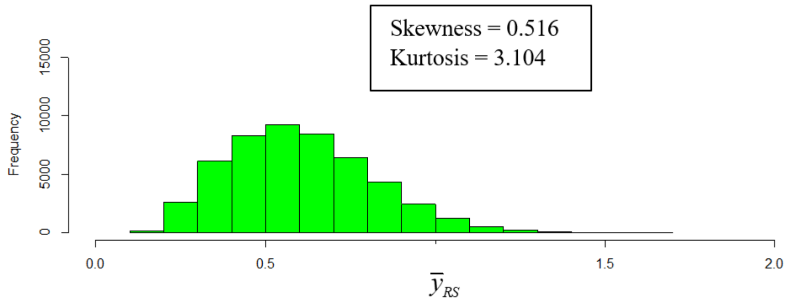

FIGURE 8. The histrogram of $\bar{y}_{RS}$ with $m_1 = m_2 = m_3 = m_4 = 6$

all three estimators are positively skewed. For a fixed m_h, the unbiased estimator $\bar{y}_{st}$ exhibits the highest skewness and kurtosis, indicating the heaviest right tail. The combined ratio estimator $\bar{y}_{RC}$ has intermediate skewness and kurtosis, while the separate ratio estimator $\bar{y}_{RSV}$ consistently shows the lowest skewness and kurtosis among the three, reflecting the most symmetric and light-tailed distribution. These results suggest that the separate ratio estimator not only achieves superior MSE performance but also has more desirable distributional properties. For all estimators, both skewness and kurtosis decrease as m_h increases from 3 to 8, indicating that the sampling distributions approach symmetry with larger samples, consistent with asymptotic normality.

CONCLUSION

This study proposed two ratio estimators for the population mean under SIRS, namely the separate ratio estimator $\bar{y}_{RS}$ and the combined ratio estimator $\bar{y}_{RC}$. Both estimators exploit the correlation between the study variable and a known auxiliary variable to improve estimation efficiency over the conventional unbiased estimator $\bar{y}_{st}$. Approximate expressions for the bias and mean squared error of each estimator were derived analytically using first-order Taylor series expansions, and consistent sample-based MSE estimators were constructed. Sufficient conditions under which the proposed estimators achieve lower MSE than unbiased estimator was also established.

The theoretical findings were validated through a Monte Carlo simulation study comprising 50,000 replications. The results demonstrated that both proposed estimators consistently outperformed the unbiased estimator across all considered values of m_h. From the relative efficiency, the separate ratio estimator achieved $\overline{RE}(\bar{y}_{RS})$ ranging from 1.586 to 2.089, while the combined ratio estimator achieved $\overline{RE}(\bar{y}_{RC})$ ranging from 1.353 to 1.563. These results indicate that the MSE of the unbiased estimator is up to approximately 2.09 times larger than that of the separate ratio estimator, confirming its substantial efficiency advantage. Furthermore, both relative efficiencies increased as m_h decreased, indicating that the efficiency gains of the proposed estimators are most pronounced when the number of sampled interested units per stratum is small - precisely the condition under which SIRS is most relevant in practice. In addition, both proposed estimators exhibited lower skewness, and kurtosis compared to the unbiased estimator, with all measures decreasing monotonically as m_h increased, consistent with the asymptotic properties implied by the large-sample approximations.

The proposed methodology is directly applicable to sample surveys involving rare populations, including epidemiological studies of uncommon medical conditions, ecological assessments of rare or endangered species, and agricultural surveys of infrequently occurring crop conditions. In such contexts, SIRS ensures adequate representation of rare units, while the proposed ratio estimators leverage auxiliary information to yield substantially more precise estimates of the population mean. Among the two proposed estimators, the separate ratio estimator is recommended when stratum-level auxiliary means are available and the stratum ratios are heterogeneous, as it consistently achieves the greatest efficiency gains. The combined ratio estimator serves as a practical alternative when only the overall population mean of the auxiliary variable is known.

Future research may pursue several promising extensions. These include the development of optimal allocation strategies for m_h across strata to minimize the MSE of the proposed estimators subject to sampling cost constraints and the generalization to regression and multi-auxiliary ratio estimators under SIRS.

ACKNOWLEDGEMENTS

We gratefully acknowledge the financial support from the Faculty of Science, Silpakorn University, Nakhon Pathom, Thailand (Grant No. SRIF-JRG-2569-02).

REFERENCES

Chao, C. 2004. Ratio estimation on adaptive cluster sampling. *Journal of the Chinese Statistical Association* 42: 307-327.

Christman, M.C. & Lan, F. 2001. Inverse adaptive cluster sampling. *Biometrics* 57(4): 1096-1105. https://doi.org/10.1111/j.0006-341X.2001.01096.x

Cochran, W.G. 1977. *Sampling Techniques*. 3rd ed. New York: John Wiley & Sons.

Finney, D.J. 1949. On a method of estimating frequencies. *Biometrika* 36(1-2): 233-234. https://doi.org/10.1093/biomet/36.1-2.233

Greco, L. & Naddeo, S. 2007. Inverse sampling with unequal selection probabilities. *Communications in Statistics - Theory and Methods* 36(5): 1039-1048. https://doi.org/10.1080/03610920601033926

Haldane, J.B.S. 1945. On a method of estimating frequencies. *Biometrika* 33(3): 222-225. https://doi.org/10.2307/2332299

Kadilar, C. & Cingi, H. 2004. Ratio estimators in simple random sampling. *Applied Mathematics and Computation* 151(3): 893-902. https://doi.org/10.1016/S0096-3003(03)00803-8

Murthy, M.N. 1964. Product method of estimation. *Sankhyā: The Indian Journal of Statistics Series A* 26: 69-74.

Salehi, M.M. & Seber, G.A.F. 2001. A new proof of Murthy's estimator which applies to sequential sampling. *Australian & New Zealand Journal of Statistics* 43(3): 281-286. https://doi.org/10.1111/1467-842X.00174

Sangngam, P. & Laoarun, W. 2025. Ratio estimators under inverse sampling. *Songklanakarin Journal of Science and Technology* 47(3): 192-199.

Sangngam, P. & Suwattee, P. 2012. Stratified inverse sampling for rare populations. *Thailand Statistician* 10(1): 69-86. https://ph02.tci-thaijo.org/index.php/thaistat/article/view/34233

Singh, H.P. & Nigam, P. 2020. Ratio-ratio-type exponential estimator of finite population mean in double sampling for stratification. *International Journal of Agricultural Statistics Sciences* 16(1): 251-257.

Sisodia, B.V.S. & Dwivedi, V.K. 1981. A modified ratio estimator using coefficient of variation of auxiliary variable. *Journal of the Indian Society of Agricultural Statistics* 33: 13-18.

Subramani, J. 2013. A new modified ratio estimator for estimation of population mean when median of the auxiliary variable is known. *Pakistan Journal of Statistics and Operation Research* 9(2): 137-145(2). https://doi.org/10.18187/pjsor.v9i2.486

Sungsuwan, S. & Suwattee, P. 2014. Model-assisted estimation in inverse sampling. *Chiang Mai Journal of Science* 41(3): 704-713.

*Corresponding author; email: LAOARUN_W@su.ac.th