Songklanakarin J. Sci. Technol.
46 (3), 246–255, May – Jun. 2024

http://www.sjst.psu.ac.th

Original Article

New transformed estimators in stratified random sampling: A case study on rubber production in Thailand

Natthapat Thongsak[1], and Nuanpan Lawson[2*]

[1] *State Audit Office of the Kingdom of Thailand, Phaya Thai, Bangkok, 10400 Thailand*

[2] *Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bang Sue, Bangkok, 10800 Thailand*

Received: 26 October 2023; Revised: 29 January 2024; Accepted: 4 January 2024

Abstract

Estimating the rubber production in Thailand, the world's leading rubber supplier, can help the Thai government to prepare for rubber cultivation in policy planning. A transformation technique can be used to improve the efficiency of estimating the average rubber yield by reducing the biases and mean square error. A group of population mean estimators has been suggested under stratified random sampling utilizing a transformed auxiliary variable. The biases and mean square errors of the proposed estimators are investigated. Simulation studies and an application to rubber production data in Thailand have been applied to assess their performances under stratified random sampling where the yield of rubber varies depending upon the region. The results show that the estimates of rubber yields with the proposed estimators had small biases and mean square errors. The best estimator gave an estimated rubber production of 1,140 kilogram/hectare, which is close to the population mean of the yields of rubber.

Keywords: rubber production, stratified random sampling, transformed auxiliary variable, bias, mean square error

1. Introduction

Rubber production in Thailand is among the largest ones in the world and gains income from natural rubber exports all year round. The southern region of Thailand has abundant rubber cultivation, as it is a suitable location in the tropical country. Knowledge of the estimated supply of rubber can be useful for assisting planning and policies of the government, in order to manage investments in the rubber industry. Rubber yields differ by region of production, located largely in the south and in some other regions of Thailand. Thongsak and Lawson (2021) applied population mean ratio estimators to rubber data in Thailand under simple random sampling without replacement (SRSWOR). They considered rubber data as the study variable and the cultivated area for the districts in Thailand as the auxiliary variable. Thongsak and Lawson (2023a) studied the biases and mean square errors (MSEs) of the population mean estimators under double sampling and applied them to rubber production data in Thailand following Thongsak and Lawson (2021).

Stratified sampling proves to be advantageous when dealing with a population characterized by heterogeneous subgroups. It divides the population into subgroups called strata, for homogeneity within a stratum and heterogeneity between different strata. This enables researchers to ensure comprehensive representation of all such subgroups within the selected sample. Therefore, this approach is suitable for conducting a survey of rubber data in Thailand, due to the differences in rubber production based on the cultivated areas in each region. One of the renowned estimators is the population mean ratio estimator suggested by Cochran (1940), which is divided into two types under stratified random sampling; a separate and a combined ratio estimator. To improve the population mean estimate of the variable of interest, several researchers proposed ratio estimators under stratified random sampling by adopting the ratio estimators

*Corresponding author
Email address: nuanpan.n@sci.kmutnb.ac.th

under SRSWOR that use the coefficients of variation, kurtosis, and mid-range. Tailor and Lone (2014) proposed four separate ratio estimators using the coefficients of variation, kurtosis, and a combination of the two by adopting the ratio estimators under SRSWOR that were proposed by Sisodia and Dwivedi (1981), Singh, Tailor, Tailor, and Kakran (2004) and Upadhyaya and Singh (1999) under stratified random sampling. Bhushan, Kumar, Lone, Anwar, and Gunaime (2023) recommend two classes of population mean estimators under stratified random sampling. Their estimators are in the form of logarithm and represent either separate or combined ratio estimators. Singh, Gupta, and Tailor (2023) introduced two new classes of population mean estimators in the form of exponentials using the transformed auxiliary variable under stratified random sampling. Their estimators are in the form of the combined estimators which use the values that make MSEs optimal (e.g., Kadilar & Cingi, 2003, 2005; Maqbool, Subzar, & Bhat, 2017).

The transformation of variables is also implemented to increase the efficacy by changing the shape of the distribution, leading to a more accurate and powerful population mean estimator. Under SRSWOR, Srivenkataramana (1980) employed the transformation technique to transform an auxiliary variable, which has been promoted by many researchers (e.g. Bandyopadhyaya, 1980; Onyeka, Nlebedim, & Izunobi, 2013; Singh & Upadhyaya, 1986; Yadav, Singh, Upadhyaya, & Yadav, 2024). Thongsak and Lawson (2021) suggested two classes of estimators using the transformation technique proposed by Srivenkataramana (1980) to transform an auxiliary variable under SRSWOR. Under suitable conditions, they were superior to the non-transformed estimators (e.g. Lawson, 2023; Thongsak & Lawson, 2023b, 2023c).

Motivated by the Thongsak and Lawson (2021) estimators, we proposed new estimators utilizing the same transformation method to change the distribution shape of an auxiliary variable in stratified random sampling. The formulas of biases and MSEs of the proposed estimators have been acquired. To compare the performances of the population mean estimators, the MSE is used as the criterion based on theory, simulation studies, and the application to rubber production data in Thailand.

2. Materials and Methods

2.1 Existing estimators

A population of size N is divided into L strata with each stratum of size $N_h\left(h=1,2,3,...,L\right)$, such that $\sum_{h=1}^{L} N_h = N$. Let $\left(x_i, y_i\right); i=1,2,3,...,N$ be the pairs of the auxiliary and study variables, respectively. A sample of size n_h is selected from each stratum using SRSWOR, such that $\sum_{h=1}^{L} n_h = n$. The $\hat{\bar{Y}}_{RS}$ is

$$\hat{\bar{Y}}_{RS} = \sum_{h=1}^{L} W_h \bar{y}_h \left(\frac{\bar{X}_h}{\bar{x}_h}\right), \tag{1}$$

where $\bar{X}_h = \sum_{i=1}^{N_h} x_{hi} \Big/ N_h$ and $\bar{Y}_h = \sum_{i=1}^{N_h} y_{hi} \Big/ N_h$ are the population mean of the auxiliary and study variables in stratum h, $\bar{x}_h = \sum_{i=1}^{n_h} x_{hi} \Big/ n_h$ and $\bar{y}_h = \sum_{i=1}^{n_h} y_{hi} \Big/ n_h$ are the sample means of the auxiliary and study variables in stratum h, respectively, and $W_h = \frac{N_h}{N}$ is the stratum weight.

The bias and MSE of $\hat{\bar{Y}}_{RS}$ are respectively

$$Bias\left(\hat{\bar{Y}}_{RS}\right) = \sum_{h=1}^{L} W_h \gamma_h \bar{Y}_h \left(C_{xh}^2 - \rho_h C_{xh} C_{yh}\right), \tag{2}$$

$$MSE\left(\hat{\bar{Y}}_{RS}\right) = \sum_{h=1}^{L} W_h^2 \gamma_h \bar{Y}_h^2 \left(C_{yh}^2 + C_{xh}^2 - 2\rho_h C_{xh} C_{yh}\right), \tag{3}$$

where $\gamma_h = \frac{1}{n_h} - \frac{1}{N_h}$, $C_{xh} = S_{xh} / \bar{X}_h$ is the population coefficient of variation of the auxiliary variable in stratum h, $\rho_h = \frac{S_{xyh}}{S_{xh} S_{yh}}$ is the population correlation coefficient between the auxiliary and study variables in stratum h, $S_{xh} = \sqrt{\frac{1}{N_h - 1}\sum_{i=1}^{N_h}\left(x_{hi} - \bar{X}_h\right)^2}$, $S_{yh} = \sqrt{\frac{1}{N_h - 1}\sum_{i=1}^{N_h}\left(y_{hi} - \bar{Y}_h\right)^2}$, and $S_{xyh} = \sqrt{\frac{1}{N_h - 1}\sum_{i=1}^{N_h}\left(x_{hi} - \bar{X}_h\right)\left(y_{hi} - \bar{Y}_h\right)}$

Tailor and Lone (2014) suggested four separate ratio estimators utilizing the advantage of the known coefficients of variation (C_{xh}), kurtosis (β_{2h} (x)), and a combination of the two by adopting the ratio estimators under SRSWOR suggested by Sisodia and Dwivedi (1981), Singh *et al.* (2004), and Upadhyaya and Singh (1999). Tailor and Lone (2014) estimators are

$$\hat{\bar{Y}}_{\text{Tailor \& Lone1}} = \sum_{h=1}^{L} W_h \bar{y}_h \left(\frac{\bar{X}_h + C_{xh}}{\bar{x}_h + C_{xh}} \right), \quad (4)$$

$$\hat{\bar{Y}}_{\text{Tailor \& Lone2}} = \sum_{h=1}^{L} W_h \bar{y}_h \left(\frac{\bar{X}_h + \beta_{2h}(x)}{\bar{x}_h + \beta_{2h}(x)} \right), \quad (5)$$

$$\hat{\bar{Y}}_{\text{Tailor \& Lone3}} = \sum_{h=1}^{L} W_h \bar{y}_h \left(\frac{\beta_{2h}(x)\bar{X}_h + C_{xh}}{\beta_{2h}(x)\bar{x}_h + C_{xh}} \right), \quad (6)$$

$$\hat{\bar{Y}}_{\text{Tailor \& Lone4}} = \sum_{h=1}^{L} W_h \bar{y}_h \left(\frac{C_{xh}\bar{X}_h + \beta_{2h}(x)}{C_{xh}\bar{x}_h + \beta_{2h}(x)} \right), \quad (7)$$

where $\beta_{2h}(x) = \frac{N_h(N_h+1)\sum_{i=1}^{N_h}\left(x_{hi} - \bar{X}\right)^4}{(N_h-1)(N_h-2)(N_h-3)S_{xh}^4} - \frac{3(N_h-1)^2}{(N_h-2)(N_h-3)}$ is the population coefficient of kurtosis of the auxiliary variable in stratum h.

The biases and MSEs of Tailor and Lone (2014)'s estimators are

$$Bias\left(\hat{\bar{Y}}_{\text{Tailor \& Lone1}}\right) = \sum_{h=1}^{L} W_h \gamma_h \bar{Y}_h \left(\left(\frac{\bar{X}_h}{\bar{X}_h + C_{xh}} \right)^2 C_{xh}^2 - \left(\frac{\bar{X}_h}{\bar{X}_h + C_{xh}} \right) \rho_h C_{xh} C_{yh} \right), \quad (8)$$

$$Bias\left(\hat{\bar{Y}}_{\text{Tailor \& Lone2}}\right) = \sum_{h=1}^{L} W_h \gamma_h \bar{Y}_h \left(\left(\frac{\bar{X}_h}{\bar{X}_h + \beta_{2h}(x)} \right)^2 C_{xh}^2 - \left(\frac{\bar{X}_h}{\bar{X}_h + \beta_{2h}(x)} \right) \rho_h C_{xh} C_{yh} \right), \quad (9)$$

$$Bias\left(\hat{\bar{Y}}_{\text{Tailor \& Lone3}}\right) = \sum_{h=1}^{L} W_h \gamma_h \bar{Y}_h \left(\left(\frac{\beta_{2h}(x)\bar{X}_h}{\beta_{2h}(x)\bar{X}_h + C_{xh}} \right)^2 C_{xh}^2 - \left(\frac{\beta_{2h}(x)\bar{X}_h}{\beta_{2h}(x)\bar{X}_h + C_{xh}} \right) \rho_h C_{xh} C_{yh} \right), \quad (10)$$

$$Bias\left(\hat{\bar{Y}}_{\text{Tailor \& Lone4}}\right) = \sum_{h=1}^{L} W_h \gamma_h \bar{Y}_h \left(\left(\frac{C_{xh}\bar{X}_h}{C_{xh}\bar{X}_h + \beta_{2h}(x)} \right)^2 C_{xh}^2 - \left(\frac{C_{xh}\bar{X}_h}{C_{xh}\bar{X}_h + \beta_{2h}(x)} \right) \rho_h C_{xh} C_{yh} \right), \quad (11)$$

$$MSE\left(\hat{\bar{Y}}_{\text{Tailor \& Lone1}}\right) = \sum_{h=1}^{L} W_h^2 \gamma_h \bar{Y}_h^2 \left(C_{yh}^2 + \left(\frac{\bar{X}_h}{\bar{X}_h + C_{xh}} \right)^2 C_{xh}^2 - 2\left(\frac{\bar{X}_h}{\bar{X}_h + C_{xh}} \right) \rho_h C_{xh} C_{yh} \right), \quad (12)$$

$$MSE\left(\hat{\bar{Y}}_{\text{Tailor \& Lone2}}\right) = \sum_{h=1}^{L} W_h^2 \gamma_h \bar{Y}_h^2 \left(C_{yh}^2 + \left(\frac{\bar{X}_h}{\bar{X}_h + \beta_{2h}(x)} \right)^2 C_{xh}^2 - 2\left(\frac{\bar{X}_h}{\bar{X}_h + \beta_{2h}(x)} \right) \rho_h C_{xh} C_{yh} \right), \quad (13)$$

$$MSE\left(\hat{\bar{Y}}_{\text{Tailor \& Lone3}}\right) = \sum_{h=1}^{L} W_h^2 \gamma_h \bar{Y}_h^2 \left(C_{yh}^2 + \left(\frac{\beta_{2h}(x)\bar{X}_h}{\beta_{2h}(x)\bar{X}_h + C_{xh}} \right)^2 C_{xh}^2 - 2\left(\frac{\beta_{2h}(x)\bar{X}_h}{\beta_{2h}(x)\bar{X}_h + C_{xh}} \right) \rho_h C_{xh} C_{yh} \right), \quad (14)$$

$$MSE\left(\hat{\bar{Y}}_{\text{Tailor \& Lone4}}\right) = \sum_{h=1}^{L} W_h^2 \gamma_h \bar{Y}_h^2 \left(C_{yh}^2 + \left(\frac{C_{xh}\bar{X}_h}{C_{xh}\bar{X}_h + \beta_{2h}(x)} \right)^2 C_{xh}^2 - 2\left(\frac{C_{xh}\bar{X}_h}{C_{xh}\bar{X}_h + \beta_{2h}(x)} \right) \rho_h C_{xh} C_{yh} \right). \quad (15)$$

Thongsak and Lawson (2021) derived two classes of ratio estimators in SRSWOR using the transformation method to ameliorate the population mean estimator. They suggested to use the transformation method to modify the general class of ratio estimators suggested by Jaroengeratikun and Lawson (2019), which used the assistance of known parameters. One of Thongsak and Lawson's (2021) estimators is

$$\hat{\bar{Y}}_{\text{Thongsak \& Lawson}} = \bar{y}\left(\frac{A\bar{x}^{*}+D}{A\bar{X}+D}\right), \quad (16)$$

where $\bar{x}^{*}=(1+\pi)\bar{X}-\pi\bar{x}$ is the transformed sample mean, $\pi = n/(N-n)$, $A \neq 0$ and D are constants or functions of the auxiliary variable.

The bias and MSE of the estimator are

$$Bias\left(\hat{\bar{Y}}_{\text{Thongsak \& Lawson}}\right) = -\gamma\pi\theta\bar{Y}\rho C_x C_y, \quad (17)$$

$$MSE\left(\hat{\bar{Y}}_{\text{Thongsak \& Lawson}}\right) = \gamma\bar{Y}^{2}\left(C_y^2+\theta^2\pi^2C_x^2-2\theta\pi\rho C_x C_y\right), \quad (18)$$

where $\theta=\frac{A\bar{X}}{A\bar{X}+D}, \gamma=\frac{1}{n}-\frac{1}{N}, \rho$ is the correlation coefficient between the auxiliary and study variables, and C_x, C_y are the coefficients of variation of the auxiliary variable and study variable, respectively.

Some of Thongsak and Lawson's (2021) estimators are shown in Table 1.

Table 1. Some of Thongsak and Lawson's (2021) estimators

Estimator	A	D
$\hat{\bar{Y}}_{\text{Thongsak \& Lawson1}}=\bar{y}\left(\frac{\bar{x}^{*}}{\bar{X}}\right)$	1	0
$\hat{\bar{Y}}_{\text{Thongsak \& Lawson2}}=\bar{y}\left(\frac{\bar{x}^{*}+C_x}{\bar{X}+C_x}\right)$	1	C_x
$\hat{\bar{Y}}_{\text{Thongsak \& Lawson3}}=\bar{y}\left(\frac{\bar{x}^{*}+\beta_2(x)}{\bar{X}+\beta_2(x)}\right)$	1	$\beta_2(x)$
$\hat{\bar{Y}}_{\text{Thongsak \& Lawson4}}=\bar{y}\left(\frac{\beta_2(x)\bar{x}^{*}+C_x}{\beta_2(x)\bar{X}+C_x}\right)$	$\beta_2(x)$	C_x
$\hat{\bar{Y}}_{\text{Thongsak \& Lawson5}}=\bar{y}\left(\frac{C_x\bar{x}^{*}+\beta_2(x)}{C_x\bar{X}+\beta_2(x)}\right)$	C_x	$\beta_2(x)$

We can see that some of Thongsak and Lawson's (2021) transformed estimators under SRSWOR are of the same form as the estimators proposed by Tailor and Lone (2014) under stratified random sampling, but they are not transformed estimators.

2.2 Proposed estimators

A class of estimators under stratified random sampling utilizing the transformed auxiliary variable was suggested following Thongsak and Lawson's (2021) idea. The class of the proposed estimators is

$$\hat{\bar{Y}}_{N} = \sum_{h=1}^{L} W_h\bar{y}_h\left(\frac{A_h\bar{x}_h^{*}+D_h}{A_h\bar{X}_h+D_h}\right), \quad (19)$$

where $\bar{x}_h^{*}=(1+\pi_h)\bar{X}_h-\pi_h\bar{x}_h$ is the transformed sample mean of an auxiliary variable in stratum h, $\pi_h = n_h/(N_h-n_h)$, $A_h \neq 0$ and D_h are constants or functions of the auxiliary variable in in stratum h.

Let $\varepsilon_{0h}=\frac{\bar{y}_h-\bar{Y}_h}{\bar{Y}_h}$ so that $\bar{y}_h=(1+\varepsilon_{0h})\bar{Y}_h$, and $\varepsilon_{1h}=\frac{\bar{x}_h-\bar{X}_h}{\bar{X}_h}$ so that $\bar{x}_h=(1+\varepsilon_{1h})\bar{X}_h$ and $\bar{x}_h^{*}=(1-\pi_h\varepsilon_{1h})\bar{X}_h$; then $E(\varepsilon_{0h})=E(\varepsilon_{1h})=0, E(\varepsilon_0^2)=\gamma C_y^2, E(\varepsilon_1^2)=\gamma C_x^2, E(\varepsilon_0\varepsilon_1)=\gamma\rho C_x C_y$.

Rewriting Equation (19) in a form using ε_{0h} and ε_{1h} we have:

$$\hat{\bar{Y}}_N = \sum_{h=1}^{L} W_h \left(1+\varepsilon_{0h}\right)\bar{Y}_h \left(\frac{\left(A_h \bar{X}_h + D_h\right) - \pi_h \varepsilon_{1h} A_h \bar{X}_h}{A_h \bar{X}_h + D_h}\right). \quad (20)$$

Let $\theta_h = \dfrac{A_h \bar{X}_h}{A_h \bar{X}_h + D_h}$, to get

$$\hat{\bar{Y}}_N = \sum_{h=1}^{L} W_h \bar{Y}_h \left(1+\varepsilon_{0h}\right)\left(\frac{\dfrac{A_h \bar{X}_h}{\theta_h} - \pi_h \varepsilon_{1h} A_h \bar{X}_h}{\dfrac{A_h \bar{X}_h}{\theta_h}}\right)$$

$$= \sum_{h=1}^{L} W_h \bar{Y}_h \left(1+\varepsilon_{0h}\right)\left(\frac{\dfrac{1-\pi_h \theta_h \varepsilon_{1h}}{\theta_h}}{\dfrac{1}{\theta_h}}\right)$$

$$= \sum_{h=1}^{L} W_h \bar{Y}_h \left(1+\varepsilon_{oh} - \pi_h \theta_h \varepsilon_{1h} - \pi_h \theta_h \varepsilon_{0h} \varepsilon_{1h}\right).$$

So the estimation error of $\hat{\bar{Y}}_N$ is

$$\hat{\bar{Y}}_N - \bar{Y} = \sum_{h=1}^{L} W_h \bar{Y}_h \left(\varepsilon_{oh} - \pi_h \theta_h \varepsilon_{1h} - \pi_h \theta_h \varepsilon_{0h} \varepsilon_{1h}\right).$$

From approximation using Taylor linearization, the bias of $\hat{\bar{Y}}_N$ to the first degree is

$$\begin{aligned} Bias\left(\hat{\bar{Y}}_N\right) &= E\left(\hat{\bar{Y}}_N - \bar{Y}\right) \\ &= -\sum_{h=1}^{L} W_h \gamma_h \pi_h \theta_h \bar{Y}_h \rho_h C_{xh} C_{yh}, \end{aligned} \quad (21)$$

and the MSE of $\hat{\bar{Y}}_N$ is

$$\begin{aligned} MSE\left(\hat{\bar{Y}}_N\right) &= E\left(\hat{\bar{Y}}_N - \bar{Y}\right)^2 \\ &\cong E\left(\sum_{h=1}^{L} W_h \bar{Y}_h \left(\varepsilon_{oh} - \pi_h \theta_h \varepsilon_{1h}\right)\right)^2 \\ &= \sum_{h=1}^{L} W_h^2 \bar{Y}_h^2 E\left(\varepsilon_{oh}^2 + \pi_h^2 \theta_h^2 \varepsilon_{1h}^2 - 2\pi_h \theta_h \varepsilon_{oh} \varepsilon_{1h}\right) \\ &= \sum_{h=1}^{L} W_h^2 \gamma_h \bar{Y}_h^2 \left(C_{yh}^2 + \theta_h^2 \pi_h^2 C_{xh}^2 - 2\theta_h \pi_h \rho_h C_{xh} C_{yh}\right). \end{aligned} \quad (22)$$

Note that from Equation (22) the unknown parameters can be estimated using the sample values. For instance, *r*, the sample correlation coefficient between the auxiliary and study variables can estimate *ρ*.

Some of the proposed estimators are in Table 2.

2.3 Efficiency comparisons

The Tailor and Lone (2014) estimators under stratified random sampling and the Thongsak and Lawson (2021) estimator under SRSWOR are compared with the proposed estimators. The details are as follows.

1) The proposed estimator is superior to the usual separate ratio estimator under certain conditions, namely:

$$MSE\left(\hat{\bar{Y}}_N\right) < MSE\left(\hat{\bar{Y}}_{RS}\right)$$

$$\sum_{h=1}^{L} W_h^2 \gamma_h \bar{Y}_h^2 C_{xh}^2 \left(\theta_h^2 \pi_h^2 - 1\right) < 2\sum_{h=1}^{L} W_h^2 \gamma_h \bar{Y}_h^2 \rho_h C_{xh} C_{yh} \left(\theta_h \pi_h - 1\right) \quad (23)$$

Table 2. The proposed estimators, $\hat{\bar{Y}}_{Ni}, i = 1,2,...,5$

Estimator	A_h	D_h
$\hat{\bar{Y}}_{N1} = \sum_{h=1}^{L} W_h \bar{y}_h \left(\frac{\bar{x}_h^*}{\bar{X}_h} \right)$	1	0
$\hat{\bar{Y}}_{N2} = \sum_{h=1}^{L} W_h \bar{y}_h \left(\frac{\bar{x}_h^* + C_{xh}}{\bar{X}_h + C_{xh}} \right)$	1	C_{xh}
$\hat{\bar{Y}}_{N3} = \sum_{h=1}^{L} W_h \bar{y}_h \left(\frac{\bar{x}_h^* + \beta_{2h}(x)}{\bar{X}_h + \beta_{2h}(x)} \right)$	1	$\beta_{2h}(x)$
$\hat{\bar{Y}}_{N4} = \sum_{h=1}^{L} W_h \bar{y}_h \left(\frac{\beta_{2h}(x)\bar{x}_h^* + C_{xh}}{\beta_{2h}(x)\bar{X}_h + C_{xh}} \right)$	$\beta_{2h}(x)$	C_{xh}
$\hat{\bar{Y}}_{N5} = \sum_{h=1}^{L} W_h \bar{y}_h \left(\frac{C_{xh}\bar{x}_h^* + \beta_{2h}(x)}{C_{xh}\bar{X}_h + \beta_{2h}(x)} \right)$	C_{xh}	$\beta_{2h}(x)$

2) The proposed estimator is superior to the Tailor and Lone (2014) estimator ($\hat{\bar{Y}}_{\text{Tailor \& Lone1}}$) under certain conditions, namely:

$$MSE\left(\hat{\bar{Y}}_N\right) < MSE\left(\hat{\bar{Y}}_{\text{Tailor \& Lone1}}\right)$$

$$\sum_{h=1}^{L} W_h^2 \gamma_h \bar{Y}_h^2 C_{xh}^2 \left(\theta_h^2 \pi_h^2 - \left(\frac{\bar{X}_h}{\bar{X}_h + C_{xh}} \right)^2 \right) < 2\sum_{h=1}^{L} W_h^2 \gamma_h \bar{Y}_h^2 \rho_h C_{xh} C_{yh} \left(\theta_h \pi_h - \frac{\bar{X}_h}{\bar{X}_h + C_{xh}} \right) \tag{24}$$

3) The proposed estimator is superior to the Tailor and Lone (2014) estimator ($\hat{\bar{Y}}_{\text{Tailor \& Lone2}}$) under certain conditions, namely:

$$MSE\left(\hat{\bar{Y}}_N\right) < MSE\left(\hat{\bar{Y}}_{\text{Tailor \& Lone2}}\right)$$

$$\sum_{h=1}^{L} W_h^2 \gamma_h \bar{Y}_h^2 C_{xh}^2 \left(\theta_h^2 \pi_h^2 - \left(\frac{\bar{X}_h}{\bar{X}_h + \beta_{2h}(x)} \right)^2 \right) < 2\sum_{h=1}^{L} W_h^2 \gamma_h \bar{Y}_h^2 \rho_h C_{xh} C_{yh} \left(\theta_h \pi_h - \frac{\bar{X}_h}{\bar{X}_h + \beta_{2h}(x)} \right) \tag{25}$$

4) The proposed estimator is superior to the Tailor and Lone (2014) estimator ($\hat{\bar{Y}}_{\text{Tailor \& Lone3}}$) under certain conditions, namely:

$$MSE\left(\hat{\bar{Y}}_N\right) < MSE\left(\hat{\bar{Y}}_{\text{Tailor \& Lone3}}\right)$$

$$\sum_{h=1}^{L} W_h^2 \gamma_h \bar{Y}_h^2 C_{xh}^2 \left(\theta_h^2 \pi_h^2 - \left(\frac{\beta_{2h}(x)\bar{X}_h}{\beta_{2h}(x)\bar{X}_h + C_{xh}} \right)^2 \right) < 2\sum_{h=1}^{L} W_h^2 \gamma_h \bar{Y}_h^2 \rho_h C_{xh} C_{yh} \left(\theta_h \pi_h - \frac{\beta_{2h}(x)\bar{X}_h}{\beta_{2h}(x)\bar{X}_h + C_{xh}} \right) \tag{26}$$

5) The proposed estimator is superior to the Tailor and Lone (2014) estimator ($\hat{\bar{Y}}_{\text{Tailor \& Lone4}}$) under certain conditions, namely:

$$MSE\left(\hat{\bar{Y}}_N\right) < MSE\left(\hat{\bar{Y}}_{\text{Tailor \& Lone4}}\right)$$

$$\sum_{h=1}^{L} W_h^2 \gamma_h \bar{Y}_h^2 C_{xh}^2 \left(\theta_h^2 \pi_h^2 - \left(\frac{C_{xh}\bar{X}_h}{C_{xh}\bar{X}_h + \beta_{2h}(x)} \right)^2 \right) < 2\sum_{h=1}^{L} W_h^2 \gamma_h \bar{Y}_h^2 \rho_h C_{xh} C_{yh} \left(\theta_h \pi_h - \frac{C_{xh}\bar{X}_h}{C_{xh}\bar{X}_h + \beta_{2h}(x)} \right) \tag{27}$$

Equations (23) to (27), can be rewritten in a general form as follows.

$$\sum_{h=1}^{L} W_h^2 \gamma_h \bar{Y}_h^2 C_{xh}^2 \left(\theta_h^2 \pi_h^2 - \Omega^2 \right) < 2\sum_{h=1}^{L} W_h^2 \gamma_h \bar{Y}_h^2 \rho_h C_{xh} C_{yh} \left(\theta_h \pi_h - \Omega \right) \tag{28}$$

If $\Omega = 1$, then $\hat{\bar{Y}}_N$ is better than $\hat{\bar{Y}}_{RS}$.

If $\Omega = \dfrac{\bar{X}_h}{\bar{X}_h + C_{xh}}$, then $\hat{\bar{Y}}_N$ is better than $\hat{\bar{Y}}_{\text{Tailor \& Lone1}}$.

If $\Omega = \dfrac{\bar{X}_h}{\bar{X}_h + \beta_{2h}(x)}$, then $\hat{\bar{Y}}_N$ is better than $\hat{\bar{Y}}_{\text{Tailor \& Lone2}}$.

If $\Omega = \dfrac{\beta_{2h}(x)\bar{X}_h}{\beta_{2h}(x)\bar{X}_h + C_{xh}}$, then $\hat{\bar{Y}}_N$ is better than $\hat{\bar{Y}}_{\text{Tailor \& Lone3}}$.

If $\Omega = \dfrac{C_{xh}\bar{X}_h}{C_{xh}\bar{X}_h + \beta_{2h}(x)}$, then $\hat{\bar{Y}}_N$ is better than $\hat{\bar{Y}}_{\text{Tailor \& Lone4}}$.

6) The proposed estimator is superior to the Thongsak and Lawson (2021) estimator ($\hat{\bar{Y}}_{\text{Thongsak \& Lawson}}$) under certain conditions, namely:

$$MSE\left(\hat{\bar{Y}}_N\right) < MSE\left(\hat{\bar{Y}}_{\text{Thongsak \& Lawson}}\right)$$

$$\sum_{h=1}^{L} W_h^2 \gamma_h \bar{Y}_h^2 \left(C_{yh}^2 + \theta_h^2 \pi_h^2 C_{xh}^2 - 2\theta_h \pi_h \rho_h C_{xh} C_{yh}\right) < \gamma \bar{Y}^2 \left(C_y^2 + \theta^2 \pi^2 C_x^2 - 2\theta\pi\rho C_x C_y\right) \tag{29}$$

3. Results and Discussion

3.1 Simulation studies

We divide the population into three strata and generate the paired variable (*X*, *Y*) from the bivariate normal distribution for each stratum following the parameters below which satisfy the conditions in Equations (23)-(29).

1st stratum: $N_1 = 1{,}000, \bar{X}_1 = 400, \bar{Y}_1 = 500, C_{x1} = 1.2, C_{y1} = 0.3, \rho_1 = 0.8$

2nd stratum: $N_2 = 600, \bar{X}_2 = 550, \bar{Y}_2 = 700, C_{x2} = 1.0, C_{y2} = 0.8, \rho_2 = 0.6$

3rd stratum: $N_3 = 400, \bar{X}_3 = 550, \bar{Y}_3 = 350, C_{x3} = 0.9, C_{y3} = 1.2, \rho_3 = 0.4$

Samples of sizes n = 100, n = 200, n = 400 are drawn from the population of size N = 2,000 using SRSWOR and give each stratum proportional allocation. The sample sizes for each strata are n_1 = 50, n_2 = 30, n_3 = 20, for n = 100, n_1 = 100, n_2 = 60, n_3 = 40 for n = 200, and n_1 = 200, n_2 = 120, n_3 = 80 for n = 400. We repeated the simulation studies 10,000 times using R program (R Core Team, 2021).

The biases and MSEs of the estimators are calculated by

$$Bias\left(\hat{\bar{Y}}\right) = \frac{1}{10{,}000}\sum_{i=1}^{10{,}000}\left|\hat{\bar{Y}}_i - \bar{Y}\right|, \tag{30}$$

$$MSE\left(\hat{\bar{Y}}\right) = \frac{1}{10{,}000}\sum_{i=1}^{10{,}000}\left(\hat{\bar{Y}}_i - \bar{Y}\right)^2. \tag{31}$$

The biases and MSEs of the estimators are represented in Table 3.

Table 3. Biases and MSEs of the estimators

Estimator		n = 100		n = 200		n = 400	
		Bias	MSE	Bias	MSE	Bias	MSE
Tailor and Lone (2014) Existing estimators (non-transformed estimators under stratified random sampling)	$\hat{\bar{Y}}_{RS}$	43.31	3229.19	28.45	1317.38	18.70	553.34
	$\hat{\bar{Y}}_{\text{Tailor \& Lone1}}$	43.19	3208.60	28.38	1310.41	18.65	550.66
	$\hat{\bar{Y}}_{\text{Tailor \& Lone2}}$	43.31	3229.33	28.45	1317.42	18.70	553.35
	$\hat{\bar{Y}}_{\text{Tailor \& Lone3}}$	41.15	2914.06	27.06	1182.79	17.74	497.77
	$\hat{\bar{Y}}_{\text{Tailor \& Lone4}}$	43.31	3229.37	28.45	1317.43	18.70	553.36

Table 3. Continued.

Estimator		n = 100		n = 200		n = 400	
		Bias	MSE	Bias	MSE	Bias	MSE
Thongsak and Lawson (2021) Existing estimators (transformed estimators under SRSWOR)	$\hat{\hat{Y}}_{\text{Thongsak \& Lawson1}}$	30.68	1471.13	20.24	637.57	12.65	252.03
	$\hat{\hat{Y}}_{\text{Thongsak \& Lawson2}}$	30.68	1471.36	20.24	637.76	12.65	252.11
	$\hat{\hat{Y}}_{\text{Thongsak \& Lawson3}}$	30.68	1471.14	20.24	637.58	12.65	252.03
	$\hat{\hat{Y}}_{\text{Thongsak \& Lawson4}}$	30.77	1480.11	20.36	645.02	12.73	255.37
	$\hat{\hat{Y}}_{\text{Thongsak \& Lawson5}}$	30.68	1471.14	20.24	637.58	12.65	252.03
Proposed estimators (transformed estimators under stratified random sampling)	$\hat{\hat{Y}}_{N1}$	28.68	1299.07	18.97	563.85	11.73	216.79
	$\hat{\hat{Y}}_{N2}$	28.68	1299.31	18.97	564.02	11.74	216.84
	$\hat{\hat{Y}}_{N3}$	28.68	1299.07	18.97	563.84	11.73	216.78
	$\hat{\hat{Y}}_{N4}$	28.69	1299.79	18.95	562.88	11.66	213.94
	$\hat{\hat{Y}}_{N5}$	28.68	1299.07	18.97	563.84	11.73	216.78

The results in Table 3 show that the proposed estimators utilizing the transformed auxiliary variable under stratified random sampling gave less biases and MSEs compared to Tailor and Lone's (2014) estimator, the non-transformed estimators under stratified random sampling and Thongsak and Lawson's (2021) transformed estimator under SRSWOR. In a comparison to Tailor and Lone's (2014) estimator, the proposed transformed estimators gave an MSE around two times smaller for all sample sizes. Bigger sample sizes resulted in smaller biases and MSEs. The biases and MSEs compared between the sample sizes n = 100 and n = 400 show at least two times smaller biases and at least a six fold reduced MSEs.

3.2 Application to rubber production in Thailand

Rubber production data in Thailand are considered in this study to see the efficiency of the estimators (Office of Agricultural Economics, 2017). The cultivated area (hectare) and the yield of rubber (kilograms/hectare) in the district are considered as the auxiliary and the study variables, respectively. The data represent a population of size N = 746 districts. The parameters are

$$\bar{Y} = 1130.37,\ \bar{X} = 4,900.92,\ C_y = 0.29,\ C_x = 1.70,\ \rho = 0.59, \text{ and } \beta_2(x) = 9.81.$$

The data are divided by regions, 1: North (N_1 = 110), 2: North East (N_2 = 308), 3: West (N_3 = 39), 4: Central (N_4 = 84), 5: East (N_5 = 54), and 6: South (N_6 = 151). A sample n = 150 is taken from the population of size N = 746. Through proportional allocation, samples of sizes n_1 = 22, n_2 = 62, n_3 = 8, n_4 = 17, n_5 = 11, n_6 = 30 are randomly acquired from each stratum. The population parameters in each stratum are summarized in Table 4.

The MSEs of the estimators are presented in Table 5.

Table 4. Population parameters for each region

Region	North	North East	West
Parameters	$N_1 = 110$	$N_2 = 308$	$N_3 = 39$
	$n_1 = 22$	$n_2 = 62$	$n_3 = 8$
	$\bar{X}_1 = 1,234.43$	$\bar{X}_2 = 2,716.86$	$\bar{X}_3 = 1,725.01$
	$\bar{Y}_1 = 888.46$	$\bar{Y}_2 = 1,107.66$	$\bar{Y}_3 = 1,074.92$
	$C_{x1} = 1.45$	$C_{x2} = 1.64$	$C_{x3} = 1.74$
	$C_{y1} = 0.34$	$C_{y2} = 0.25$	$C_{y3} = 0.21$
	$\rho_1 = 0.61$	$\rho_2 = 0.55$	$\rho_3 = 0.66$
	$\beta_{21}(x) = 2.23$	$\beta_{22}(x) = 15.35$	$\beta_{23}(x) = 5.90$

Table 4. Continued.

Region	North	North East	West
Parameters	$N_4 = 84$	$N_5 = 54$	$N_6 = 151$
	$n_4 = 17$	$n_5 = 11$	$n_6 = 30$
	$\bar{X}_4 = 953.87$	$\bar{X}_5 = 6,979.89$	$\bar{X}_6 = 14,621.15$
	$\bar{Y}_4 = 845.12$	$\bar{Y}_5 = 1,119.89$	$\bar{Y}_6 = 1,529.65$
	$C_{x4} = 2.96$	$C_{x5} = 1.31$	$C_{x6} = 0.82$
	$C_{y4} = 0.22$	$C_{y5} = 0.24$	$C_{y6} = 0.09$
	$\rho_4 = 0.26$	$\rho_5 = 0.49$	$\rho_6 = 0.34$
	$\beta_{24}(x) = 28.83$	$\beta_{25}(x) = 7.31$	$\beta_{26}(x) = 1.86$

Table 5. Estimated values of rubber production, biases, and MSEs of the estimators

Estimator		Estimated values of rubber production	Bias	MSE
Tailor and Lone (2014) Existing estimators (non-transformed estimators under stratified random sampling)	$\hat{\bar{Y}}_{RS}$	1226.56	96.19	9253.41
	$\hat{\bar{Y}}_{\text{Tailor \& Lone1}}$	1226.46	96.09	9233.05
	$\hat{\bar{Y}}_{\text{Tailor \& Lone2}}$	1225.37	95.00	9024.37
	$\hat{\bar{Y}}_{\text{Tailor \& Lone3}}$	1226.57	96.20	9254.12
	$\hat{\bar{Y}}_{\text{Tailor \& Lone4}}$	1226.16	95.79	9175.32
Thongsak and Lawson (2021) Existing estimators (transformed estimators under SRSWOR)	$\hat{\bar{Y}}_{\text{Thongsak \& Lawson1}}$	1165.34	34.97	1223.19
	$\hat{\bar{Y}}_{\text{Thongsak \& Lawson2}}$	1165.33	34.96	1222.52
	$\hat{\bar{Y}}_{\text{Thongsak \& Lawson3}}$	1165.29	34.92	1219.33
	$\hat{\bar{Y}}_{\text{Thongsak \& Lawson4}}$	1165.34	34.97	1223.12
	$\hat{\bar{Y}}_{\text{Thongsak \& Lawson5}}$	1165.31	34.94	1220.91
Proposed estimators (transformed estimators under stratified random sampling)	$\hat{\bar{Y}}_{N1}$	1140.97	10.60	112.40
	$\hat{\bar{Y}}_{N2}$	1140.98	10.61	112.53
	$\hat{\bar{Y}}_{N3}$	1140.86	10.49	110.11
	$\hat{\bar{Y}}_{N4}$	1140.98	10.61	112.52
	$\hat{\bar{Y}}_{N5}$	1140.95	10.58	111.96

Table 5 reveals that the proposed estimators performed much better than Tailor and Lone's (2014) estimator, the non-transformed estimators under stratified random sampling and Thongsak and Lawson's (2021) transformed estimator under SRSWOR in terms of both smaller biases and MSEs. The proposed estimators gave similar estimated values for rubber production and also biases, and $\hat{\bar{Y}}_{N3}$ performed the best in terms of biases and MSEs for the rubber data production in Thailand. We can see that the estimated rubber production in Thailand from the proposed estimators is 1,140 kilograms/hectare in this situation, which is close to the population mean of the yields of rubber.

4. Conclusions

New transformed auxiliary variable estimators were presented in this study under stratified random sampling. The results from rubber data in Thailand show that the proposed estimators gave better estimates for rubber production than the existing estimators, the Tailor and Lone (2014) estimator, the non-transformed estimators under stratified random sampling, and the Thongsak and Lawson (2021) estimator, transformed estimators under SRSWOR. The proposed transformed estimators produced smaller biases and MSEs among all the tested estimators. The available parameters of the auxiliary variable gave a similar average rubber yield and also biases and MSEs. The best estimator uses the known coefficient of kurtosis based on the transformation technique. In future works, available parameters of the auxiliary variable can be applied to the proposed estimators to estimate the study variable. This class of proposed population mean estimators can be helpful for estimates from agricultural, economic, environmental, and other data related to real world problems.

Acknowledgements

This research was funded by King Mongkut's University of Technology North Bangkok, Contract no. KMUTNB-67-BASIC-30. We would like to thank all unknown referees for valuable comments.

References

Bandyopadhyaya, S. (1980). Improved ratio and product estimators. *Sankhya, Series C*, *42*, 45-49.

Bhushan, S., Kumar, A., Lone, S. A., Anwar, S., & Gunaime, N. M. (2023). An efficient class of estimators in stratified random sampling with an application to real data. *Axioms*, *12*(6), 576. doi:10.3390/axioms12060576

Cochran, W. G. (1940). *Sampling techniques* (3rd ed). India: Wiley Eastern Limited.

Jaroengeratikun, U. & Lawson, N. (2019). A combined family of ratio estimators for population mean using an auxiliary variable in simple random sampling. *Journal of Mathematical and Fundamental Sciences*, *51*(1), 1-12. doi:10.5614/j.math.fund.sci.2019.51.1.1

Kadilar, C., & Cingi, H. (2003). Ratio estimators in stratified random sampling. *Biometrical Journal*, *45*(2), 218-225. doi:10.1002/bimj.200390007

Kadilar, C., & Cingi, H. (2005). A new ratio estimator in stratified random sampling. *Communications in Statistics-Theory and Methods*, *34*(3), 597-602. doi:10.1081/STA-200052156

Lawson, N. (2023). An improved family of estimators for estimating population mean
using a transformed auxiliary variable under double sampling. *Songklanakarin*
Journal of Science and Technology, *45*(2), 165-172.

Maqbool, S., Subzar, M., & Bhat, M. A. (2017). Ratio estimator in stratified random sampling using mid-range as auxiliary information. *International Journal for Research Trends and Innovation*, *2*(8), 15-18. doi:10.6084/m9.doione.IJRTI1708004

Office of Agricultural Economics. (2020). Rubber production of Thailand. Retrieved from http://www.oae.go.th/assets/portals/1/fileups/prcaidata/files/1_rubber_dit%2060.pdf

Onyeka, A. C., Nlebedim, V. U., & Izunobi, C. H. (2013). Estimation of population ratio in simple random sampling using variable transformation. *Global Journal of Science Frontier Research Mathematics and Decision Sciences*, *13*(4), 57-65.

R Core Team. (2021). R: A language and environment for statistical computing. Vienna, Austri: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/.

Singh, H. P., Gupta, A. & Tailor, R. (2023). Efficient class of estimators for finite population mean using auxiliary attribute in stratified random sampling. *Scientific Reports*, *13*, 10253. doi:10.1038/s41598-023-34603-z

Singh, H. P., Tailor, R., Tailor, R., & Kakran , M. S. (2004). An improved estimator of population mean using power transformation. *Journal of the Indian Society of Agricultural Statistics*, *58*(2), 223-230.

Singh, H. P., & Upadhyaya, L. N. (1986). A dual to modified ratio estimator using coefficient of variation of auxiliary variable. *Proceedings of the National Academy of Sciences, India-Section A*, *56*(A), 336-340.

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*(1), 13-18.

Srivenkataramana, T. (1980). A dual to ratio estimator in sample surveys. *Biometrika*, *67*, 199-204. doi:10.2307/2335334

Tailor, R., & Lone, H. (2014). A. Separate ratio-type estimators of population mean in stratified random sampling. *Journal of Modern Applied Statistical Methods*, *13*(1), 223-233. doi:10.22237/jmasm/1398917580

Thongsak, N., & Lawson, N. (2021). Classes of dual to modified ratio estimators for estimating population mean in simple random sampling. *Proceedings of the 2021 Research, Invention, and Innovation Congress: Innovation Electricals and Electronics (RI2C)*, Bangkok, Thailand, 211-215. doi:10.1109/RI2C51727.2021.9559798

Thongsak, N. & Lawson, N. (2022). Classes of combined population mean estimators utilizing transformed variables under double sampling: An application to air pollution in Chiang Rai, Thailand. *Songklanakarin Journal of Science and Technology*, *44*(5), 1390-1398.

Thongsak, N., & Lawson, N. (2023a). Classes of population mean estimators using transformed variables in double sampling. *Gazi University Journal of Science*, *36*(4), 1834-1852. doi:10.35378/gujs.1056453

Thongsak, N., & Lawson, N. (2023b). Bias and mean square error reduction by changing the shape of the distribution of an auxiliary variable: Application to air pollution data in Nan, Thailand. *Mathematical Population Studies*, *3*(3), 180-190. doi:10.1080/08898480.2022.2145790

Thongsak, N. & Lawson, N. (2023c). A new imputation method for population mean in the presence of missing data based on a transformed variable with applications to air pollution data in Chiang Mai, Thailand. *Journal of Air Pollution and Health*, *8*(3), 285- 298.

Upadhyaya, L. N., & Singh, H. P. (1999). Use of transformed auxiliary variable in estimating the finite population mean. *Biometrical Journal*, *41*(5), 627-636.

Yadav, R., Singh, H. P., Upadhyaya, L. N., & Yadav, S. K. (2024). Adroit family of estimators of population mean using known auxiliary parameters. *Journal of Computational and Applied Mathematics*, *437*, 115455. doi:10.1016/j.cam.2023.115455