A Class of Population Mean Estimators in Stratified Random Sampling: A Case Study on Fine Particulate Matter in the North of Thailand

NUANPAN LAWSON[1*], NATTHAPAT THONGSAK[2]
[1]Department of Applied Statistics, Faculty of Applied Science,
King Mongkut's University of Technology North Bangkok,
1518 Pracharat 1 Road, Wongsawang, Bangsue, Bangkok 10800,
THAILAND

[2]State Audit Office of the Kingdom of Thailand,
Bangkok, 10400,
THAILAND

Abstract: - Residents of Thailand's upper northern have been facing hazardous air quality with the amount of fine particulate matter rising several times higher than the standards of the World Health Organization for many years which is classified as a level that severely affects public health. The dust problem is an urgent issue in Thailand that needs to be solved. Assessment of pollution data in advance can help the Thai government in planning to abolish and prevent ongoing dust problems for Thai citizens. A new class of population mean estimators is proposed under stratified random sampling. The bias and mean square error of the proposed estimators are studied using a Taylor series approximation. A simulation study and an application to air pollution data in the north of Thailand to investigate the performance of the estimators. The results from the air pollution data in the north of Thailand present that the proposed estimators offer the highest efficiency concerning others.

Key-Words: - Ratio estimator, Fine particulate matter, Mean square error, Stratified random sampling, Population mean, Thailand.

Received: November 3, 2023. Revised: November 19, 2023. Accepted: December 15, 2023. Published: March 8, 2024.

1 Introduction

Thailand's air pollution has become an increasing concern at present due to the amount of dust exceeding Thailand's and the World Health Organization's guidelines. Multiple provinces in the north where a 24-hour average of fine particulate matter with a diameter less than 2.5 microns (PM2.5) and fine particulate matter with a diameter less than 10 microns (PM10) on a 24-hour average exceed the standard of 50 micrograms per cubic meter and the air quality index (AQI) exceeds the standard value of 100 which is classified as levels that severely affect human health. In northern Thailand, there is an abundance of provinces tackling pollution from dust, due to seasonal agricultural practices and geography consisting of mountainous terrain accounting for the trapping of pollution that passes into the area. There is a significant correlation between pollution and citizens' health and quality of life, as shown by an increased incidence of upper respiratory symptoms and lung cancer, a worrying problem. Lung cancer due to PM2.5 pollution affects genetically susceptible people at a rapid and unprecedented rate, which is an increasing concern amongst the population. The Thai government needs to monitor the air quality data and assess the situation closely including implementing measures to control burning in forest areas, farmland, and areas along the highway strictly. Nevertheless, the government must enact a strict policy requesting people's cooperation to refrain from burning waste and agricultural waste in preparation for farming to prevent forest fires which can lead to dust issues in the north of Thailand.

Possessing air pollution data can assist in policy planning and preventing harm to life. Thailand's air pollution data have been investigated by many researchers. Carbon monoxide is used to estimate the average PM2.5 in Dindang, Thailand using the known prior information in simple random sampling without replacement (SRSWOR), [1]. The best estimators with the smallest mean square error (MSE) that were suggested by [1], are based on the known median of the auxiliary variable and quartile average. The [1], estimators assist in saving budget and time with a small sampling fraction. The transformation technique to create combined estimators was called upon to estimate the average PM2.5 via nitrogen dioxide in Chiang Rai under double sampling which results in improving the performance of the combined estimators concerning the single ones, [2]. The air pollution from vehicles in Selangor, Malaysia, an area with heavy traffic congestion has been studied in [3]. They found that PM10 and ozone are the key factors contributing to air pollution in Selangor although the air pollution there seemed to be reduced in the period of their study, [4], [5].

In sample surveys the ratio and regression estimators for estimating population mean, an average of a specified characteristic of a group, are well known in assisting to heighten the efficiency of estimators when an available auxiliary variable has a positive relation to a study variable, a ratio estimator was introduced for estimating the study variable population mean under SRSWOR when the population mean of the auxiliary variable is present, [6]. To increase the precision of the population mean estimator, some parameters of an auxiliary variable such as coefficient of variation, correlation, and kurtosis are applied to modify the usual ratio estimator, [7], [8], [9]. Five ratio estimators using a regression estimator and known parameters to estimate the population mean were proposed by [10], [11], [12], [13].

Stratified random sampling is one of the sampling techniques that is suitable for use when the units in the population are homogenous within the same stratum and heterogeneous between different strata. In the technique of stratified sampling, a population comprising N units is subdivided into distinct subpopulations, for example, it could be divided into geographical characteristics, age group, and gender. The subpopulations are called strata. Dividing strata is useful when we want to find the size of each stratum. The samples in each stratum are selected via simple random sampling. Under stratified random sampling, there are two types of ratio estimator; a separate and combined one. Ratio estimators were adopted by [14], under SRSWOR that were suggested by [7], [8], [9], which resulted in four separate ratio estimators under stratified random sampling and found that the proposed estimators are more efficient than the usual separate ratio estimator. More works that proposed to modify ratio estimator under stratified random sampling can be seen in [15], [16], [17], [18].

In the prevailing study, we aim to propose a new class of separate ratio estimators utilizing some parameters of an auxiliary variable under stratified random sampling. We acquired the formula of bias and mean square error of the proposed class of estimators up to the first-degree Taylor approximation. A simulation study and an application to air pollution in the north of Thailand are considered to demonstrate the capacity of the proposed class of estimators.

2 Existing Estimators

2.1 The Usual Separate Ratio Estimator

The usual separate ratio estimator for the population mean is

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

where $\bar{X}_h = \sum_{i=1}^{N_h} x_i \Big/ N_h$ is the population mean of an auxiliary variable in stratum h; $h = 1,2,3,...,L$, $\bar{x}_h = \sum_{i=1}^{n_h} x_i \Big/ n_h$ and $\bar{y}_h = \sum_{i=1}^{n_h} y_i \Big/ n_h$ are the sample means of the auxiliary and study variables in stratum h based on a sample of size n_h, respectively, $W_h = \frac{N_h}{N}$ is the stratum weight, and N_h is the population size in stratum h, such that $\sum_{h=1}^{L} N_h = N.$

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

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

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

where $C_{xh} = S_{xh} \big/ \bar{X}_h$ is the auxiliary variable's population coefficient of variation 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, $S_{xh} = \sqrt{\frac{1}{N_h - 1} \sum_{i=1}^{N_h} \left(x_i - \bar{X}_h \right)^2}$,

$$S_{yh}=\sqrt{\frac{1}{N_h-1}\sum_{i=1}^{N_h}\left(y_i-\bar{Y}_h\right)^2}\text{ , and}$$

$$S_{xyh}=\sqrt{\frac{1}{N_h-1}\sum_{i=1}^{N_h}\left(x_i-\bar{X}_h\right)\left(y_i-\bar{Y}_h\right)}$$

2.2 The Adjusted Ratio, [14], Estimators

Ratio estimators proposed by [8], [9], were adopted under SRSWOR to stratified random sampling, [14]. The [14], estimators are:

$$\hat{\bar{Y}}_{\mathrm{TL1}}=\sum_{h=1}^{L}W_h\bar{y}_h\left(\frac{\bar{X}_h+C_{xh}}{\bar{x}_h+C_{xh}}\right), \qquad (2)$$

$$\hat{\bar{Y}}_{\mathrm{TL2}}=\sum_{h=1}^{L}W_h\bar{y}_h\left(\frac{\bar{X}_h+\beta_{2h}(x)}{\bar{x}_h+\beta_{2h}(x)}\right), \qquad (3)$$

$$\hat{\bar{Y}}_{\mathrm{TL3}}=\sum_{h=1}^{L}W_h\bar{y}_h\left(\frac{\beta_{2h}(x)\bar{X}_h+C_{xh}}{\beta_{2h}(x)\bar{x}_h+C_{xh}}\right), \qquad (4)$$

$$\hat{\bar{Y}}_{\mathrm{TL4}}=\sum_{h=1}^{L}W_h\bar{y}_h\left(\frac{C_{xh}\bar{X}_h+\beta_{2h}(x)}{C_{xh}\bar{x}_h+\beta_{2h}(x)}\right), \qquad (5)$$

where

$$\beta_{2h}(x)=\frac{N_h(N_h+1)\sum_{i=1}^{N_h}\left(x_i-\bar{X}\right)^4}{(N_h-1)(N_h-2)(N_h-3)S_{xh}^4}-\frac{3(N_h-1)^2}{(N_h-2)(N_h-3)}$$

is the population coefficient of kurtosis of the auxiliary variable in stratum h.

The biases and MSEs of the [14], estimators are:

$$Bias\left(\hat{\bar{Y}}_{\mathrm{TL1}}\right)=\sum_{h=1}^{L}W_h\gamma_h\bar{Y}_h\left(\left(\frac{\bar{X}_h}{\bar{X}_h+C_{xh}}\right)^2C_{xh}^2-\left(\frac{\bar{X}_h}{\bar{X}_h+C_{xh}}\right)\rho_hC_{xh}C_{yh}\right),$$

$$Bias\left(\hat{\bar{Y}}_{\mathrm{TL2}}\right)=\sum_{h=1}^{L}W_h\gamma_h\bar{Y}_h\left(\left(\frac{\bar{X}_h}{\bar{X}_h+\beta_{2h}(x)}\right)^2C_{xh}^2-\left(\frac{\bar{X}_h}{\bar{X}_h+\beta_{2h}(x)}\right)\rho_hC_{xh}C_{yh}\right),$$

$$Bias\left(\hat{\bar{Y}}_{\mathrm{TL3}}\right)=\sum_{h=1}^{L}W_h\gamma_h\bar{Y}_h\left(\left(\frac{\beta_{2h}(x)\bar{X}_h}{\beta_{2h}(x)\bar{X}_h+C_{xh}}\right)^2C_{xh}^2-\left(\frac{\beta_{2h}(x)\bar{X}_h}{\beta_{2h}(x)\bar{X}_h+C_{xh}}\right)\rho_hC_{xh}C_{yh}\right),$$

$$Bias\left(\hat{\bar{Y}}_{\mathrm{TL4}}\right)=\sum_{h=1}^{L}W_h\gamma_h\bar{Y}_h\left(\left(\frac{C_{xh}\bar{X}_h}{C_{xh}\bar{X}_h+\beta_{2h}(x)}\right)^2C_{xh}^2-\left(\frac{C_{xh}\bar{X}_h}{C_{xh}\bar{X}_h+\beta_{2h}(x)}\right)\rho_hC_{xh}C_{yh}\right),$$

$$MSE\left(\hat{\bar{Y}}_{\mathrm{TL1}}\right)=\sum_{h=1}^{L}W_h^2\gamma_h\bar{Y}_h^2\left(C_{yh}^2+\left(\frac{\bar{X}_h}{\bar{X}_h+C_{xh}}\right)^2C_{xh}^2-2\left(\frac{\bar{X}_h}{\bar{X}_h+C_{xh}}\right)\rho_hC_{xh}C_{yh}\right),$$

$$MSE\left(\hat{\bar{Y}}_{\mathrm{TL2}}\right)=\sum_{h=1}^{L}W_h^2\gamma_h\bar{Y}_h^2\left(C_{yh}^2+\left(\frac{\bar{X}_h}{\bar{X}_h+\beta_{2h}(x)}\right)^2C_{xh}^2-2\left(\frac{\bar{X}_h}{\bar{X}_h+\beta_{2h}(x)}\right)\rho_hC_{xh}C_{yh}\right),$$

$$MSE\left(\hat{\bar{Y}}_{\mathrm{TL3}}\right)=\sum_{h=1}^{L}W_h^2\gamma_h\bar{Y}_h^2\left(C_{yh}^2+\left(\frac{\beta_{2h}(x)\bar{X}_h}{\beta_{2h}(x)\bar{X}_h+C_{xh}}\right)^2C_{xh}^2-2\left(\frac{\beta_{2h}(x)\bar{X}_h}{\beta_{2h}(x)\bar{X}_h+C_{xh}}\right)\rho_hC_{xh}C_{yh}\right),$$

$$MSE\left(\hat{\bar{Y}}_{\mathrm{TL4}}\right)=\sum_{h=1}^{L}W_h^2\gamma_h\bar{Y}_h^2\left(C_{yh}^2+\left(\frac{C_{xh}\bar{X}_h}{C_{xh}\bar{X}_h+\beta_{2h}(x)}\right)^2C_{xh}^2-2\left(\frac{C_{xh}\bar{X}_h}{C_{xh}\bar{X}_h+\beta_{2h}(x)}\right)\rho_hC_{xh}C_{yh}\right).$$

The usual separate ratio in (1) and [14], estimators in (2) to (5) in a general form are:

$$\hat{\bar{Y}}_{\mathrm{R}}=\sum_{h=1}^{L}W_h\bar{y}_h\left(\frac{A_h\bar{X}_h+D_h}{A_h\bar{x}_h+D_h}\right), \qquad (6)$$

The bias and MSE of the $\hat{\bar{Y}}_{\mathrm{R}}$ in general forms are"

$$Bias\left(\hat{\bar{Y}}_{\mathrm{R}}\right)=\sum_{h=1}^{L}W_h\gamma_h\bar{Y}_h\left(\theta_h^2C_{xh}^2-\theta_h\rho_hC_{xh}C_{yh}\right), \qquad (7)$$

$$MSE\left(\hat{\bar{Y}}_{\mathrm{R}}\right)=\sum_{h=1}^{L}W_h^2\gamma_h\bar{Y}_h^2\left(C_{yh}^2+\theta_h^2C_{xh}^2-2\theta_h\rho_hC_{xh}C_{yh}\right). \qquad (8)$$

where $A_h\neq 0$ and D_h are constants or functions of the auxiliary variable in stratum h, and $\theta_h=\dfrac{A_h\bar{X}_h}{A_h\bar{X}_h+D_h}$.

3 Proposed Estimator

A new group of separate ratio estimators is shown under stratified random sampling inspired by the work of [10], under SRS.

$$\hat{\bar{Y}}_{\mathrm{Reg}}=\sum_{h=1}^{L}W_h\left[\bar{y}_h+b_h\left(\bar{X}_h-\bar{x}_h\right)\right]\left(\frac{A_h\bar{X}_h+D_h}{A_h\bar{x}_h+D_h}\right), \qquad (9)$$

where b_h is a sample regression coefficient of β_h in stratum h.

Let $\varepsilon_{0h}=\dfrac{\bar{y}_h-\bar{Y}_h}{\bar{Y}_h}$ then $\bar{y}_h=\left(1+\varepsilon_{0h}\right)\bar{Y}_h$, let $\varepsilon_{1h}=\dfrac{\bar{x}_h-\bar{X}_h}{\bar{X}_h}$ then $\bar{x}_h=\left(1+\varepsilon_{1h}\right)\bar{X}_h$, such that $E\left(\varepsilon_{0h}\right)=E\left(\varepsilon_{1h}\right)=0, E\left(\varepsilon_0^2\right)=\gamma C_y^2, E\left(\varepsilon_1^2\right)=\gamma C_x^2$, $E\left(\varepsilon_0\varepsilon_1\right)=\gamma\rho C_xC_y$.

Rewriting (9) in terms of ε_{0h} and ε_{1h},

$$\hat{\bar{Y}}_{\text{Reg}} = \sum_{h=1}^{L} W_h \left[\left(1+\varepsilon_{0h}\right)\bar{Y}_h - \varepsilon_{1h} b_h \bar{X}_h \right] \left(\frac{A_h \bar{X}_h + D_h}{A_h\left(1+\varepsilon_{1h}\right)\bar{X}_h + D_h} \right)$$

$$= \sum_{h=1}^{L} W_h \left[\bar{Y}_h + \varepsilon_{0h}\bar{Y}_h - \varepsilon_{1h} b_h \bar{X}_h \right] \left(1+\varepsilon_{1h}\theta_h\right)^{-1}$$

$$\cong \sum_{h=1}^{L} W_h \left[\bar{Y}_h + \varepsilon_{0h}\bar{Y}_h - \varepsilon_{1h} b_h \bar{X}_h \right] \left(1-\varepsilon_{1h}\theta_h + \varepsilon_{1h}^2\theta_h^2\right)$$

$$= \sum_{h=1}^{L} W_h \left(\bar{Y}_h + \varepsilon_{0h}\bar{Y}_h - \varepsilon_{1h} b_h \bar{X}_h - \varepsilon_{1h}\bar{Y}_h\theta_h \right.$$
$$\left. -\varepsilon_{0h}\varepsilon_{1h}\bar{Y}_h\theta_h + \varepsilon_{1h}^2\theta_h b_h \bar{X}_h + \varepsilon_{1h}^2\theta_h^2\bar{Y}_h \right)$$

The estimation error of $\hat{\bar{Y}}_{\text{Reg}}$ is:

$$\hat{\bar{Y}}_{\text{Reg}} - \bar{Y} = \sum_{h=1}^{L} W_h \left(\varepsilon_{0h}\bar{Y}_h - \varepsilon_{1h} b_h \bar{X}_h - \varepsilon_{1h}\bar{Y}_h\theta_h \right.$$
$$\left. -\varepsilon_{0h}\varepsilon_{1h}\bar{Y}_h\theta_h + \varepsilon_{1h}^2\theta_h b_h \bar{X}_h + \varepsilon_{1h}^2\theta_h^2\bar{Y}_{hh} \right)$$

The $\hat{\bar{Y}}_{\text{Reg}}$ bias

$$Bias\left(\hat{\bar{Y}}_{\text{Reg}}\right) = E\left(\hat{\bar{Y}}_{\text{Reg}} - \bar{Y}\right)$$

$$\cong \sum_{h=1}^{L} W_h\gamma_h\theta_h\bar{Y}_h \left[\left(\beta_h K_h + \theta_h\right) C_{xh}^2 - \rho_h C_{xh} C_{yh} \right], \tag{10}$$

The $\hat{\bar{Y}}_{\text{Reg}}$ MSE

$$MSE\left(\hat{\bar{Y}}_{\text{Reg}}\right) = E\left(\hat{\bar{Y}}_{\text{Reg}} - \bar{Y}\right)^2$$

$$\cong E\left(\sum_{h=1}^{L} W_h \left(\varepsilon_{0h}\bar{Y}_h - \varepsilon_{1h} b_h \bar{X}_h - \varepsilon_{1h}\theta_h\bar{Y}_h \right) \right)^2$$

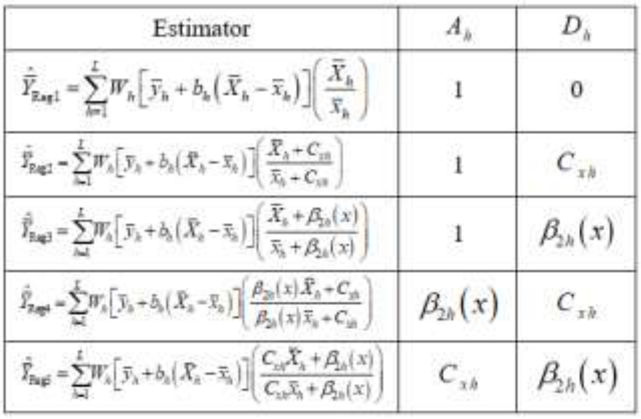
$$= \sum_{h=1}^{L} W_h^2 E\left(\varepsilon_{0h}\bar{Y}_h - \varepsilon_{1h} b_h \bar{X}_h - \varepsilon_{1h}\bar{Y}_h\theta_h \right)^2$$

$$= \sum_{h=1}^{L} W_h^2 E\left(\varepsilon_{0h}^2\bar{Y}_h^2 + \varepsilon_{1h}^2 b_h^2 \bar{X}_h^2 + \varepsilon_{1h}^2\theta_h^2\bar{Y}_h^2 - 2\varepsilon_{0h}\varepsilon_{1h} b_h \bar{X}_h \bar{Y}_h \right.$$
$$\left. -2\varepsilon_{0h}\varepsilon_{1h}\theta_h\bar{Y}_h^2 + 2\varepsilon_{1h}^2 b_h\theta_h\bar{X}_h\bar{Y}_h \right)$$

$$= \sum_{h=1}^{L} W_h^2\gamma_h\bar{Y}_h^2 \left(C_{yh}^2 + \left(\theta_h + \beta_h K_h\right)^2 C_{xh}^2 - 2\left(\theta_h + \beta_h K_h\right)\rho_h C_{xh} C_{yh} \right) \tag{11}$$

Some members of $\hat{\bar{Y}}_{\text{Reg}}$ are represented in Table 1.

Table 1. Some of the proposed estimators, $\hat{\bar{Y}}_{\text{Reg}i}, i = 1,2,...,5$



Estimator	A_h	D_h
$\hat{\bar{Y}}_{\text{Reg}1} = \sum_{h=1}^{L} W_h\left[\bar{y}_h + b_h\left(\bar{X}_h - \bar{x}_h\right)\right]\left(\frac{\bar{X}_h}{\bar{x}_h}\right)$	1	0
$\hat{\bar{Y}}_{\text{Reg}2} = \sum_{h=1}^{L} W_h\left[\bar{y}_h + b_h\left(\bar{X}_h - \bar{x}_h\right)\right]\left(\frac{\bar{X}_h + C_{xh}}{\bar{x}_h + C_{xh}}\right)$	1	C_{xh}
$\hat{\bar{Y}}_{\text{Reg}3} = \sum_{h=1}^{L} W_h\left[\bar{y}_h + b_h\left(\bar{X}_h - \bar{x}_h\right)\right]\left(\frac{\bar{X}_h + \beta_{2h}(x)}{\bar{x}_h + \beta_{2h}(x)}\right)$	1	$\beta_{2h}(x)$
$\hat{\bar{Y}}_{\text{Reg}4} = \sum_{h=1}^{L} W_h\left[\bar{y}_h + b_h\left(\bar{X}_h - \bar{x}_h\right)\right]\left(\frac{\beta_{2h}(x)\bar{X}_h + C_{xh}}{\beta_{2h}(x)\bar{x}_h + C_{xh}}\right)$	$\beta_{2h}(x)$	C_{xh}
$\hat{\bar{Y}}_{\text{Reg}5} = \sum_{h=1}^{L} W_h\left[\bar{y}_h + b_h\left(\bar{X}_h - \bar{x}_h\right)\right]\left(\frac{C_{xh}\bar{X}_h + \beta_{2h}(x)}{C_{xh}\bar{x}_h + \beta_{2h}(x)}\right)$	C_{xh}	$\beta_{2h}(x)$

4 Theoretical Comparison

The effectiveness of the suggested estimator is evaluated in comparison to the currently employed estimators; the usual separate ratio, and Tailor and Lone estimators. The MSE is used as a criterion in comparison with the new estimator in (11) with the existing ones in the general form in (8) as follows. The suggested $\hat{\bar{Y}}_{\text{Reg}}$ exhibits superior performance to the existing $\hat{\bar{Y}}_{\text{R}}$ if:

$$MSE\left(\hat{\bar{Y}}_{\text{Reg}}\right) < MSE\left(\hat{\bar{Y}}_{\text{R}}\right)$$

$$\sum_{h=1}^{L} W_h^2\gamma_h\bar{Y}_h^2 \begin{pmatrix} C_{yh}^2 + \left(\theta_h + \beta_h K_h\right)^2 C_{xh}^2 \\ -2\left(\theta_h + \beta_h K_h\right)\rho_h C_{xh} C_{yh} \end{pmatrix} < \sum_{h=1}^{L} W_h^2\gamma_h\bar{Y}_h^2 \left(C_{yh}^2 + \theta_h^2 C_{xh}^2 - 2\theta_h\rho_h C_{xh} C_{yh} \right)$$

$$\sum_{h=1}^{L} W_h^2\gamma_h\bar{Y}_h^2 \begin{pmatrix} \left(\theta_h + \beta_h K_h\right)^2 C_{xh}^2 \\ -2\left(\theta_h + \beta_h K_h\right)\rho_h C_{xh} C_{yh} \end{pmatrix} < \sum_{h=1}^{L} W_h^2\gamma_h\bar{Y}_h^2 \left(\theta_h^2 C_{xh}^2 - 2\theta_h\rho_h C_{xh} C_{yh} \right)$$

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

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

5 A Simulation Study

The simulation studies illustrate and compare the effectiveness of the suggested estimators against those already in existence. A population of size $N = 2{,}000$ is divided into 3 strata and variable (X,Y) are generated via bivariate normal distribution for each stratum. The strata parameters are:

1stYstratum:

$N_1 = 1{,}000, \bar{X}_1 = 400,\ \bar{Y}_1 = 500,\ C_{x1} = 0.3,\ C_{y1} = 1.4,\ \rho_1 = 0.8$

2nd stratum:

$N_2 = 600, \bar{X}_2 = 550,\ \bar{Y}_2 = 700,\ C_{x2} = 1.0,\ C_{y2} = 1.6,\ \rho_2 = 0.6$

3rd stratum:

$N_3 = 400,\ \bar{X}_3 = 550,\ \bar{Y}_3 = 350,\ C_{x3} = 1.1,\ C_{y3} = 0.9,\ \rho_3 = 0.4$

A sample of sizes $n = 100,\ 200,\ 400,$ and 600 are allocated from the population $N = 2,000$ using SRSWOR. We allocate the sample sizes *n* to each stratum using proportional allocation. For $n = 100$, the allocated sample sizes for the 1st, 2nd, and 3rd strata are $n_1 = 50,\ n_2 = 30,\ n_3 = 20$, for $n = 200$, the allocated sample sizes for the 1st, 2nd and 3rd strata are $n_1 = 100,\ n_2 = 60,\ n_3 = 40$, for $n = 400$ the allocated sample sizes for the 1st, 2nd and 3rd strata are $n_1 = 200,\ n_2 = 120,\ n_3 = 80$ and for $n = 600$ the allocated sample sizes for the 1st, 2nd and 3rd stratums are $n_1 = 300,\ n_2 = 180,\ n_3 = 120$, respectively. R program, [19], is applied to repeat the simulation 10,000 times.

The new and existing biases and MSEs are calculated by:

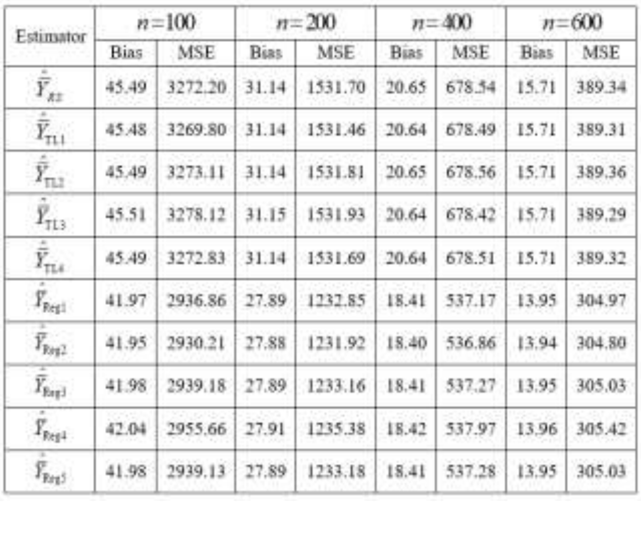
$$Bias\left(\hat{\bar{Y}}\right) = \frac{1}{10,000}\sum_{i=1}^{10,000}\left|\hat{\bar{Y}}_i - \bar{Y}\right|,$$

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

The biases and MSEs are represented in Table 2.

The results in Table 2 showed that the introduced estimators worked well because they gave both less bias and MSE than the existing ones. All proposed calculations assisting with unique known parameters of the auxiliary variable gave similar results for both biases and MSEs. The larger sample sizes gave smaller biases and MSEs concerning smaller sample sizes. We can see that the introduced estimators utilizing the sample regression estimator are more effective than the usual separate ratio and [14], ones under stratified random sampling.

Table 2. Biases and MSEs of the estimators



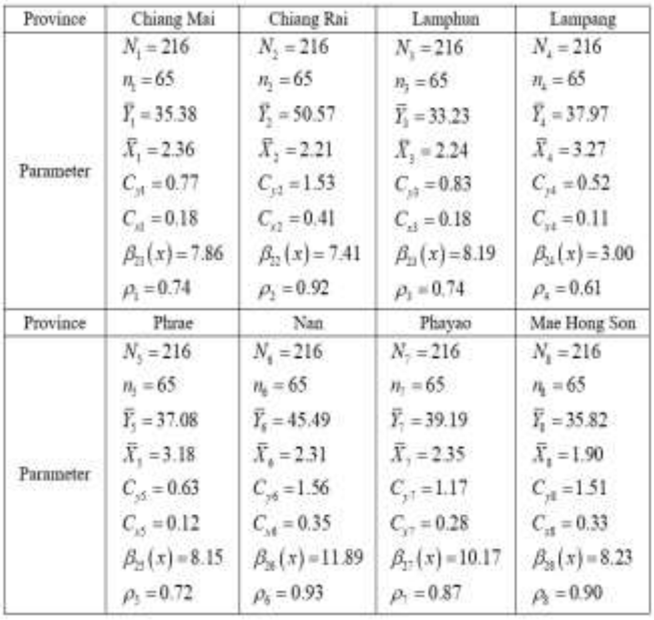
Estimator	$n=100$		$n=200$		$n=400$		$n=600$	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\hat{\bar{Y}}_{RS}$	45.49	3272.20	31.14	1531.70	20.65	678.54	15.71	389.34
$\hat{\bar{Y}}_{TL1}$	45.48	3269.80	31.14	1531.46	20.64	678.49	15.71	389.31
$\hat{\bar{Y}}_{TL2}$	45.49	3273.11	31.14	1531.81	20.65	678.56	15.71	389.36
$\hat{\bar{Y}}_{TL3}$	45.51	3278.12	31.15	1531.93	20.64	678.42	15.71	389.29
$\hat{\bar{Y}}_{TL4}$	45.49	3272.83	31.14	1531.69	20.64	678.51	15.71	389.32
$\hat{\bar{Y}}_{Reg1}$	41.97	2936.86	27.89	1232.85	18.41	537.17	13.95	304.97
$\hat{\bar{Y}}_{Reg2}$	41.95	2930.21	27.88	1231.92	18.40	536.86	13.94	304.80
$\hat{\bar{Y}}_{Reg3}$	41.98	2939.18	27.89	1233.16	18.41	537.27	13.95	305.03
$\hat{\bar{Y}}_{Reg4}$	42.04	2955.66	27.91	1235.38	18.42	537.97	13.96	305.42
$\hat{\bar{Y}}_{Reg5}$	41.98	2939.13	27.89	1233.18	18.41	537.28	13.95	305.03

6 An Application to Air Pollution Data in Northern Thailand

Northern Thailand air pollution data from 2003-2020 are applied to assess the effectiveness of the newest estimators, [20]. The monthly average density of nitrogen dioxide NO$_2$ (mg per square metre) is to be the auxiliary variable and the concentration of PM2.5 (micron per cubic metre) is considered as the study variable with a population $N = 1,728$ units. The population parameters are summarized below.

$\bar{Y} = 39.34,\ \bar{X} = 2.48,\ C_y = 1.23,\ C_x = 0.30,\ \beta_2(x) = 3.36,$ $\rho = 0.68.$ We considered eight provinces in northern Thailand as the strata to study the average amount of PM2.5. The provinces that are included in the study are Chiang Mai, Chiang Rai, Lamphun, Lampang, Phrae, Phayao, Nan, and Mae Hong Son ($N_h = 216,\ h = 1, 2, ..., 8$). A sample $n = 520$ is randomly picked from a population $N = 1,728$ using SRSWOR. The samples of sizes $n_h = 65$ based on proportional allocation are randomly taken from each strata using SRSWOR. The parameters in each stratum are showcased in Table 3.

Table 3. Population parameters for each province in northern Thailand

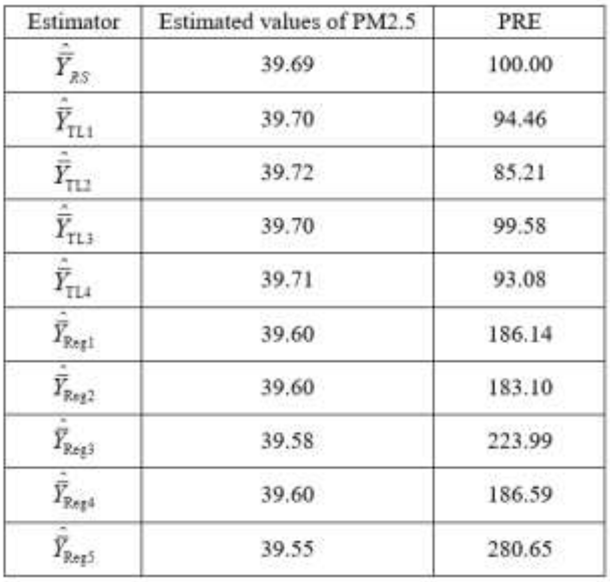


Province	Chiang Mai	Chiang Rai	Lamphun	Lampang
Parameter	$N_1 = 216$ $n_1 = 65$ $\bar{Y}_1 = 35.38$ $\bar{X}_1 = 2.36$ $C_{y1} = 0.77$ $C_{x1} = 0.18$ $\beta_{21}(x) = 7.86$ $\rho_1 = 0.74$	$N_2 = 216$ $n_2 = 65$ $\bar{Y}_2 = 50.57$ $\bar{X}_2 = 2.21$ $C_{y2} = 1.53$ $C_{x2} = 0.41$ $\beta_{22}(x) = 7.41$ $\rho_2 = 0.92$	$N_3 = 216$ $n_3 = 65$ $\bar{Y}_3 = 33.23$ $\bar{X}_3 = 2.24$ $C_{y3} = 0.83$ $C_{x3} = 0.18$ $\beta_{23}(x) = 8.19$ $\rho_3 = 0.74$	$N_4 = 216$ $n_4 = 65$ $\bar{Y}_4 = 37.97$ $\bar{X}_4 = 3.27$ $C_{y4} = 0.52$ $C_{x4} = 0.11$ $\beta_{24}(x) = 3.00$ $\rho_4 = 0.61$
Province	Phrae	Nan	Phayao	Mae Hong Son
Parameter	$N_5 = 216$ $n_5 = 65$ $\bar{Y}_5 = 37.08$ $\bar{X}_5 = 3.18$ $C_{y5} = 0.63$ $C_{x5} = 0.12$ $\beta_{25}(x) = 8.15$ $\rho_5 = 0.72$	$N_6 = 216$ $n_6 = 65$ $\bar{Y}_6 = 45.49$ $\bar{X}_6 = 2.31$ $C_{y6} = 1.56$ $C_{x6} = 0.35$ $\beta_{26}(x) = 11.89$ $\rho_6 = 0.93$	$N_7 = 216$ $n_7 = 65$ $\bar{Y}_7 = 39.19$ $\bar{X}_7 = 2.35$ $C_{y7} = 1.17$ $C_{x7} = 0.28$ $\beta_{27}(x) = 10.17$ $\rho_7 = 0.87$	$N_8 = 216$ $n_8 = 65$ $\bar{Y}_8 = 35.82$ $\bar{X}_8 = 1.90$ $C_{y8} = 1.51$ $C_{x8} = 0.33$ $\beta_{28}(x) = 8.23$ $\rho_8 = 0.90$

The estimated PM2.5 and percentage relative efficiencies (PREs) of the estimators against the usual separate ratio estimator are depicted in Table 4.

The new group of estimators gave a better performance which produced higher PREs compared to all existing estimators. The proposed estimator $\hat{\bar{Y}}_{\text{Reg5}}$ that used the benefit of both C_{xh} and $\beta_{2h}\left(x\right)$ gave the highest PRE in this situation and gave a closer estimated PM2.5 to the population. The results revealed that the recommended estimators achieved much more than the existing ones based on the air pollution dataset.

Table 4. Estimated values of PM2.5 and PREs of the estimators



Estimator	Estimated values of PM2.5	PRE
$\hat{\bar{Y}}_{RS}$	39.69	100.00
$\hat{\bar{Y}}_{\text{TL1}}$	39.70	94.46
$\hat{\bar{Y}}_{\text{TL2}}$	39.72	85.21
$\hat{\bar{Y}}_{\text{TL3}}$	39.70	99.58
$\hat{\bar{Y}}_{\text{TL4}}$	39.71	93.08
$\hat{\bar{Y}}_{\text{Reg1}}$	39.60	186.14
$\hat{\bar{Y}}_{\text{Reg2}}$	39.60	183.10
$\hat{\bar{Y}}_{\text{Reg3}}$	39.58	223.99
$\hat{\bar{Y}}_{\text{Reg4}}$	39.60	186.59
$\hat{\bar{Y}}_{\text{Reg5}}$	39.55	280.65

7 Conclusion

A new class of separate ratio estimators for predicting population mean are investigated through this study under stratified random sampling. Some available insights on the auxiliary variable have been implemented to increase the efficiency of the population mean estimator. The outcomes of the simulation study and the application to air pollution data in northern Thailand indicated that the suggested estimators performed more effectively than the typical separate ratio approach and [14] estimators under stratified random sampling. As expected, larger sample sizes resulted in smaller MSEs for all situations. The top-performing estimator outperformed all other existing estimators, delivering nearly thrice the efficiency. In subsequent research, additional established auxiliary variables may be employed to predict the population means of the variable under study, and the new estimators can be formulated to suit more intricate survey frameworks. e.g., cluster sampling and stratified single-stage cluster sampling. Nevertheless, the proposed estimators can help estimate other application data in many areas of interest.

Acknowledgement:
We appreciate all unknown referees for their valuable comments.

References:

[1] Lawson, N., Improved ratio type estimators using some prior information in sample surveys: a case study of fine particulate matter in Thailand, WSEAS Transactions on Systems, Vol. 22, 2023. pp. 538-542.

[2] Thongsak, N. and Lawson, N., A combined family of dual to ratio estimators using a transformed auxiliary variable, Lobachevskii Journal of Mathematics, Vol.43, No.9, 2022, pp. 2621-2633.

[3] Fadzil, A., Shuhaili, A., Ihsan, S. Z. and Faris, W. F., Air pollution study of vehicles emission in high volume traffic: Selangor, Malaysia as a case study, WSEAS Transactions on Systems, Vol. 12, 2013. pp. 67-84.

[4] Austin, W., Carattini, S, Gomez-Mahecha, J. and Pesko, M.F., The effects of contemporaneous air pollution on COVID-19 morbidity and mortality, Journal of Environmental Economics and Management, Vol. 119, 2023, pp. 102815.

[5] Beloconi, A. and Vounatsou, P, Long-term air pollution exposure and COVID-19 case-severity: An analysis of individual-level data from Switzerland. Environmental Research, Vol. 216, 2023, pp. 114481.

[6] Cochran, W.G. *Sampling Techniques*. 3rd edition. India: Wiley Eastern Limited, 1940.

[7] Sisodia, B. V. S., and Dwivedi ,V. K., A modified ratio estimator using coefficient of variation of auxiliary variable, *Journal of the Indian Society of Agricultural Statistics*, Vol.33, No.1, 1981, pp.13-18.

[8] Upadhyaya, L. N., and Singh, H. P., Use of transformed auxiliary variable in estimating the finite Population Mean, *Biometrical Journal*, Vol.41, No.5, 1999, pp. 627-636.

[9] Singh, H. P., Tailor, R., Tailor, R., and Kakran, M. S., An improved estimator of population mean using power transformation, *Journal of the Indian Society of Agricultural Statistics*, Vol. 58, No. 2, 2004, pp. 223-230.

[10] Kadilar, C. and Cingi, H., Ratio estimators in simple random sampling, *Applied Mathematics and Computation*, 2004, Vol. 151, pp. 893-902.

[11] Kadilar, C. and Cingi, H., An improvement in estimating the population mean by using the correlation coefficient, *Hacettepe Journal of Mathematics and Statistics*, 2006, Vol. 35, pp. 103 -109.

[12] Nangsue, N., Adjusted ratio and regression type estimators for estimation of population mean when some observations are missing, *International Scholarly and Scientific Research & Innovation*, Vo.3, No.5, 2009, pp. 334-337.

[13] Koç, T. and Koç, H. A new class of quantileregression ratio-type estimators for finite population mean in stratified random sampling, *Axioms*, Vol. 12, No. 7, 2023, pp. 713.

[14] Tailor, R. and Lone, H. A., Separate ratio-type estimators of population mean in stratified random sampling, *Journal of Modern Applied Statistical Methods*, 2014, Vol. 13, No.1, pp.223-233.

[15] Kadilar, C. and Cingi, H., Ratio estimators in stratified random sampling, *Journal of Modern Applied Statistical Methods*, 2003, Vol.45, No.2, pp. 218-225.

[16] Kadilar, C. and Cingi, H., A new ratio estimator in stratified random sampling, *Communications in Statistics-Theory and Methods*, 2005, Vol. 34, No.3, pp. 597-602.

[17] Singh, R.V.K, and Ahmed, A., Ratio-type estimators in stratified random sampling using auxiliary attribute, *Proceedings of the International MultiConference of Engineers and Computer Scientists* 2014 Vol I, IMECS 2014, March 12 - 14, 2014, Hong Kong.

[18] Sharma, V. and Kumar, S., Simulation study of ratio type estimators in stratified randomsampling using multi-auxiliary information, *Thailand Statistician*, 2020, Vol. 18, No. 3, pp.281-289.

[19] R Core Team, R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria, 2021, [Online], https://www.R-project.org/ (Accessed Date: November 5, 2023)

[20] Inness, A., Ades, M., Agustí-Panareda, A., Barré, J., Benedictow, A., Blechschmidt, A.-M., Dominguez, J. J., Engelen, R., Eskes, H., Flemming, J., Huijnen, V., Jones, L., Kipling, Z., Massart, S., Parrington, M., Peuch, V.-H., Razinger, M., Remy, S., Schulz, M., and Suttie, M. The CAMS Reanalysis of Atmospheric Composition. *Atmospheric Chemistry and Physics*, 2019, Vol. 19. No. 6, pp. 3515-3556.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of funding for research presented in a scientific article or scientific article itself

This work was funded by King Mongkut's University of Technology North Bangkok, Thailand. Contract number 671159.

Conflict of Interest

The author has no conflicts of interest to declare.

Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)