Songklanakarin J. Sci. Technol.
45 (2), 165–172, Mar. – Apr. 2023

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

Original Article

An improved family of estimators for estimating population mean using a transformed auxiliary variable under double sampling

Nuanpan Lawson*

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

Received: 30 November 2022; Revised: 10 January 2023; Accepted: 30 January 2023

Abstract

The performance of a population mean estimator can be improved by transformation techniques using a subset of the population data. An improved family of estimators for population mean has been proposed under double sampling using a transformed auxiliary variable. The biases and mean square errors of the proposed family of estimators up to the first order of approximation have been investigated. Simulation studies and an application to fine particulate matter in Chiang Rai, Thailand, are used to study the efficiency of the proposed estimators. The results from an application to air pollution in Chiang Rai showed that the proposed estimators gave smaller biases, by at least a half of the existing ones, and gave at least four times less mean square error than the existing ones.

Keywords: transformed auxiliary variable, double sampling, population mean, mean square error, bias

1. Introduction

It is important to study how to improve the efficiency of the estimators using sample survey data in order to develop precise estimators of population parameters. There are many techniques to use for this purpose, including the transformation techniques. For example, Srivenkataramana (1980) suggested to transform an auxiliary variable *X* to increase the performance of the population mean estimator under simple random sampling without replacement (SRSWOR) (see for example Tailor and Sharma (2009), Onyeka *et al.* (2013). Later, Adewara *et al.* (2012) proposed to transform both the auxiliary variable and the study variable *Y* with the same purpose to increase the efficiency of the population mean estimator.

Moreover, many parameters of the auxiliary variable are usually unavailable. Thongsak and Lawson (2021) suggested general forms of ratio estimators using the transformation method to gain more efficiency for population mean estimator under SRSWOR. Later, Thongsak and Lawson (2022a) developed new estimators based on Thongsak and Lawson (2021), when the population mean of the auxiliary variable is not known under double sampling. Double sampling is a sampling method applying two stages in order to gain information on the unknown population mean of the auxiliary variable, which is usually unknown in practice from the first phase of sampling suggested by Neyman (1938). The Thongsak and Lawson (2022a) estimators are

$$\hat{\bar{Y}}_{\mathrm{R}} = \bar{y}\left(\frac{A\bar{x}^{*d} + D}{A\bar{x}' + D}\right)^{b}, \tag{1}$$

$$\hat{\bar{Y}}_{\mathrm{Reg}} = \left[\bar{y} + b\left(\bar{x}' - \bar{x}^{*d}\right)\right]\left(\frac{G\bar{x}^{*d} + H}{G\bar{x}' + H}\right)^{b}, \tag{2}$$

where $\bar{x}^{*d} = \dfrac{n'\bar{x}' - n\bar{x}}{n' - n} = (1+\pi)\bar{x}' - \pi\bar{x}, \quad \pi = \dfrac{n}{n' - n}$,

$\bar{x}'$ and $\bar{x}$ are the sample means of the auxiliary variable based on the first phase sample of size n' and the second phase sample of size n, $\bar{y}$ is the sample mean of the study variable

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

and b is a sample regression coefficient. A, D, G, H are constants or some known parameters such as the coefficient of variation or the correlation coefficient.

The biases and mean square errors (MSEs) of $\hat{\hat{\bar{Y}}}_{\mathrm{R}}$ and $\hat{\hat{\bar{Y}}}_{\mathrm{Reg}}$ are respectively

$$Bias\left(\hat{\hat{\bar{Y}}}_{\mathrm{R}}\right) \cong \bar{Y}\left(\gamma-\gamma^{*}\right)\left[\frac{\beta(\beta-1)}{2}\pi^{2}\theta_{1}^{2}C_{x}^{2}-\beta\pi\theta_{1}\rho C_{x}C_{y}\right], \tag{3}$$

$$Bias\left(\hat{\hat{\bar{Y}}}_{\mathrm{Reg}}\right) \cong \bar{Y}\left(\gamma-\gamma^{*}\right)\left[\left(\frac{\beta(\beta-1)}{2}\theta_{2}^{2}-\beta\beta K\theta_{2}\right)\pi^{2}C_{x}^{2}-\beta\pi\theta_{2}\rho C_{x}C_{y}\right], \tag{4}$$

$$MSE\left(\hat{\hat{\bar{Y}}}_{\mathrm{R}}\right) \cong \bar{Y}^{2}\left[\gamma C_{y}^{2}+\left(\gamma-\gamma^{*}\right)\left(\beta^{2}\theta_{1}^{2}\pi^{2}C_{x}^{2}-2\beta\theta_{1}\pi\rho C_{x}C_{y}\right)\right], \tag{5}$$

$$MSE\left(\hat{\hat{\bar{Y}}}_{\mathrm{Reg}}\right) \cong \bar{Y}^{2}\left[\gamma C_{y}^{2}+\left(\gamma-\gamma^{*}\right)\left\{\left(\beta\theta_{2}-\beta K\right)^{2}\pi^{2}C_{x}^{2}-2\left(\beta\theta_{2}-\beta K\right)\pi\rho C_{x}C_{y}\right\}\right], \tag{6}$$

where $\theta_1=\frac{A\bar{X}}{A\bar{X}+D}$, $\theta_2=\frac{G\bar{X}}{G\bar{X}+H}$, $K=\frac{\bar{X}}{\bar{Y}}$, $\beta=\frac{\rho S_y}{S_x}$, $\gamma=\frac{1}{n}-\frac{1}{N}$, $\gamma^{*}=\frac{1}{n'}-\frac{1}{N}$, ρ is the correlation coefficient between X and Y, and C_x, C_y are the coefficients of variation of X and Y, respectively.

Members of $\hat{\hat{\bar{Y}}}_{\mathrm{R1}}$ and $\hat{\hat{\bar{Y}}}_{\mathrm{Reg1}}$ with some auxiliary parameters are in Table 1.

Table 1. Some estimators proposed by Thongsak and Lawson (2022a).

Estimator		A or G	D or H
$\hat{\hat{\bar{Y}}}_{\mathrm{R1}}=\bar{y}\left(\frac{\bar{x}^{*d}}{\bar{x}'}\right)^{b}$	$\hat{\hat{\bar{Y}}}_{\mathrm{Reg1}}=\left[\bar{y}+b\left(\bar{x}'-\bar{x}^{*d}\right)\right]\left(\frac{\bar{x}^{*d}}{\bar{x}'}\right)^{b}$	1	0
$\hat{\hat{\bar{Y}}}_{\mathrm{R2}}=\bar{y}\left(\frac{\bar{x}^{*d}+C_x}{\bar{x}'+C_x}\right)^{b}$	$\hat{\hat{\bar{Y}}}_{\mathrm{Reg2}}=\left[\bar{y}+b\left(\bar{x}'-\bar{x}^{*d}\right)\right]\left(\frac{\bar{x}^{*d}+C_x}{\bar{x}'+C_x}\right)^{b}$	1	C_x
$\hat{\hat{\bar{Y}}}_{\mathrm{R3}}=\bar{y}\left(\frac{\bar{x}^{*d}+\rho}{\bar{x}'+\rho}\right)^{b}$	$\hat{\hat{\bar{Y}}}_{\mathrm{Reg3}}=\left[\bar{y}+b\left(\bar{x}'-\bar{x}^{*d}\right)\right]\left(\frac{\bar{x}^{*d}+\rho}{\bar{x}'+\rho}\right)^{b}$	1	ρ

Recently, Thongsak and Lawson (2022b) recommended the class of ratio estimators taking benefit of the known parameters of the auxiliary variable and benefit of a transformed auxiliary variable under SRSWOR, developed from the non-transformed estimators proposed by Jaroengeratikun and Lawson (2018). The results found in the study of Thongsak and Lawson (2022b) showed that using the transformation technique can gain increased efficiency of the estimators over non-transformed estimators (see for example Thongsak and Lawson, 2022c).

In this study, an improved family of estimators for population mean is proposed using a transformed auxiliary variable to improve the efficiency of the estimators under double sampling. The Taylor series approximation is considered to investigate the biases and mean square errors of the proposed estimators. Simulation studies and an empirical study on fine particulate matter 2.5 data for Chiang Rai, Thailand, are performed to examine the performances of the proposed estimators.

2. Materials and Methods

Motivated by Thongsak and Lawson (2022a), we suggest improving the population mean estimator by combining the estimators $\hat{\hat{\bar{Y}}}_{\mathrm{R}}$ in equation (1) and $\hat{\hat{\bar{Y}}}_{\mathrm{Reg}}$ in equation (2) with a constant α that minimizes the mean square error of the proposed estimator. A combined family of estimators is suggested based on double sampling when the auxiliary variable is not available.

The proposed estimator is

$$\hat{\hat{\bar{Y}}}_{N}=\alpha\bar{y}\left(\frac{A\bar{x}^{*d}+D}{A\bar{x}'+D}\right)^{b}+(1-\alpha)\left[\bar{y}+b\left(\bar{x}'-\bar{x}^{*d}\right)\right]\left(\frac{G\bar{x}^{*d}+H}{G\bar{x}'+H}\right)^{b}, \tag{7}$$

where α is a constant that minimizes MSE of the proposed estimator, $\bar{x}^{*d}=\dfrac{n'\bar{x}'-n\bar{x}}{n'-n}=\left(1+\pi\right)\bar{x}'-\pi\bar{x},$ $\bar{x}'$ and $\bar{x}$ are the sample means of the auxiliary variable based on the first phase sample of size n' and second phase sample of size *n*, $\bar{y}$ is the sample mean of the study variable and *b* is a sample regression coefficient, and *A*, *D*, *G*, *H* are constants or some known parameters such as the coefficient of variation (C_x) or the correlation coefficient (ρ).

To obtain the biases and MSEs of the proposed estimators, the following notation is employed: $\varepsilon_0=\left(\bar{y}-\bar{Y}\right)/\bar{Y}$ then $\bar{y}=\left(1+\varepsilon_0\right)\bar{Y}$, $\varepsilon_1=\left(\bar{x}-\bar{X}\right)/\bar{X}$ then $\bar{x}=\left(1+\varepsilon_1\right)\bar{X}$, $\varepsilon_2=\left(\bar{x}'-\bar{X}\right)/\bar{X}$ then $\bar{x}'=\left(1+\varepsilon_2\right)\bar{X}$ and $\bar{x}^{*d}=\left(1+\varepsilon_2+\pi\varepsilon_2-\pi\varepsilon_1\right)\bar{X}$, $\varepsilon_3=\left(\bar{y}'-\bar{Y}\right)/\bar{Y}$ then $\bar{y}'=\left(1+\varepsilon_3\right)\bar{Y}$ and $\bar{y}^{*d}=\left(1+\varepsilon_3+\pi\varepsilon_3-\pi\varepsilon_0\right)\bar{Y}$ such that $E\left(\varepsilon_0\right)=E\left(\varepsilon_1\right)=E\left(\varepsilon_2\right)=E\left(\varepsilon_3\right)=0$, $E\left(\varepsilon_0^2\right)=\gamma C_y^2$, $E\left(\varepsilon_1^2\right)=\gamma C_x^2$, $E\left(\varepsilon_2^2\right)=E\left(\varepsilon_1\varepsilon_2\right)=\gamma^* C_x^2$, $E\left(\varepsilon_3^2\right)=E\left(\varepsilon_0\varepsilon_3\right)=\gamma^* C_y^2$, $E\left(\varepsilon_0\varepsilon_1\right)=\gamma\rho C_x C_y$, $E\left(\varepsilon_0\varepsilon_2\right)=E\left(\varepsilon_1\varepsilon_3\right)=E\left(\varepsilon_2\varepsilon_3\right)=\gamma^*\rho C_x C_y$.

Rewriting equation (7) in terms of $\varepsilon_0,\ \varepsilon_1,\varepsilon_2$ we have:

$$\hat{\hat{\bar{Y}}}_N=\alpha\left(1+\varepsilon_0\right)\bar{Y}\left(\frac{\left(A\bar{X}+D\right)+\left(\varepsilon_2+\pi\varepsilon_2-\pi\varepsilon_1\right)A\bar{X}}{\left(A\bar{X}+D\right)+\varepsilon_2 A\bar{X}}\right)^b +\left(1-\alpha\right)\left[\left(1+\varepsilon_0\right)\bar{Y}+b\left(\varepsilon_1-\varepsilon_2\right)\pi\bar{X}\right]\left(\frac{\left(G\bar{X}+H\right)+\left(\varepsilon_2+\pi\varepsilon_2-\pi\varepsilon_1\right)G\bar{X}}{\left(G\bar{X}+H\right)+\varepsilon_2 G\bar{X}}\right)^b \tag{8}$$

The biases and MSEs of $\hat{\hat{\bar{Y}}}_N$ up to the first degree of approximation are acquired

$$Bias\left(\hat{\hat{\bar{Y}}}_N\right)\cong\left(\gamma-\gamma^*\right)\bar{Y}\left[\begin{array}{l}\left(\dfrac{\beta\left(\beta-1\right)}{2}\left(\alpha\theta_1^2+\left(1-\alpha\right)\theta_2^2\right)-\left(1-\alpha\right)\beta^2K\theta_2\right)\pi^2C_X^2\\ -\left(\alpha\theta_1+\left(1-\alpha\right)\theta_2\right)\beta\pi\rho C_x C_y\end{array}\right], \tag{9}$$

$$MSE\left(\hat{\hat{\bar{Y}}}_N\right)\cong\bar{Y}^2\left[\gamma C_y^2+\left(\gamma-\gamma^*\right)\left(\begin{array}{l}\left[\alpha\theta_1+\left(\alpha-1\right)\left(K-\theta_2\right)\right]^2\pi^2\beta^2C_x^2\\ -2\left[\alpha\theta_1+\left(\alpha-1\right)\left(K-\theta_2\right)\right]\pi\beta\rho C_x C_y\end{array}\right)\right], \tag{10}$$

where the difference $E\left(b\right)-\beta$ was omitted (Cochran, 1977).

2.1 Optimum choice of scalar α

In order to find the minimum value of MSE in the proposed family of estimators $\hat{\hat{\bar{Y}}}_N$ in equation (7), we find the optimum value of α by taking a partial derivative of the MSE in equation (10) with respect to α and equating it to zero.

The optimum value of α for $\hat{\hat{\bar{Y}}}_N$ is

$$\alpha=\frac{\rho C_y+\left(K-\theta_2\right)\pi\beta C_x}{\left(\theta_1+K-\theta_2\right)\pi\beta C_x}=\alpha_{\mathrm{N}}^{\mathrm{opt}},\ \left(\text{say}\right). \tag{11}$$

Substituting equation (11) into equation (7), the optimum $\hat{\hat{\bar{Y}}}_N$ is

$$\hat{\hat{\bar{Y}}}_{\mathrm{N}}^{\mathrm{opt}}=\alpha_{\mathrm{N}}^{\mathrm{opt}}\hat{\bar{Y}}_{\mathrm{R}}+\left(1-\alpha_{\mathrm{N}}^{\mathrm{opt}}\right)\hat{\bar{Y}}_{\mathrm{Reg}}. \tag{12}$$

Substituting equation (11) into equation (10), the optimum MSE of estimator $\hat{\hat{\bar{Y}}}_{\mathrm{N}}^{\mathrm{opt}}$ is

$$MSE_{\min}\left(\hat{\hat{\bar{Y}}}_{\mathrm{N}}^{\mathrm{opt}}\right) \cong \bar{Y}^2\left(\gamma C_y^2 - \left(\gamma-\gamma^*\right)\rho^2 C_y^2\right). \tag{13}$$

Some proposed estimators are displayed in Table 2.

Table 2. Some proposed estimators

Estimator
$\hat{\hat{\bar{Y}}}_{\mathrm{N1}}^{\mathrm{opt}} = \alpha_{\mathrm{N1}}^{\mathrm{opt}}\hat{\hat{\bar{Y}}}_{\mathrm{R1}} + \left(1-\alpha_{\mathrm{N1}}^{\mathrm{opt}}\right)\hat{\hat{\bar{Y}}}_{\mathrm{Reg1}}$
$\hat{\hat{\bar{Y}}}_{\mathrm{N2}}^{\mathrm{opt}} = \alpha_{\mathrm{N2}}^{\mathrm{opt}}\hat{\hat{\bar{Y}}}_{\mathrm{R2}} + \left(1-\alpha_{\mathrm{N2}}^{\mathrm{opt}}\right)\hat{\hat{\bar{Y}}}_{\mathrm{Reg2}}$
$\hat{\hat{\bar{Y}}}_{\mathrm{N3}}^{\mathrm{opt}} = \alpha_{\mathrm{N3}}^{\mathrm{opt}}\hat{\hat{\bar{Y}}}_{\mathrm{R3}} + \left(1-\alpha_{\mathrm{N3}}^{\mathrm{opt}}\right)\hat{\hat{\bar{Y}}}_{\mathrm{Reg3}}$

2.2 Efficiency comparisons

The proposed estimators are compared with the usual ratio estimator ($\hat{\hat{\bar{Y}}}_{\mathrm{Neyman}}$) and Thongsak and Lawson's (2022a) estimators ($\hat{\hat{\bar{Y}}}_{\mathrm{R}}$ and $\hat{\hat{\bar{Y}}}_{\mathrm{Reg}}$) under the double sampling scheme, and the MSEs are used as a criterion.

1) $\hat{\hat{\bar{Y}}}_{\mathrm{N}}^{\mathrm{opt}}$ performs better than $\hat{\hat{\bar{Y}}}_{\mathrm{Neyman}}$, $\hat{\hat{\bar{Y}}}_{\mathrm{R}}$ and $\hat{\hat{\bar{Y}}}_{\mathrm{Reg}}$ if

$$\left(\omega C_x - \rho C_y\right)^2 > 0. \tag{14}$$

For $\omega = 1,$ $\hat{\hat{\bar{Y}}}_{\mathrm{N}}^{\mathrm{opt}}$ is more efficient than $\hat{\hat{\bar{Y}}}_{\mathrm{Neyman}}$.

For $\omega = \beta\theta_1\pi,$ $\hat{\hat{\bar{Y}}}_{\mathrm{N}}^{\mathrm{opt}}$ is more efficient than $\hat{\hat{\bar{Y}}}_{\mathrm{R}}$.

For $\omega = \beta\left(\theta_2 - K\right)\pi,$ $\hat{\hat{\bar{Y}}}_{\mathrm{N}}^{\mathrm{opt}}$ is more efficient than $\hat{\hat{\bar{Y}}}_{\mathrm{Reg}}$.

From equation (14) we can see that the proposed estimators always perform better than all existing estimators because the condition is always true.

3. Results and Discussion

3.1 Simulation results

The paired variables (*X, Y*) from the bivariate normal distribution were generated with the following parameters: $N = 1{,}500,\ \bar{Y} = 55,\ \bar{X} = 45,\ C_y = 0.6,\ C_x = 1.5,\ \rho = 0.3,\ 0.5,\ 0.8$.

In the first phase of sampling, the samples of sizes $n' = 150,\ 300,$ and 600 units were selected from N population units under SRSWOR scheme, then in the second phase of sampling, the sample of sizes $n = 45,\ 90,$ and 180 units were selected under the SRSWOR scheme from $n' = 150,\ 300,$ and 600, respectively. The simulation was repeated 10,000 times in the R program (R Core Team (2021)). The biases and MSEs of the proposed and existing estimators were calculated by

$$Bias\left(\hat{\hat{\bar{Y}}}\right) = \frac{1}{10{,}000}\sum_{i=1}^{10{,}000}\left|\hat{\hat{\bar{Y}}}_i - \bar{Y}\right|, \tag{15}$$

$$MSE\left(\hat{\hat{\bar{Y}}}\right) = \frac{1}{10{,}000}\sum_{i=1}^{10{,}000}\left(\hat{\hat{\bar{Y}}}_i - \bar{Y}\right)^2, \tag{16}$$

The biases and MSEs of the proposed and existing estimators are represented in Tables 3-5 and the mean square errors of all estimators are presented in Figure 1.

The results in Tables 3-5 are similar. We can see that all the proposed estimators using combined estimators gave both smaller biases and MSEs at all levels of sample sizes and correlation coefficients between the auxiliary and study variables. The larger is ρ , the higher the efficiencies are for the new combined estimators.

From Figure 1, the results indicate that the mean square errors of the proposed estimates are smaller than for all other existing estimators at different levels of ρ. A larger ρ gave smaller mean square errors than a smaller ρ, at all sample sizes.

Table 3. Biases and MSEs of the proposed estimators and existing estimators when ρ = 0.3

Estimator	n'=150, n=45		n'=300, n=90		n'=600, n=180	
	Bias	MSE	Bias	MSE	Bias	MSE
$\hat{\bar{Y}}_{Neyman}$	9.026	178.910	5.936	60.147	4.122	27.583
$\hat{\bar{Y}}_{R1}$	3.770	22.372	2.619	10.801	1.786	4.972
$\hat{\bar{Y}}_{R2}$	3.772	22.400	2.621	10.817	1.787	4.979
$\hat{\bar{Y}}_{R3}$	3.770	22.379	2.620	10.805	1.786	4.974
$\hat{\bar{Y}}_{Reg1}$	3.858	23.400	2.685	11.333	1.831	5.216
$\hat{\bar{Y}}_{Reg2}$	3.863	23.457	2.688	11.362	1.833	5.230
$\hat{\bar{Y}}_{Reg3}$	3.859	23.413	2.686	11.339	1.831	5.219
$\hat{\bar{Y}}_{N1}^{opt}$	3.749	22.149	2.591	10.597	1.770	4.889
$\hat{\bar{Y}}_{N2}^{opt}$	3.750	22.159	2.591	10.598	1.770	4.889
$\hat{\bar{Y}}_{N3}^{opt}$	3.749	22.151	2.591	10.597	1.770	4.889

Table 4. Biases and MSEs of the proposed estimators and existing estimators when ρ = 0.5

Estimator	n'=150, n=45		n'=300, n=90		n'=600, n=180	
	Bias	MSE	Bias	MSE	Bias	MSE
$\hat{\bar{Y}}_{Neyman}$	8.242	146.555	5.421	49.903	3.758	22.878
$\hat{\bar{Y}}_{R1}$	3.605	20.417	2.501	9.824	1.695	4.482
$\hat{\bar{Y}}_{R2}$	3.612	20.489	2.506	9.864	1.698	4.500
$\hat{\bar{Y}}_{R3}$	3.608	20.442	2.503	9.837	1.696	4.488
$\hat{\bar{Y}}_{Reg1}$	3.841	23.138	2.669	11.210	1.817	5.155
$\hat{\bar{Y}}_{Reg2}$	3.853	23.285	2.678	11.284	1.824	5.191
$\hat{\bar{Y}}_{Reg3}$	3.845	23.189	2.672	11.235	1.819	5.168
$\hat{\bar{Y}}_{N1}^{opt}$	3.529	19.649	2.430	9.290	1.645	4.235
$\hat{\bar{Y}}_{N2}^{opt}$	3.530	19.665	2.430	9.291	1.645	4.235
$\hat{\bar{Y}}_{N3}^{opt}$	3.530	19.656	2.430	9.291	1.645	4.235

Table 5. Biases and MSEs of the proposed estimators and existing estimators when ρ = 0.8

Estimator	n'=150, n=45		n'=300, n=90		n'=600, n=180	
	Bias	MSE	Bias	MSE	Bias	MSE
$\hat{\bar{Y}}_{Neyman}$	6.979	104.213	4.580	35.463	3.159	16.126
$\hat{\bar{Y}}_{R1}$	3.157	15.604	2.163	7.370	1.448	3.281
$\hat{\bar{Y}}_{R2}$	3.174	15.782	2.176	7.465	1.458	3.328
$\hat{\bar{Y}}_{R3}$	3.166	15.698	2.170	7.420	1.454	3.306
$\hat{\bar{Y}}_{Reg1}$	3.771	22.408	2.609	10.764	1.780	4.982
$\hat{\bar{Y}}_{Reg2}$	3.801	22.767	2.631	10.945	1.796	5.073
$\hat{\bar{Y}}_{Reg3}$	3.787	22.596	2.620	10.859	1.788	5.030
$\hat{\bar{Y}}_{N1}^{opt}$	2.892	13.301	1.967	6.054	1.294	2.633
$\hat{\bar{Y}}_{N2}^{opt}$	2.892	13.315	1.967	6.052	1.294	2.633
$\hat{\bar{Y}}_{N3}^{opt}$	2.893	13.312	1.967	6.054	1.294	2.633

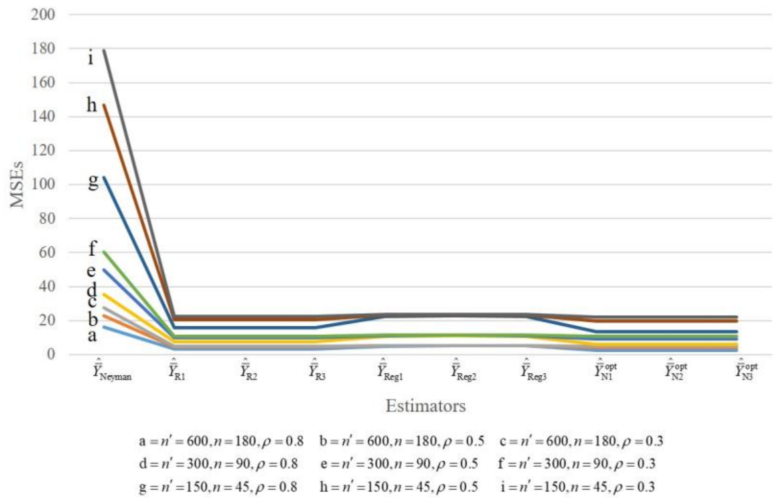

Figure 1. The mean square errors of the proposed and existing estimators at different levels of ρ.

3.2 Application to air pollution in Chiang Rai, Thailand

Air pollution is among the most important issues in Thailand nowadays, especially in northern Thailand. Therefore, we consider the air pollution data for Chiang Rai as an application case. The fine particulate matter 2.5, PM2.5 levels (μg/m^3) and nitrogen oxide, NO_2 levels (mg/m^2) on a time-scale of once per month (monthly averages) from the Copernicus Atmosphere Monitoring Service (CAMS), The European Centre for Medium-Range Weather Forecasts (ECMWF) in 2003-2020 [25] were used in the study. The data belong to a population of size 216 units. The concentration of NO_2 is considered as the study variable Y and the concentration of PM2.5 is considered as the auxiliary variable X. The population parameters are

$$N = 216,\ \bar{Y} = 2.214,\ \bar{X} = 50.570,\ C_y = 0.410,\ C_x = 1.531,\ \rho = 0.921.$$

In the first phase of sampling, a sample of size $n' = 75$ is selected from the population size n = 216 using the SRSWOR scheme. In the second phase of sampling a sample of size n = 20 is selected from $n' = 75$ using the SRSWOR scheme. The biases and MSEs of the proposed estimators and existing estimators are presented in Table 6.

Table 6. Biases and MSEs of the proposed estimators and existing estimators in an application of pollution data in Chiang Rai

Estimator	Bias	MSE
$\hat{\bar{Y}}_{\text{Neyman}}$	0.3108	0.0966
$\hat{\bar{Y}}_{\text{R1}}$	0.2920	0.0852
$\hat{\bar{Y}}_{\text{R2}}$	0.2921	0.0853
$\hat{\bar{Y}}_{\text{R3}}$	0.2920	0.0853
$\hat{\bar{Y}}_{\text{Reg1}}$	0.3582	0.1283
$\hat{\bar{Y}}_{\text{Reg2}}$	0.3584	0.1284
$\hat{\bar{Y}}_{\text{Reg3}}$	0.3583	0.1284
$\hat{\bar{Y}}_{\text{N1}}^{\text{opt}}$	0.1417	0.0201
$\hat{\bar{Y}}_{\text{N2}}^{\text{opt}}$	0.1417	0.0201
$\hat{\bar{Y}}_{\text{N3}}^{\text{opt}}$	0.1417	0.0201

Table 6 shows that the results from an application to fine particulate matter in Chiang Rai also support the results found in the simulation studies. The proposed combined estimators performed the best as the biases and mean square errors of the proposed estimators were smaller than those of all existing estimators. The biases were reduced by at least a half, and gave at least four times less mean square errors than the existing ones which can be considered a large improvement. The proposed estimators work well in this application with a high correlation between NO_2 and the concentration of PM2.5 (ρ = 0.921) in Chiang Rai, Thailand, and again the case study supports what we found in the simulation studies.

4. Conclusions

A family of ratio estimators for population mean have been proposed in the case that the population mean of an auxiliary variable is unknown. Double sampling scheme is considered under this situation to estimate the unknown population mean of the auxiliary variable. The biases and mean square errors of the improved estimators are displayed. The simulation results and an application to fine particulate matter data in Chiang Rai, Thailand, support the finding that the improved estimators gave the lowest biases and mean square errors for all levels of sample sizes and all levels of correlation coefficients between the auxiliary variable and study variable. The proposed estimators always perform the best and they can be useful for estimating the population mean or population total of the variable of interest when the researchers have no information on the auxiliary variable. Therefore, this approach can be applied in real world problems which leads to more powerful estimation.

Acknowledgements

This research was funded by Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Thailand, contract no 661148. Thank you to all the reviewers for their comments for improving this paper.

References

Adewara, A. A., Singh, R., & Kumar, M. (2012). Efficiency of some modified ratio and product estimators using known value of some population parameters. *International Journal of Applied Science and Technology*, *2*(2), 76-79.

Jaroengeratikun, U., & Lawson, N. (2018). 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. Retrieved from https://doi.org/10.5614/j.math.fund.sci.2019.51.1.1

Neyman J. (1938). Contribution to the theory of sampling human populations. *Journal of the American Statistical Association*, *33*, 101-116. Retrieved from https://doi.org/10.2307/2279117

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*, *13*, 57-65. Retrieved from https://doi.org/10.17654/TS057020103

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

Srivenkataramana, T. (1980). A dual to ratio estimator in sample surveys. *Biometrika*, *67*(1), 199-204. Retrieved from https://doi.org/10.2307/2335334

Tailor, R., & Sharma, B. (2009). A modified ratio-cum-product estimator of finite population mean using known coefficient of variation and coefficient of kurtosis. *Statistics in Transition-New Series*, *10*(1), 15-24.

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*, Bangkok, Thailand, 1-2 September 2021. Retrieved from https://doi.org/10.1109/RI2C51727.2021.9559798

Thongsak, N., & Lawson, N. (2022a). Classes of population mean estimators using transformed variables in double sampling. *Gazi Journal of Science and Technology*.

Thongsak, N. & Lawson, N. (2022b). A combined family of dual to ratio estimators using a transformed auxiliary variable. *Lobachevskii Journal of Mathematics, 43*(9), 2621-2633. Retrieved from https://doi.org/10.1134/S1995080222120253

Thongsak, N., & Lawson, N. (2022c). 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.* Retrieved from https://doi.org/10.1080/08898480.2022.2145790