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

Nuanpan Lawson[1*]

[1]King Mongkut's University of Technology North Bangkok, Faculty of Applied Science, 10800 Bangkok, Thailand

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

Abstract

The performance of population mean estimators can be improved by transformation techniques using 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. The 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 show that the proposed estimators gave smaller biases, at least half of the existing ones, and gave at least four times less mean square errors 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 to estimate 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 e.g., 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 in the same purpose in increasing the efficiency of the population mean estimator.

Moreover, many parameters of the auxiliary variable are usually unavailable. Thongsak and Lawson (2021) suggested the general forms for 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 to use under 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

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

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

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

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

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

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

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

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

where $\theta_{1} = \dfrac{A\bar{X}}{A\bar{X} + D}$, $\theta_{2} = \dfrac{G\bar{X}}{G\bar{X} + H}$, $K = \dfrac{\bar{X}}{\bar{Y}}$, $\beta = \dfrac{\rho S_{y}}{S_{x}}$, $\gamma = \dfrac{1}{n} - \dfrac{1}{N}$, $\gamma^{*} = \dfrac{1}{n'} - \dfrac{1}{N}$, ρ is the correlation coefficient between X and Y, C_{x}, C_{y} are the coefficients of variation of X and Y, respectively.

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

Recently, Thongsak and Lawson (2022b) recommended the class of ratio estimators using the benefit of the known parameters of the auxiliary variable and the benefit of transformed auxiliary variable under SRSWOR which developed from the non-transformed estimators proposed by Jaroengeratikun and Lawson (2018). The results found from the study of Thongsak and Lawson (2022b) showed that using the transformation technique can be gained in increasing the efficiency of the estimators over non-transformed estimators (see e.g., Thongsak and Lawson, 2022c).

In this study, an improved family of estimators for population mean has been 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. The simulation studies and an empirical study on the fine particulate matter 2.5 data in Chiang Rai, Thailand are applied to examine the performance of the proposed estimators.

2. Materials and Methods

Motivated by Thongsak and Lawson (2022a), we suggested to improve the population mean estimator by combining the estimators $\hat{\bar{Y}}_{\mathrm{R}}$ in equation (1) and $\hat{\bar{Y}}_{\mathrm{Reg}}$ 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

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

where α is a constant that minimizes MSE of the proposed estimator, $\bar{x}^{*d}=\dfrac{n'\bar{x}'-n\bar{x}}{n'-n}=(1+\pi)\bar{x}'-\pi\bar{x}$, $\bar{x}'$ 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) and the correlation coefficient (ρ).

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

Rewriting equation (7) in terms of ε_0, $\varepsilon_1, \varepsilon_2$ we have:

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

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

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

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

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

2.1 Optimum Choice of Scalar α

In order to find the minimum value of MSE of the proposed family of estimators $\hat{\hat{Y}}_N$ 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.

Optimum value of α for $\hat{\hat{Y}}_N$ is

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

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

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

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

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

Some proposed estimators are displayed in Table 2.

2.2 Efficiency Comparisons

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

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

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

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

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

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

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 are 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 are 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 are selected under the SRSWOR scheme from $n' = 150,\ 300,$ and 600, respectively. The simulation is repeated 10,000 times by the R program (R Core Team (2021)). The biases and MSEs of the proposed and existing estimators are calculated by

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

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

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 represented in Figure 1.

The results from tables 3-5 are similar. We can see that all 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 ore higher ρ, the higher efficiencies of the new combined estimators.

From figure 1, the results found that the mean square errors of the proposed are smaller than all other existing estimators at different levels of ρ. A larger ρ gave smaller mean square errors compared to a smaller ρ at all sample sizes.

3.2 Application to Air Pollution in Chiang Rai, Thailand

The air pollution is one of the most important issues in Thailand nowadays especially in the north of Thailand therefore, we consider the air pollution data in Chiang Rai in the application. The fine particulate matter 2.5, $\mathrm{PM2.5}$ levels ($\mu\mathrm{g/m^3}$) and nitrogen oxide, $\mathrm{NO_2}$ levels ($\mathrm{mg/m^2}$) on a time-scale of one per month (monthly average) from the Copernicus Atmosphere Monitoring Service (CAMS), The European Centre for Medium-Range Weather Forecasts (ECMWF) in 2003-2020 [25] are used in the study. The data belongs to a population of size 216 units. The concentration of $\mathrm{NO_2}$ is considered as the study variable Y and the concentration of $\mathrm{PM2.5}$ 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 showed the results for an application to fine particulate matter in Chiang Rai also supported the results found in the simulation studies. The proposed combined estimators performed the best as you can see the biases and mean square errors of the proposed estimators were smaller than all existing estimators. The biases were at least half, and gave at least four times less mean square errors than the existing ones which can be considered high improvement. The proposed estimators work well in this application with a high correlation between NO_2 and the concentration of PM2.5 ($\rho = 0.921$) in Chiang Rai, Thailand and again, it 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 supports 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 and therefore can be applied in the real world problems which leads to more powerful estimation.

Acknowledgments

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. URL 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. and 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. In 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, *Accepted for publication in 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

Table

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

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

Table 2 Some proposed estimators.

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

Table 3 Biases and MSEs of the proposed estimators and existing estimators when $\rho = 0.3$.

Estimator	n'=150, n=45		n'=300, n=90		n'=600, n=180	
	Bias	MSE	Bias	MSE	Bias	MSE
$\hat{\bar{Y}}_{\mathrm{Neyman}}$	9.026	178.910	5.936	60.147	4.122	27.583
$\hat{\bar{Y}}_{\mathrm{R1}}$	3.770	22.372	2.619	10.801	1.786	4.972
$\hat{\bar{Y}}_{\mathrm{R2}}$	3.772	22.400	2.621	10.817	1.787	4.979
$\hat{\bar{Y}}_{\mathrm{R3}}$	3.770	22.379	2.620	10.805	1.786	4.974
$\hat{\bar{Y}}_{\mathrm{Reg1}}$	3.858	23.400	2.685	11.333	1.831	5.216
$\hat{\bar{Y}}_{\mathrm{Reg2}}$	3.863	23.457	2.688	11.362	1.833	5.230
$\hat{\bar{Y}}_{\mathrm{Reg3}}$	3.859	23.413	2.686	11.339	1.831	5.219
$\hat{\bar{Y}}_{\mathrm{N1}}^{\mathrm{opt}}$	3.749	22.149	2.591	10.597	1.770	4.889
$\hat{\bar{Y}}_{\mathrm{N2}}^{\mathrm{opt}}$	3.750	22.159	2.591	10.598	1.770	4.889
$\hat{\bar{Y}}_{\mathrm{N3}}^{\mathrm{opt}}$	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 $\rho = 0.5$.

Estimator	n'=150, n=45		n'=300, n=90		n'=600, n=180	
	Bias	**MSE**	**Bias**	**MSE**	**Bias**	**MSE**
$\hat{\bar{Y}}_{\mathrm{Neyman}}$	8.242	146.555	5.421	49.903	3.758	22.878
$\hat{\bar{Y}}_{\mathrm{R1}}$	3.605	20.417	2.501	9.824	1.695	4.482
$\hat{\bar{Y}}_{\mathrm{R2}}$	3.612	20.489	2.506	9.864	1.698	4.500
$\hat{\bar{Y}}_{\mathrm{R3}}$	3.608	20.442	2.503	9.837	1.696	4.488
$\hat{\bar{Y}}_{\mathrm{Reg1}}$	3.841	23.138	2.669	11.210	1.817	5.155
$\hat{\bar{Y}}_{\mathrm{Reg2}}$	3.853	23.285	2.678	11.284	1.824	5.191
$\hat{\bar{Y}}_{\mathrm{Reg3}}$	3.845	23.189	2.672	11.235	1.819	5.168
$\hat{\bar{Y}}_{\mathrm{N1}}^{\mathrm{opt}}$	3.529	19.649	2.430	9.290	1.645	4.235
$\hat{\bar{Y}}_{\mathrm{N2}}^{\mathrm{opt}}$	3.530	19.665	2.430	9.291	1.645	4.235
$\hat{\bar{Y}}_{\mathrm{N3}}^{\mathrm{opt}}$	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 $\rho = 0.8$.

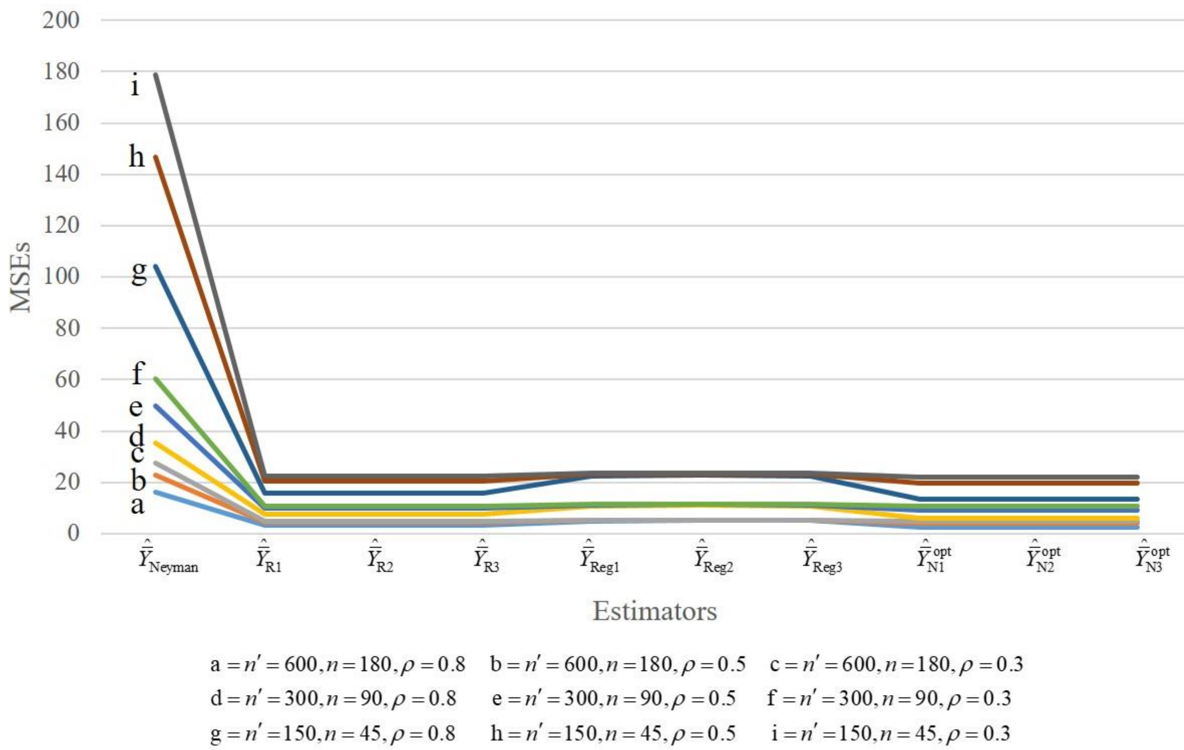
Estimator	n'=150, n=45		n'=300, n=90		n'=600, n=180	
	Bias	MSE	Bias	MSE	Bias	MSE
$\hat{\bar{Y}}_{\text{Neyman}}$	6.979	104.213	4.580	35.463	3.159	16.126
$\hat{\bar{Y}}_{\text{R1}}$	3.157	15.604	2.163	7.370	1.448	3.281
$\hat{\bar{Y}}_{\text{R2}}$	3.174	15.782	2.176	7.465	1.458	3.328
$\hat{\bar{Y}}_{\text{R3}}$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

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

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

Figure

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



Response to Reviewer#1

1. Some abbreviations are used without their definition, e.g., MSE, PRE.

Answer: The definitions of MSE and PRE have been added.

2. It is not clear about constants A,D, G,H, what are they?

Answer: A, D, G, H are constants or some known parameters such as the coefficient of variation (C_x) and the correlation coefficient (ρ) and they have been added in the text.

3. In simulation study, it would be expect to see the results based on ρ =0.5, 0.6, 0.7, 0.8, 0.9, 0.95.

Answer: As you can see similar results for ρ = 0.3, 0.5 and 0.8 for small, medium and high levels of correlation between the auxilary and study variables that is why it is not necessary to display many levels. However, as you requtested I run the simulation for $\rho = 0.95$ as below, you can see that they are similar to the results for ρ = 0.3, 0.5 and 0.8 and a higher ρ gave better results for the proposed estimators than the existing ones.

Biases and MSEs of the proposed estimators and existing estimators when $\rho = 0.95$.

Estimator	n'=150, n=45		n'=300, n=90		n'=600, n=180	
	Bias	MSE	Bias	MSE	Bias	MSE
$\hat{\bar{Y}}_{\text{Neyman}}$	6.307	85.735	4.130	28.740	2.827	12.896
$\hat{\bar{Y}}_{\text{R1}}$	2.805	12.328	1.897	5.681	1.254	2.467
$\hat{\bar{Y}}_{\text{R2}}$	2.832	12.577	1.919	5.812	1.270	2.533
$\hat{\bar{Y}}_{\text{R3}}$	2.822	12.483	1.911	5.762	1.264	2.508
$\hat{\bar{Y}}_{\text{Reg1}}$	3.702	21.843	2.559	10.379	1.753	4.846
$\hat{\bar{Y}}_{\text{Reg2}}$	3.744	22.342	2.590	10.631	1.776	4.973
$\hat{\bar{Y}}_{\text{Reg3}}$	3.728	22.149	2.578	10.535	1.767	4.926
$\hat{\bar{Y}}_{\text{N1}}^{\text{opt}}$	2.356	8.912	1.570	3.859	0.994	1.551
$\hat{\bar{Y}}_{\text{N2}}^{\text{opt}}$	2.354	8.921	1.569	3.855	0.994	1.550
$\hat{\bar{Y}}_{\text{N3}}^{\text{opt}}$	2.355	8.923	1.570	3.858	0.994	1.550

4. In Table 4, some imprecise results are presented, what is your explanation.

Answer: As reviewer#2 requested to show the bias and MSE instead of the PRE, now you can see clearer results for Table 4.

5. Mathematically, the following equation may be fixed.

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

Answer: It is correct.

6. Some applications must be added.

Answer: The proposed estimators are applied to the fine particulate matter 2.5 data in Chiang Rai, Thailand to see their performance. It is an example of an application just to support the results found theoretically and in the simulation studies. I think this should be enough just to back up what we found in both theory and simulation studies. Hope it is ok.

7. Since the proposed estimator is extended from Thongsak and Lawson (2021) [1], the comparison may be nonsense.

Answer: The proposed estimators are the combined estimators based on Thongsak and Lawson's (2022a) estimators ($\hat{\bar{Y}}_{\mathrm{R}}$ and $\hat{\bar{Y}}_{\mathrm{Reg}}$) which are single estimators, that is why we compared the combined estimators with the single estimators in both theory, simulation studies and an application to real data to see how the proposed combined estimators perform better than the single ones.

8. Please check the reference item. Proceeding must be Proceedings, Date and venue of the conference must be include.

Answer: It is corrected as below.

Thongsak, N. & Lawson, N. (2021). Classes of dual to modified ratio estimators for estimating population mean in simple random sampling. In 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

9. Based on the application (only one), the following paragraph may be revised.

Some specified value of population parameters should be mentioned.

Answer: The population parameters are already stated in this section, section 3.2. However, I have rewritten the sentences by adding $\rho = 0.921$ in the text and explanation of the results in term of biases and MSEs as below.

"Table 6 showed the results from an application to fine particulate matter in Chiang Rai also supported the results found in the simulation studies. The proposed combined estimators performed the best as you can see the biases and mean square errors of the proposed estimators were smaller than all existing estimators. The biases were at least half, and gave at least four times less mean square errors than the existing ones which can be considered high improvement. The proposed estimators work well in this application with a high correlation between NO_2 and the concentration of PM2.5 ($\rho = 0.921$) in Chiang Rai, Thailand and again, it supports what we found in the simulation studies."

Thanks a lot for your valuable comments.

Response to Reviewer#2

1. The proposed estimator, equation (7), is the weighted means of 2 estimators of Thongsak and Lawson (2022). Equation (1) – (2), except there is superscript d in equation (2) but not in equation (7). To be reasonable to do so (i.e., weighting 2 estimators), one estimator should trend to be under-estimated and over-estimated for the other. Or there should be at least a reason for smoothing these values of estimators. Verify it.

Answer: Superscript "d" has been added to equations (1) and (7).

The proposed estimator is the combined estimator from the single estimators proposed by Thongsak and Lawson (2022). A constant α is used as a weight to find the minimum value for the proposed combined estimators which of course we expect a better performance from the combined estimator using the optimum value of α that is why we implemented the proposed estimators. I have already stated this in section 2 Materials and Methods on line 3.

2. Typo: Line 46, there should be "respectively" after "are".
Line 79, there are 2 "in"s before "Chiang Rai".
Line 112, change "equation (10)" to "equation (7)".

Answer: They have been corrected.

3. In this paper, the correlation ρ between X and Y is known. I think it is not quite realistic.

Answer: Some parameters are known or available which could come from census or past data.

4. Notation for the proposed estimators should be $\hat{\hat{Y}}_{\mathrm{Nij}}^{\mathrm{opt}}$ as is defined in equation (12) rather than $\hat{\hat{Y}}_{\mathrm{Nij}}$ (without superscript) where i and j are 1, 2, 3.

Answer: Corrected.

5. For more precise, "π" should be defined as $\pi = \dfrac{n}{n'-n}$ and "K" is not yet defined.

Answer: "π" has been defined, but "K" has already been defined after equation (6).

6. The situation when the proposed estimator performs better is mentioned in Equation (14). Hence, simulation should be done for both situations where equation (14) is true and not true in order to support the findings.

Answer: From equation (14) we can see that the condition is always true and therefore the proposed estimators always perform the best compared to all existing estimators. I have added this at the end of section 2.2.

7. In table 3 – 5, what is the "existing estimator" to what proposed estimator compare?

Answer: The existing estimators are the Neyman(1938) and the Thongsak and Lawson (2022a) estimators which we have already stated in section 2.2 Efficiency Comparisons.

8. Figure 1, What is estimator 1 – 16?

Answer: The estimators have been added to Figure 1.

My opinion: I do not agree with

(1) Proposed all 9 estimators from 3x3 combinations of 3 $\hat{\hat{Y}}_{\mathrm{R}}$'s and 3 $\hat{\hat{Y}}_{\mathrm{Reg}}$'s, too many and unreasonable.

Suggestion: Determine carefully and reasonably and choose some interesting ones. Or choose at most 3 proposed estimators with the same values of A/G and D/H, namely $\hat{\bar{Y}}_{\mathrm{N11}}^{\mathrm{opt}}$, $\hat{\bar{Y}}_{\mathrm{N22}}^{\mathrm{opt}}$, $\hat{\bar{Y}}_{\mathrm{N33}}^{\mathrm{opt}}$.

(2) Determine only PRE/MSE. **Suggestion:** Biased is also important.

(3) Compare to $\hat{\bar{Y}}_{\mathrm{Neyman}}$. Neyman had proposed ratio estimator since 1938. Many researchers including Thongsak and Lawson (2022a), as referred in this paper and was the recent study, developed class of ratio estimators accordingly from Neyman's idea. So, the recent ones performed much better with no doubt.

Suggestion: Compare $\hat{\bar{Y}}_{\mathrm{Nij}}^{\mathrm{opt}}$ to $\hat{\bar{Y}}_{\mathrm{R}}$ and $\hat{\bar{Y}}_{\mathrm{Reg}}$, both for biased and PRE/MSE.

Answer: I agree with you and have corrected the results following all your suggestions above. Thanks a lot for your valuable comments.