Research article

New Ratio Estimators for Population Mean under Unequal Probability Sampling Without Replacement in the Presence of Missing Data: A Case Study on Fine Particulate Matter in Bangkok, Thailand

Chugiat Ponkaew[1] and Nuanpan Lawson[2]*

[1]Department of Mathematics and Data Science, Faculty of Science and Technology, Phetchabun Rajabhat University, Phetchabun, Thailand
[2]Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand

Curr. Appl. Sci. Technol. 2024, Vol. 24 (No. 3), e0258414; https://doi.org/10.55003/cast.2024.258414

Received: 14 April 2023, Revised: 1 July 2023, Accepted: 15 November 2023, Published: 15 February 2024

Abstract

Keywords

missing data;

nonresponse;

ratio estimator;

reverse framework;

unequal probability sampling

Missing data are frequently present in datasets and give rise to a myriad of issues that significantly affect data utilization. The missing data needs to be handled before data can be efficiently estimated and applied. New ratio estimators for population mean were proposed for use when data are missing completely at random and for a more flexible situation where missing data are missing at random in the study variable under unequal probability sampling without replacement. Furthermore, the variance estimators of the proposed ratio estimators were investigated under a reverse framework. We show theoretically that the proposed estimators were approximately unbiased estimators. The proposed estimators were utilized in simulation studies and were applied to the study of fine particulate matter data in Suan Luang District, Bangkok, Thailand in order to see how the proposed estimators performed. The results from the application to fine particulate matter showed that the ratio estimators and their variance estimators worked better than existing estimators, producing less estimated variances. Therefore, they could be applied to estimate the average fine particulate matter even when missing values appeared.

*Corresponding author: Tel.: (+66) 25552000 ext. 4903
E-mail: nuanpan.n@sci.kmutnb.ac.th

1. Introduction

The available population mean of an auxiliary variable that is positively related to the study variable can assist in estimating the population mean and make it more accurate. Cochran [1] invented a ratio estimator for estimating population mean under SRSWOR using the known population mean of an auxiliary variable in the case of a full response. Bacanli and Kadilar [2] suggested a ratio estimator for population mean of a study variable based on the Horvitz and Thompson estimator [3]. The Horvitz and Thompson estimator is a very popular linear estimator for population total that makes use of the first order inclusion probability (π_i) in its estimation. Later, Hájek [4] introduced a nonlinear estimator for population mean and it is better than the Horvitz and Thompson estimator in some situations including situations where π_i are not strong or are negatively correlated with y_i (see [5]). The ratio estimator is biased although the bias is small and can be negligible for large sample size. Thongsak and Lawson [6] investigated the bias and the mean square error (MSE) of a ratio estimator by introducing a transformation technique to solve bias and MSE under SRSWOR. They studied how bias and MSE are reduced by applying new estimators to pollution data in Nan, Thailand. They found that a transformed auxiliary variable can help decrease the bias and MSE when compared to existing estimators with no transformed format (see [7-9]).

Nonresponse, which usually occurs in sample surveys, may occur as missing completely at random (MCAR) or uniform nonresponse. MCAR occurs when the observed and missing values do not depend on the missingness. However, when there is a connection between the missingnesses and the observed values but not for missing values, nonresponse is called missing at random (MAR). Ratio estimators have also been adapted for nonresponse. Lawson [10] suggested an approximately unbiased estimator for estimating the population mean and total under probability proportional to size sampling with replacement (PPSWR) when the sampling fraction was omitted, and nonresponse was missing completely at random (MCAR). Lawson [10] invented a new way for creating a ratio estimator that did not require known response probabilities, but it was under MCAR which does not occur in practice. Ponkaew and Lawson [11] introduced a new ratio estimator for the population total under unequal probability sampling without replacement (UPWOR) using a reverse framework when the nonresponse mechanism was MCAR. Their suggested ratio estimator was based on the estimator proposed by Särndal and Lundström [12], which was an unbiased estimator for the population total based on Horvitz and Thompson's estimator [3]. Ponkaew and Lawson's estimator [11] considered a different framework when missing data. However, Sarndal and Lundstorm's [12] considered the two-phrase framework which is more complex as it uses a variance estimator under nonresponse. Both available and unavailable response probability were considered under MCAR and the sampling fraction was assumed to be small and negligible in Ponkaew and Lawson's study [11]. The results indicated that their suggested estimator performed better than the Särndal and Lundström estimator, based on the relative root mean square error. Ponkaew and Lawson [13] suggested two estimators for estimating population total linear and ratio estimators based on Särndal and Lundström [12] and Lawson [10]. It is under the reverse framework for UPWOR in cases where response probabilities are either known or unknown. The ratio estimator was in a nonlinear form and it, therefore, needed to be transformed into a linear form using more complex methods. Lawson and Ponkaew [14] introduced a new generalized regression estimator (GREG) by transforming the nonlinear estimator suggested by Lawson [10]. They suggested the use of Lawson's estimator [10] that was free from response probabilities unlike GREG estimator [3]. The Lawson and Ponkaew estimator [14] could only be used when nonresponse was under MCAR and only for a small sampling fraction which is more restricted for use. Later, Lawson and Siripanich [15] improved the GREG estimator proposed by Lawson and Ponkaew [14] by extending its applicability to a more flexible situation when nonresponse occurred under missing at random (MAR) where the response probabilities were not uniform and there was no need to ignore the sampling fraction. The Lawson and Siripanich estimator [15] could be used practically in real life situations and was based on the work of Lawson and Ponkaew [14] and on assumption that the nonresponse was MCAR. Ponkaew and Lawson [16] proposed new ratio GREG estimators in more flexible situations where the sampling fraction was both large and small and also where the response probabilities were known and unknown. Their estimators were in a general form when compared to previous ones and could be applied more flexibly. Ponkaew and Lawson [17] suggested an improvement to the ratio estimator proposed by Ponkaew and Lawson [11] that gave more flexibility when nonresponse occurred under MAR. They proposed new ratio estimators in the presence of nonresponse under UPWOR under a reverse framework when the population mean of the auxiliary variable was known and unknown and needed to be estimated from the calibration variable based on the generalized regression estimator. The results from water demand data in Thailand showed that their suggested estimators gave smaller errors, especially for large response rates.

Pollution in Thailand attributable to the deleterious repercussions of alarming levels of fine particulate matter has been perpetuated for years. This problem is destructive to human health and affects numerous sectors of industry on a large-scale. The air quality in different locations across the country is variable and tends to fluctuate between seasons as a result of disparate activity leading to emissions. Data concerning these levels must be acquired to specifically solve the cause of poor air quality in each location. An intriguing point that impedes the solution to this obstinate issue is the fact that missing data persist in reports. It is critical for missing data to be dealt with before analysis.

Estimates of air pollution data assist in planning and preparing solutions for this issue. Chodjuntug and Lawson [18] applied a new imputation method to estimate missing values of fine particulate matter with a diameter of 2.5 microns (PM2.5) in Bangkok, Thailand and then estimated the mean of the fine particulate matter under simple random sampling (SRSWOR). The estimator was in exponential form which was complex when compared to common ratio estimators but had a higher efficiency than the ratio estimators. The results indicated that the average PM2.5 was 48.20 $\mu g/m^3$ with MSE equal to 0.90 $\mu g/m^3$. Chodjuntug and Lawson [19] suggested using a response rate which was free from known parameters and therefore easy to use and also a constant to minimize MSE to estimate the PM2.5 at Kanchana Phisek Road in Bangkok. The Chodjuntug and Lawson estimator [19] was in the form of a regression and exponential estimator under SRSWOR. The results showed that the mean PM2.5 from their suggested estimator was 42.22 $\mu g/m^3$ with MSE of 0.34 $\mu g/m^3$. Lawson [20] suggested a new imputation method to estimate carbon monoxide and nitrogen dioxide based on PM2.5 in Bangkok, Thailand under SRSWOR. The Lawson estimator used the response rate, a sample regression coefficient and an optimum constant in the estimator which yielded better results than the existing estimator, producing a lower mean square error. Similar to Chodjuntug and Lawson [19], Lawson [20] also suggested the use of response rate and sample regression coefficient in the estimator so when there were no unknown auxiliary parameters the Lawson estimator could still be applied to estimate the population mean with missing data.

The purpose of this study was to introduce ratio estimators in the case of nonresponse in the study variables under the nonresponse mechanism MCAR and under a flexible situation when nonresponse was not uniformly nonresponse under MAR. We also suggested variance estimators under the reverse framework. The simulation studies and an application to air pollution data in Thailand enabled us to see the performance of the proposed estimators.

2. Materials and Methods

2.1 Basic setup

Unequal probability sampling without replacement is considered in this study. Let y be the study variable, x and z be the auxiliary variables where y has a positive correlation with x and has a negative correlation with z. Let k be a size variable which is correlated with y and it is used to define the first and joint inclusion probabilities. Let $U=\{1,2,\ldots,N\}$ be the finite population of size N and s be a set of sample size n selected from population U with unequal probability sampling without replacement. Let $\mathcal{F}$ be the set of all possible subsets of U and the sampling design $P(\bullet)$ is the probability measure on the possible s, i.e., $P(s)\geq 0$ for all $s\in\mathcal{F}$. For all i and $j\in U$, we define the notation as follows: let $\pi_i=P(i\in s)=\sum_{s\ni i}P(s)$ be the first order inclusion probability and $\pi_{ij}=P(i\wedge j\in s)=\sum_{s\supset\{i,j\}}P(s)$ be the second order inclusion probability.

Under a reverse framework, let r_i be a response indicator variable of y_i where $r_i=1$ if unit i responds to item y_i otherwise $r_i=0$. Let $\boldsymbol{R}=(r_1\ r_2\ \cdots\ r_N)'$ be the vector of the response indicator and $p_i=p=P(r_i=1)$ be the response probability under MAR. Let $E_q(\bullet)$ and $V_q(\bullet)$ be the expectation and variance operators with respect to the nonresponse mechanism. Let $E_p(\bullet)$ and $V_p(\bullet)$ be the expectation and variance operators with respect to sampling design. The overall expectation and variance operators are defined by $E(\bullet)$ and $V(\bullet)$ respectively. The finite population U is randomly classified into subpopulations according to the nonresponse mechanism that includes both respondent and non-respondent population subtotals in the first phase. Then in the second phase, a random sample is selected from the two subpopulations. Assume that $\hat{\bar{Y}}$ is the estimator of the population mean. Under a reverse framework, the variance of $\hat{\bar{Y}}$ can be obtained by

$$V(\hat{\bar{Y}})=E_qV_p\left(\hat{\bar{Y}}\middle|\boldsymbol{R}\right)+V_qE_p\left(\hat{\bar{Y}}\middle|\boldsymbol{R}\right). \quad (1)$$

2.2 The existing estimators

2.2.1 The estimators in the full response

In a full response case, Horvitz and Thompson [3] proposed a linear estimator for population total of the study variable based on sample s of size n and it is defined by

$$\hat{Y}_{HT}=\sum_{i\in s}\frac{y_i}{\pi_i}.$$

The population mean estimator of Horvitz and Thompson [3] is

$$\hat{\bar{Y}}_{HT}=\frac{1}{N}\hat{Y}_{HT}=\frac{1}{N}\sum_{i\in s}\frac{y_i}{\pi_i}. \quad (2)$$

The variance of $\hat{\bar{Y}}_{HT}$ is

$$V\left(\hat{\bar{Y}}_{HT}\right)=\frac{1}{N^2}\left[\sum_{i\in U}F_i y_i^2+\sum_{i\in U}\sum_{j\backslash\{i\}\in U}J_{ij}y_i y_j\right],$$

where $F_i=(1-\pi_i)\pi_i$, $J_{ij}=(\pi_{ij}-\pi_i\pi_j)(\pi_i\pi_j)^{-1}$.

The estimator of $V\left(\hat{\bar{Y}}_{HT}\right)$ is

$$\hat{V}\left(\hat{\bar{Y}}_{HT}\right)=\frac{1}{N^2}\left[\sum_{i\in s}\hat{F}_i y_i^2+\sum_{i\in s}\sum_{j\backslash\{i\}\in s}\hat{J}_{ij}y_i y_j\right],$$

where $\hat{F}_i=(1-\pi_i)\pi_i^2$, $\hat{J}_{ij}=(\pi_{ij}-\pi_i\pi_j)(\pi_{ij}\pi_i\pi_j)^{-1}$.

Hájek [4] proposed a nonlinear estimator for estimating population mean and it is defined by

$$\hat{\bar{Y}}_{Haj}=\frac{\sum_{i\in s}\frac{y_i}{\pi_i}}{\sum_{i\in s}\frac{1}{\pi_i}}. \tag{3}$$

The variance of $\hat{\bar{Y}}_{Haj}$ is

$$V\left(\hat{\bar{Y}}_{Haj}\right)=\frac{1}{N^2}\left[\sum_{i\in U}F_i(y_i-\bar{Y})^2+\sum_{i\in U}\sum_{j\backslash\{i\}\in U}J_{ij}(y_i-\bar{Y})(y_j-\bar{Y})\right],$$

The estimator of $V\left(\hat{\bar{Y}}_{Haj}\right)$ is

$$\hat{V}\left(\hat{\bar{Y}}_{Haj}\right)=\frac{1}{N^2}\left[\sum_{i\in s}\hat{F}_i(y_i-\bar{Y})^2+\sum_{i\in s}\sum_{j\backslash\{i\}\in s}\hat{J}_{ij}(y_i-\hat{\bar{Y}}_{Haj})(y_j-\hat{\bar{Y}}_{Haj})\right].$$

In the situations that an auxiliary variable x is available and highly correlated with the study variable y and the population mean of x is known, Bacanli and Kadilar [2] proposed a ratio estimator for estimating the population mean of the study variable y and it is given by

$$\hat{\bar{Y}}_R=\frac{\hat{\bar{Y}}_{HT}}{\hat{\bar{X}}_{HT}}\bar{X}=\hat{R}\bar{X}, \tag{4}$$

where $\hat{\bar{Y}}_{HT}=N^{-1}\sum_{i\in s}y_i\pi_i^{-1}$, $\hat{\bar{X}}_{HT}=N^{-1}\sum_{i\in s}x_i\pi_i^{-1}$ and $\bar{X}=N^{-1}\sum_{i\in U}x_i$.

The variance of $\hat{\bar{Y}}_R$ is

$$V\left(\hat{\bar{Y}}_R\right)=\frac{1}{N^2}\left[\sum_{i\in U}F_i(y_i-Rx_i)^2+\sum_{i\in U}\sum_{j\backslash\{i\}\in U}J_{ij}(y_i-Rx_i)(y_j-Rx_j)\right],$$

where $R=\dfrac{\bar{Y}}{\bar{X}}$.

The estimator of $V\left(\hat{\bar{Y}}_R\right)$ is

$$\hat{V}\left(\hat{\bar{Y}}_R\right)=\frac{1}{N^2}\left[\sum_{i\in s}\hat{F}_i(y_i-\hat{R}x_i)^2+\sum_{i\in s}\sum_{j\backslash\{i\}\in s}\hat{J}_{ij}(y_i-\hat{R}x_i)(y_j-\hat{R}x_j)\right],$$

where $\hat{R}=\dfrac{\hat{\bar{Y}}_{HT}}{\hat{\bar{X}}_{HT}}$, $\hat{\bar{X}}_{HT}=\dfrac{1}{N}\sum_{i\in s}\dfrac{x_i}{\pi_i}$ and $\hat{\bar{Y}}_{HT}$ is defined in equation (2).

When data are missing, the Horvitz and Thompson [3] and the Bacanli and Kadilar [2] estimators cannot be used to estimate population mean because they require the values of all units in the sample s.

2.2.2 The existing estimators in the presence of nonresponse

Based on a set of respondents and a reverse framework, Ponkaew and Lawson [13] suggested a linear estimator that is defined by

$$\hat{Y}_r^{\prime\prime(1)}=\sum_{i\in s}\frac{r_i y_i}{\pi_i p_i}, \tag{5}$$

If p_i is unknown under the MAR mechanism, the logistic regression or probit models can be used to approximate it. Then, the linear estimator to estimate population mean is defined by, $\hat{\bar{Y}}_r^{\prime\prime(1)}=\dfrac{1}{N}\sum_{i\in s}\dfrac{r_i y_i}{\pi_i \hat{p}_i}$.

Under a reverse framework, the variance of $\hat{\bar{Y}}_r^{\prime\prime(1)}$ is defined by

$$V\left(\hat{\bar{Y}}_r^{\prime\prime(1)}\right)=\frac{1}{N^2}\left[\sum_{i\in U}F_i y_i^2+\sum_{i\in U}\sum_{j\backslash\{i\}\in U}J_{ij}y_i y_j+\sum_{i\in U}E_i^{\prime\prime}y_i^2\right],$$

where $F_i=(1-\pi_i)\pi_i$, $J_{ij}=(\pi_{ij}-\pi_i\pi_j)(\pi_i\pi_j)^{-1}$ and $E_i^{\prime\prime}=\dfrac{1-p_i}{p_i}$.

The estimator of $V\left(\hat{\bar{Y}}_r^{\prime\prime(1)}\right)$ is given by

$$\hat{V}\left(\hat{\bar{Y}}_r^{\prime\prime(1)}\right)=\frac{1}{N^2}\left[\sum_{i\in s}\frac{r_i\hat{F}_i y_i^2}{p_i}+\sum_{i\in s}\sum_{j\backslash\{i\}\in s}\frac{r_i r_j\hat{J}_{ij}y_i y_j}{p_i p_j}+\sum_{i\in s}\frac{r_i\hat{E}_i^{\prime\prime}y_i^2}{p_i}\right], \tag{6}$$

where $\hat{F}_i = (1-\pi_i)\pi_i^2$, $\hat{J}_{ij} = (\pi_{ij} - \pi_i\pi_j)(\pi_{ij}\pi_i\pi_j)^{-1}$. In equation (6), if p_i for $i \in s$ is unknown then the estimator values of p_i are required and are obtained by the logistic regression or probit models.

For simplicity, Ponkaew and Lawson [13] considered the case when $p_i = p$ for all $i \in s$ (MCAR nonresponse mechanism) and the $\hat{\bar{Y}}_r'^{(1)}$ is

$$\hat{\bar{Y}}_r'^{(1)} = \frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i p} . \tag{7}$$

If p is unknown, it can be estimated by $\hat{p} = \sum_{i\in s}\frac{r_i}{\pi_i} \Big/ \sum_{i\in s}\frac{1}{\pi_i}$ or $\frac{r}{n}$. As a result, $\hat{\bar{Y}}_r'^{(1)} = \frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i \hat{p}}$ is the linear estimator.

Under the reverse framework with MCAR nonresponse mechanism, the variance of Ponkaew and Lawson [13] estimator is

$$V\left(\hat{\bar{Y}}_r'^{(1)}\right) = \frac{1}{N^2}\left[\sum_{i\in U} F_i y_i^2 + \sum_{i\in U}\sum_{j\backslash\{i\}\in U} J_{ij} y_i y_j + \sum_{i\in U} E_i' y_i^2\right].$$

The estimator of $V\left(\hat{\bar{Y}}_r''^{(1)}\right)$ is

$$\hat{V}\left(\hat{\bar{Y}}_r'^{(1)}\right) = \frac{1}{N^2}\left[\sum_{i\in s}\frac{r_i \hat{F}_i y_i^2}{p} + \sum_{i\in s}\sum_{j\backslash\{i\}\in s}\frac{r_i r_j \hat{J}_{ij} y_i y_j}{p^2} + \sum_{i\in s}\frac{r_i \hat{E}_i' y_i^2}{p}\right], \tag{8}$$

where $\hat{F}_i = (1-\pi_i)\pi_i^2$, $\hat{J}_{ij} = (\pi_{ij} - \pi_i\pi_j)(\pi_{ij}\pi_i\pi_j)^{-1}$ and $\hat{E}' = \frac{1-p}{p\pi_i}$.

The estimator $\hat{\bar{Y}}_r'^{(1)}$ in equation (7) requires the response probability. Therefore, Ponkaew and Lawson [13] also suggested a nonlinear estimator for population mean that did not require the response probability as follows.

$$\hat{\bar{Y}}_r'^{(2)} = \frac{\sum_{i\in s}\frac{r_i y_i}{\pi_i p}}{\sum_{i\in s}\frac{r_i p}{\pi_i}} = \frac{\sum_{i\in s}\frac{r_i y_i}{\pi_i}}{\sum_{i\in s}\frac{r_i}{\pi_i}} . \tag{9}$$

The variance of $\hat{\bar{Y}}_r'^{(2)}$ is

$$V\left(\hat{\bar{Y}}_r'^{(2)}\right) = \frac{1}{N^2}\left[\sum_{i\in U} F_i\left(y_i - \bar{Y}\right)^2 + \sum_{i\in U}\sum_{j\backslash\{i\}\in U} J_{ij}\left(y_i - \bar{Y}\right)\left(y_j - \bar{Y}\right) + \sum_{i\in U} E_i'(y_i - \bar{Y})^2\right],$$

where $E_i' = \frac{1-p}{p}$.

The estimator of $V\left(\hat{\bar{Y}}_r'^{(2)}\right)$ is

$$\hat{V}\left(\hat{\bar{Y}}_r'^{(2)}\right) = \frac{1}{N^2}\left[\sum_{i\in s}\frac{r_i\hat{F}_i(y_i-\hat{\bar{Y}}_r'^{(2)})^2}{p} + \sum_{i\in s}\sum_{j\backslash\{i\}\in s}\frac{r_ir_j\hat{J}_{ij}(y_i-\hat{\bar{Y}}_r'^{(2)})(y_i-\hat{\bar{Y}}_r'^{(2)})}{p^2} + \sum_{i\in s}\frac{r_i\hat{E}_i'(y_i-\hat{\bar{Y}}_r'^{(2)})^2}{p}\right]. \tag{10}$$

Lawson [13] proposed a general form of the nonlinear estimator given by

$$\hat{\bar{Y}}_r''^{(2)} = \frac{\sum_{i\in s}\frac{r_iy_i}{\pi_ip_i}}{\sum_{i\in s}\frac{r_i}{\pi_ip_i}}. \tag{11}$$

However, the estimator of Lawson [10] was considered under PPSWR and could be used to investigate the variance only when the sampling fraction was negligible. We investigate the variance and associated estimator of Lawson [10] under UPWOR when the sampling fraction is not negligible as follows.

The variance of $\hat{\bar{Y}}_r''^{(2)}$ is

$$V\left(\hat{\bar{Y}}_r''^{(2)}\right) = \frac{1}{N^2}\left[\sum_{i\in U}F_i\left(y_i-\bar{Y}\right)^2 + \sum_{i\in U}\sum_{j\backslash\{i\}\in U}J_{ij}\left(y_i-\bar{Y}\right)\left(y_j-\bar{Y}\right) + \sum_{i\in U}E_i''(y_i-\bar{Y})^2\right].$$

The estimator of $V\left(\hat{\bar{Y}}_r''^{(2)}\right)$ is

$$\hat{V}\left(\hat{\bar{Y}}_r''^{(2)}\right) = \frac{1}{N^2}\left[\sum_{i\in s}\frac{r_i\hat{F}_i(y_i-\hat{\bar{Y}}_r''^{(2)})^2}{p_i} + \sum_{i\in s}\sum_{j\backslash\{i\}\in s}\frac{r_ir_j\hat{J}_{ij}(y_i-\hat{\bar{Y}}_r''^{(2)})(y_j-\hat{\bar{Y}}_r''^{(2)})}{p_ip_j} + \sum_{i\in s}\frac{r_i\hat{E}_i''(y_i-\hat{\bar{Y}}_r''^{(2)})^2}{p_i}\right]. \tag{12}$$

When the data are missing in the study variable y under MCAR and the auxiliary variable x is available and the population total of x is known, Ponkaew and Lawson [11] proposed a ratio estimator for population total based on a linear estimator given by

$$\hat{Y}_R'^{(1)} = \frac{\sum_{i\in s}\frac{r_iy_i}{\pi_ip}}{\sum_{i\in s}\frac{x_i}{\pi_i}}X.$$

Then, the ratio estimator for estimating population mean is

$$\hat{\bar{Y}}_R'^{(1)} = \frac{\frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i p}}{\frac{1}{N}\sum_{i\in s}\frac{x_i}{\pi_i}}\bar{X} = \frac{\hat{\bar{Y}}_r'^{(1)}}{\hat{\bar{X}}_{HT}}\bar{X} \;, \tag{13}$$

where $\hat{\bar{Y}}_r'^{(1)} = \frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i p}$.

Under the reverse framework, the variance of $\hat{\bar{Y}}_R'^{(1)}$ is

$$V\left(\hat{\bar{Y}}_R'^{(1)}\right) = \frac{1}{N^2}\left[\sum_{i\in U} F_i\left(y_i - Rx_i\right)^2 + \sum_{i\in U}\sum_{j\backslash\{i\}\in U} J_{ij}\left(y_i - Rx_i\right)\left(y_j - Rx_j\right) + \sum_{i\in U} E_i'(y_i - Rx_i)^2\right].$$

The estimator of $V\left(\hat{\bar{Y}}_R'^{(1)}\right)$ is

$$\hat{V}\left(\hat{\bar{Y}}_R'^{(1)}\right) = \frac{1}{N^2}\left[\sum_{i\in s}\frac{r_i\hat{F}_i(y_i - \hat{R}'^{(1)}x_i)^2}{p} + \sum_{i\in s}\sum_{j\backslash\{i\}\in s}\frac{r_i r_j \hat{J}_{ij}(y_i - \hat{R}'^{(1)}x_i)(y_j - \hat{R}'^{(1)}x_j)}{p^2} + \sum_{i\in s}\frac{r_i\hat{E}_i'(y_i - \hat{R}'^{(1)}x_i)^2}{p}\right], \tag{14}$$

where $\hat{R}'^{(1)} = \frac{\frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i p}}{\frac{1}{N}\sum_{i\in s}\frac{x_i}{\pi_i}}$.

Ponkaew and Lawson [17] adapted the ratio estimator proposed by Ponkaew and Lawson [11] for a more flexible situation when nonresponse occurs under MAR. The Ponkaew and Lawson equation [17] is

$$\hat{\bar{Y}}_R''^{(1)} = \frac{\frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i p_i}}{\frac{1}{N}\sum_{i\in s}\frac{x_i}{\pi_i}}\bar{X} = \frac{\hat{\bar{Y}}_r''^{(1)}}{\frac{1}{N}\sum_{i\in s}\frac{x_i}{\pi_i}}\bar{X} \;, \tag{15}$$

where $\hat{\bar{Y}}_r''^{(1)} = \frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i p_i}$.

Under the reverse framework, the variance of $\hat{\bar{Y}}_R'^{(1)}$ is

$$V\left(\hat{\bar{Y}}_R''^{(1)}\right)=\frac{1}{N^2}\left[\sum_{i\in U}F_i\left(y_i-Rx_i\right)^2+\sum_{i\in U}\sum_{j\backslash\{i\}\in U}J_{ij}\left(y_i-Rx_i\right)\left(y_j-Rx_j\right)+\sum_{i\in U}E_i''(y_i-Rx_i)^2\right].$$

The estimator of $V\left(\hat{\bar{Y}}_R''^{(1)}\right)$ is

$$\hat{V}\left(\hat{\bar{Y}}_R''^{(1)}\right)=\frac{1}{N^2}\left[\sum_{i\in s}\frac{r_i\hat{F}_i(y_i-\hat{R}''^{(1)}x_i)^2}{p_i}+\sum_{i\in s}\sum_{j\backslash\{i\}\in s}\frac{r_ir_j\hat{J}_{ij}(y_i-\hat{R}''^{(1)}x_i)(y_j-\hat{R}''^{(1)}x_j)}{p_ip_j}\right.$$
$$\left.+\sum_{i\in s}\frac{r_i\hat{E}_i''(y_i-\hat{R}''^{(1)}x_i)^2}{p_i}\right], \qquad (16)$$

where $\hat{R}''^{(1)}=\dfrac{\frac{1}{N}\sum_{i\in s}\frac{r_iy_i}{\pi_ip_i}}{\frac{1}{N}\sum_{i\in s}\frac{x_i}{\pi_i}}$.

3. Results and Discussion

3.1 The proposed estimators

In this section, we assumed that the population mean of the auxiliary variable $\bar{X}$ is known and nonresponse occurs in the study variable y. The nonresponse mechanisms are considered under both MCAR and MAR.

Under MCAR, we proposed a new estimator for estimating population mean by adjusting the Ponkaew and Lawson estimator [11] using Ponkaew and Lawson estimator [13] by replacing $\hat{\bar{Y}}_r'^{(1)}$ from equation (13) with $\hat{\bar{Y}}_r'^{(2)}$ from equation (9). Then the proposed population mean estimator is

$$\hat{\bar{Y}}_R'^{*}=\frac{\sum_{i\in s}\frac{r_iy_i}{\pi_i}\left(\sum_{i\in s}\frac{r_i}{\pi_i}\right)^{-1}}{\frac{1}{N}\sum_{i\in s}\frac{x_i}{\pi_i}}\bar{X}=\frac{\hat{\bar{Y}}_r'^{(2)}}{\hat{\bar{X}}_{HT}}\bar{X}=\hat{R}'^{(2)}\bar{X}, \qquad (17)$$

where $\hat{\bar{Y}}_r'^{(2)}=\sum_{i\in s}\frac{r_iy_i}{\pi_i}\left(\sum_{i\in s}\frac{r_i}{\pi_i}\right)^{-1}$ and $\hat{R}'^{(2)}=\dfrac{\hat{\bar{Y}}_r'^{(2)}}{\hat{\bar{X}}_{HT}}$.

We suggested a new ratio estimator under MCAR in equation (17) to estimate the population mean. In Theorem 1, we then demonstrated that the proposed estimator $\hat{\bar{Y}}_R'^{*}$ is an approximately unbiased estimator of $\bar{Y}$.

Theorem 1. *Assume that the nonresponse mechanism is MCAR, and* $\hat{\tilde{\bar{Y}}}_R'^{*}$ is an approximately unbiased estimator of $\bar{Y}$.

Proof. We show that $E(\hat{\tilde{\bar{Y}}}_R'^{*}) = E_q E_p\left(\hat{\tilde{\bar{Y}}}_R'^{*} \middle| \boldsymbol{R}\right) \cong \bar{Y}$.

Let $\hat{\tilde{\bar{Y}}}_R'^{*} = \dfrac{\hat{\tilde{\bar{Y}}}_r'^{(2)}}{\hat{\bar{X}}_{HT}} \bar{X}$.

Under the reverse framework the overall expectation of $\hat{\tilde{\bar{Y}}}_R'^{*}$ is defined by

$$E(\hat{\tilde{\bar{Y}}}_R'^{*}) = E_q E_p\left(\hat{\tilde{\bar{Y}}}_R'^{*} \middle| \boldsymbol{R}\right).$$

However, $\hat{\tilde{\bar{Y}}}_R'^{*}$ is in a form of a nonlinear function. The Taylor linearization approach is applied to transform this estimator into a linear function defined by

$$\hat{\tilde{\bar{Y}}}_{R,lin}'^{*} \cong \tilde{\bar{Y}}_r'^{(2)} + \frac{1}{T_2}\left[(\hat{T}_1 - T_1) - \tilde{\bar{Y}}_r'^{(2)}(\hat{T}_2 - T_2) - \tilde{R}'(\hat{T}_3 - T_3)\right], \tag{18}$$

where $\hat{T}_1 = \sum_{i\in s} \frac{r_i y_i}{\pi_i p_i}$, $T_1 = \sum_{i\in U} \frac{r_i y_i}{p_i}$, $\hat{T}_2 = \sum_{i\in s} \frac{r_i}{\pi_i p_i}$, $T_2 = \sum_{i\in U} \frac{r_i}{p_i}$, $\hat{T}_3 = \frac{1}{N}\sum_{i\in s} \frac{x_i}{\pi_i}$, $T_3 = \bar{X}$, $\tilde{\bar{Y}}_r'^{(2)} = \frac{T_1}{T_2}$ and $\tilde{R}' = \frac{T_1}{T_3}$.

Then, $E(\hat{\tilde{\bar{Y}}}_R'^{*})$ can be approximated from

$$E(\hat{\tilde{\bar{Y}}}_R'^{*}) \cong E_q E_p\left(\hat{\tilde{\bar{Y}}}_{R,lin}'^{*} \middle| \boldsymbol{R}\right) = E_q E_p\left(\tilde{\bar{Y}}_r'^{(2)} + \frac{1}{T_2}\left[(\hat{T}_1 - T_1) - \tilde{\bar{Y}}_r'^{(2)}(\hat{T}_2 - T_2) - \tilde{R}(\hat{T}_3 - T_3)\right] \middle| \boldsymbol{R}\right)$$

$$= E_q\left(\tilde{\bar{Y}}_r'^{(2)} + \frac{1}{T_2}\left[(T_1 - T_1) - \tilde{\bar{Y}}_r'^{(2)}(T_2 - T_2) - \tilde{R}(T_2 - T_3)\right] \middle| \boldsymbol{R}\right)$$

$$= E_q\left(\tilde{\bar{Y}}_r'^{(2)} \middle| \boldsymbol{R}\right) = E_q\left(\frac{\sum_{i\in U} \frac{r_i y_i}{p_i}}{\sum_{i\in U} \frac{r_i}{p_i}} \middle| \boldsymbol{R}\right) \cong \frac{1}{N}\sum_{i\in U} y_i = \bar{Y}.$$

Therefore, $E(\hat{\tilde{\bar{Y}}}_R'^{*}) \cong \bar{Y}$ and in other words, $\hat{\tilde{\bar{Y}}}_R'^{*}$ is an approximately unbiased estimator of $\bar{Y}$.

Under MAR, we proposed a new ratio estimator for estimating population mean by adjusting the Ponkaew and Lawson estimator [17] using the Lawson estimator [11] by replacing $\hat{\tilde{\bar{Y}}}_r''^{(1)}$ from equation (15) with $\hat{\tilde{\bar{Y}}}_r''^{(2)}$ from equation (11). Then, the proposed population mean estimator is

$$\hat{\bar{Y}}_R''^{*} = \frac{\sum_{i\in s}\frac{r_i y_i}{\pi_i p_i}\left(\sum_{i\in s}\frac{r_i}{\pi_i p_i}\right)^{-1}}{\frac{1}{N}\sum_{i\in s}\frac{x_i}{\pi_i}}\bar{X} = \frac{\hat{\bar{Y}}_r''^{(2)}}{\hat{\bar{X}}_{HT}}\bar{X} = \hat{R}''^{(2)}\bar{X}, \tag{19}$$

where $\hat{\bar{Y}}_r''^{(2)} = \sum_{i\in s}\frac{r_i y_i}{\pi_i p_i}\left(\sum_{i\in s}\frac{r_i}{\pi_i p_i}\right)^{-1}$ and $\hat{R}''^{(2)} = \frac{\hat{\bar{Y}}_r''^{(2)}}{\hat{\bar{X}}_{HT}}$.

In Theorem 2 as follows, we demonstrate that $\hat{\bar{Y}}_R''^{*}$ is an approximately unbiased estimator of $\bar{Y}$.

Theorem 2. *Assume that the nonresponse mechanism is MAR, and $\hat{\bar{Y}}_R''^{*}$ is an approximately unbiased estimator of $\bar{Y}$.*

Proof. The proof is similar to Theorem 1.

In Theorem 1 and Theorem 2, we showed that the proposed ratio estimators both under MCAR and MAR are approximately unbiased estimators of the population mean. Since the new ratio estimators are in nonlinear forms, Taylor linearization is applied to transform them into linear forms. In the next step, we investigate the variance and associated estimators of the proposed estimators under the reverse framework.

3.2 The proposed variance estimators

Under MCAR, Theorem 3 illustrates the variance and its estimator for the proposed ratio estimator $\hat{\bar{Y}}_R'^{*}$.

Theorem 3. *Under a reverse framework with the nonresponse mechanism MCAR.*

(1) The variance of $\hat{\bar{Y}}_R'^{*}$ is

$$V\left(\hat{\bar{Y}}_R'^{*}\right) = \frac{1}{N^2 p^2}\left[\sum_{i\in U} F_i Z_i'^2 + \sum_{i\in U}\sum_{j\backslash\{i\}\in U} J_{ij} Z_i' Z_j' + \sum_{i\in U} E_i'(y_i - \bar{Y})^2\right],$$

where $F_i = (1-\pi_i)\pi_i$, $J_{ij} = (\pi_{ij} - \pi_i\pi_j)(\pi_i\pi_j)^{-1}$, $E_i' = (1-p)p$, $Z_i' = p(y_i - \bar{Y}) - \frac{R}{N}x_i$, $Z_j' = p(y_j - \bar{Y}) - \frac{R}{N}x_j$ and $R = \frac{\bar{Y}}{\bar{X}}$.

(2) The estimator of $\hat{V}'\left(\hat{\bar{Y}}_R'^{*}\right)$ is

$$\hat{V}\left(\hat{\bar{Y}}_R'^{*}\right) = \frac{1}{\left(\sum_{i\in s}\frac{r_i}{\pi_i}\right)^2}\left[\sum_{i\in s}\hat{F}_i\hat{Z}_i'^2 + \sum_{i\in s}\sum_{j\backslash\{i\}\in s}\hat{J}_{ij}\hat{Z}_i'\hat{Z}_j' + \sum_{i\in s}\frac{r_i}{p}\hat{E}_i'(y_i - \hat{\bar{Y}}_R'^{(2)})^2\right],$$

where $\hat{Z}_i' = r_i(y_i - \hat{\bar{Y}}_R'^*) - \frac{\hat{R}'^{(2)}}{N} x_i$, $\hat{Z}_j' = r_j(y_j - \hat{\bar{Y}}_R'^*) - \frac{\hat{R}'^{(2)}}{N} x_j$, $\hat{R}'^{(2)} = \frac{\hat{\bar{Y}}_r'^{(2)}}{\hat{\bar{X}}_{HT}}$,
$\hat{\bar{Y}}_r'^{(2)} = \sum_{i\in s} \frac{r_i y_i}{\pi_i} \left(\sum_{i\in s} \frac{r_i}{\pi_i} \right)^{-1}$, $\hat{F}_i = (1-\pi_i)\pi_i^2$, $\hat{J}_{ij} = (\pi_{ij} - \pi_i \pi_j)(\pi_{ij}\pi_i\pi_j)^{-1}$, $\hat{E}_i' = \frac{(1-p)p}{\pi_i}$.

(3) If p is known, then the estimator of $V\left(\hat{\bar{Y}}_R'^*\right)$ is

$$\hat{V}\left(\hat{\bar{Y}}_R'^*\right) = \frac{1}{\left(\sum_{i\in s} \frac{r_i}{\pi_i} \right)^2} \left[\sum_{i\in s} \hat{F}_i \hat{Z}_i'^2 + \sum_{i\in s} \sum_{j\setminus\{i\}\in s} \hat{J}_{ij} \hat{Z}_i' \hat{Z}_j' + \sum_{i\in s} \frac{r_i}{\hat{p}} \hat{E}_i' (y_i - \hat{\bar{Y}}_R'^{(2)})^2 \right],$$

where $\hat{p} = \frac{r}{n}$ or $\hat{p} = \sum_{i\in s} \frac{r_i}{\pi_i} \left(\sum_{i\in s} \frac{1}{\pi_i} \right)^{-1}$ is the estimator of p and $\hat{E}_i' = \frac{(1-\hat{p})\hat{p}}{\pi_i}$.

Proof. Assume that the nonresponse mechanism is MCAR.
(1) Let

$$\hat{\bar{Y}}_R'^* = \frac{\sum_{i\in s} \frac{r_i y_i}{\pi_i} \left(\sum_{i\in s} \frac{r_i}{\pi_i} \right)^{-1}}{\frac{1}{N} \sum_{i\in s} \frac{x_i}{\pi_i}} \bar{X} = \frac{\hat{\bar{Y}}_r'^{(2)}}{\hat{\bar{X}}_{HT}} \bar{X} = \hat{R}'^{(2)} \bar{X} .$$

Under the reverse framework, the variance of $\hat{\bar{Y}}_R'^*$ is equal to

$$V(\hat{\bar{Y}}_R'^*) = V'(\hat{\bar{Y}}_R'^*) = E_q V_p (\hat{\bar{Y}}_R'^* \mid \boldsymbol{R}) + V_q E_p (\hat{\bar{Y}}_R'^* \mid \boldsymbol{R}) = V_1 + V_2 , \tag{20}$$

where $V_1 = E_q V_p (\hat{\bar{Y}}_R'^* \mid \boldsymbol{R})$ and $V_2 = V_q E_p (\hat{\bar{Y}}_R'^* \mid \boldsymbol{R})$.

Step 1: Investigate the formula of $V_1 = E_q V_p (\hat{\bar{Y}}_R'^* \mid \boldsymbol{R})$.

Since $\hat{\bar{Y}}_R'^*$ is in the form of a nonlinear function, the Taylor linearization approach is used to transform this estimator into a linear function.

$$\hat{\bar{Y}}_{R,lin}'^* \cong \tilde{\bar{Y}}_r'^{(2)} + \frac{1}{T_2} \left[(\hat{T}_1 - T_1) - \tilde{\bar{Y}}_r'^{(2)} (\hat{T}_2 - T_2) - \tilde{R}'^{(2)} (\hat{T}_3 - T_3) \right], \tag{21}$$

where $\hat{T}_1 = \sum_{i\in s} \frac{r_i y_i}{\pi_i}$, $T_1 = \sum_{i\in U} r_i y_i$, $\hat{T}_2 = \sum_{i\in s} \frac{r_i}{\pi_i}$, $T_2 = \sum_{i\in U} r_i$, $\hat{T}_3 = \frac{1}{N} \sum_{i\in s} \frac{x_i}{\pi_i}$, $T_3 = \bar{X}$, $\tilde{\bar{Y}}_r'^{(2)} = \frac{T_1}{T_2}$ and $\tilde{R}'^{(2)} = \frac{T_1}{T_3}$.

We may rewrite $\hat{\tilde{Y}}'^{*}_{R,lin}$ in (21) as

$$\hat{\tilde{Y}}'^{*}_{R,lin} \cong C_o + \frac{1}{T_2}\sum_{i\in s}\frac{\tilde{Z}'_i}{\pi_i}, \tag{22}$$

where $\tilde{Z}'_i = r_i\left(y_i - \tilde{\bar{Y}}'^{(2)}_r\right) - \frac{\tilde{R}'^{(2)}}{N}x_i$, $\tilde{\bar{Y}}'^{(2)}_r = \frac{\sum_{i\in U} r_i y_i}{\sum_{i\in U} r_i}$, $\tilde{R}' = \frac{\sum_{i\in U} r_i y_i}{\bar{X}}$ and C_o is constant.

Next, we consider $E_q\left(\tilde{Z}'_i \middle| \boldsymbol{R}\right) = Z'_i = p(y_i - \bar{Y}) - \frac{R}{N}x_i$.

Then, $V_1 = E_q V_p(\hat{\tilde{Y}}'^{*}_R \mid \boldsymbol{R})$ can be approximated by

$$V_1 = E_q V_p\left(\hat{\tilde{Y}}'^{*}_R \middle| \boldsymbol{R}\right) \cong E_q V_p\left(\hat{\tilde{Y}}'^{*}_{R,lin} \middle| \boldsymbol{R}\right) = E_q V_p\left(C_o + \frac{1}{T_2}\sum_{i\in s}\frac{\tilde{Z}'_i}{\pi_i} \middle| \boldsymbol{R}\right) = E_q V_p\left(\frac{1}{T_2}\sum_{i\in s}\frac{\tilde{Z}'_i}{\pi_i} \middle| \boldsymbol{R}\right)$$
$$= E_q\left(\frac{1}{T_2^2}\left[\sum_{i\in U} F_i \tilde{Z}'^2_i + \sum_{i\in U}\sum_{j\backslash\{i\}\in U} J_{ij}\tilde{Z}'_i\tilde{Z}'_j\right] \middle| \boldsymbol{R}\right) = \frac{1}{N^2p^2}\left[\sum_{i\in U} F_i Z'^2_i + \sum_{i\in U}\sum_{j\backslash\{i\}\in U} J_{ij} Z'_i Z'_j\right].$$

Therefore,

$$V_1 \cong \frac{1}{N^2p^2}\left[\sum_{i\in U} F_i Z'^2_i + \sum_{i\in U}\sum_{j\backslash\{i\}\in U} J_{ij} Z'_i Z'_j\right]. \tag{23}$$

Step 2: Investigate the formula of $V_2 = V_q E_p\left(\hat{\tilde{Y}}'^{*}_R \middle| \boldsymbol{R}\right)$**.**

The formula of $V_2 = V_q E_p\left(\hat{\tilde{Y}}'^{*}_R \middle| \boldsymbol{R}\right)$ can be obtained by

$$V_2 = V_q E_p\left(\hat{\tilde{Y}}'^{*}_R \middle| \boldsymbol{R}\right)$$
$$\cong V_q E_p\left(\hat{\tilde{Y}}'^{*}_{R,lin} \middle| \boldsymbol{R}\right) = V_q E_p\left(\tilde{\bar{Y}}'^{(2)}_r + \frac{1}{T_2}\left[(\hat{T}_1 - T_1) - \tilde{\bar{Y}}'^{(2)}_r(\hat{T}_2 - T_2) - \tilde{R}'(\hat{T}_3 - T_3)\right] \middle| \boldsymbol{R}\right)$$
$$= V_q\left(\tilde{\bar{Y}}'^{(2)}_r \middle| \boldsymbol{R}\right) = V_q\left(\frac{\sum_{i\in U} r_i y_i}{\sum_{i\in U} r_i} \middle| \boldsymbol{R}\right). \tag{24}$$

Since the term $\sum_{i\in U} r_i y_i\left(\sum_{i\in U} r_i\right)^{-1}$ in equation (24) is in the form of a nonlinear function, the Taylor linearization approach is used to transform this estimator into a linear function defined by

$$\frac{\sum_{i\in U} r_i y_i}{\sum_{i\in U} r_i} \cong C_1 + \frac{1}{Np}\sum_{i\in U} r_i(y_i - \bar{Y}),$$

where C_1 is a constant.

Therefore, $V_2 = V_q E_p\left(\hat{\bar{Y}}_R'^{*}\middle|\boldsymbol{R}\right)$ can be approximated from

$$V_2 \cong V_q E_p\left(C_1 + \frac{1}{Np}\sum_{i\in U} r_i(y_i - \bar{Y})\boldsymbol{R}\right) = \frac{1}{N^2 p^2}\sum_{i\in U} E_i'(y_i - \bar{Y})^2. \tag{25}$$

Substitute equations (23) and (25) in equation (20), the variance of $\hat{\bar{Y}}_R'^{*}$ can then be approximated by

$$V\left(\hat{\bar{Y}}_R'^{*}\right) = \frac{1}{N^2 p^2}\left[\sum_{i\in U} F_i Z_i'^2 + \sum_{i\in U}\sum_{j\backslash\{i\}\in U} J_{ij} Z_i' Z_j' + \sum_{i\in U} E_i'(y_i - \bar{Y})^2\right],$$

where $F_i = (1-\pi_i)\pi_i$, $J_{ij} = (\pi_{ij} - \pi_i\pi_j)(\pi_i\pi_j)^{-1}$, $E_i' = (1-p)p$, $Z_i' = p(y_i - \bar{Y}) - \frac{R}{N}x_i$ and $R = \frac{\bar{Y}}{\bar{X}}$.

(2) Assume that p is known.

Recall from equation (20), the variance of $\hat{\bar{Y}}_R'^{*}$ is equal to

$$V(\hat{\bar{Y}}_R'^{*}) = E_q V_p(\hat{\bar{Y}}_R'^{*} \mid \boldsymbol{R}) + V_q E_p(\hat{\bar{Y}}_R'^{*} \mid \boldsymbol{R}) = V_1 + V_2, \tag{26}$$

where $V_1 = E_q V_p(\hat{\bar{Y}}_R'^{*} \mid \boldsymbol{R})$ and $V_2 = V_q E_p(\hat{\bar{Y}}_R'^{*} \mid \boldsymbol{R})$.

Let $\hat{V}(\hat{\bar{Y}}_R'^{*})$ be the estimator of $V(\hat{\bar{Y}}_R'^{*})$ under the reverse framework defined by

$$\hat{V}(\hat{\bar{Y}}_R'^{*}) = \hat{V}_1 + \hat{V}_2, \tag{27}$$

where $\hat{V}_m$ are the estimators of V_m and $m = 1,2$. Next, we investigate the estimator of V_m where $m = 1,2$ as follows.

Step 1: Estimate $V_1 = E_q V_p(\hat{\bar{Y}}_R'^{*} \mid \boldsymbol{R})$.

Let $\hat{V}_1$ be the estimator of $V_1 = E_q V_p(\hat{\bar{Y}}_R'^{*} \mid \boldsymbol{R})$. To estimate V_1, we use the procedure from Ponkaew and Lawson [14] defined by

$$\hat{V}_1 = \hat{V}_p(\hat{\tilde{Y}}_R'^{*} \mid \boldsymbol{R}) \cong \hat{V}_p\left(\frac{1}{T_2}\sum_{i\in s}\frac{\tilde{Z}_i'}{\pi_i}\middle|\boldsymbol{R}\right), \tag{28}$$

where $T_2 = \sum_{i\in U} r_i$ and $\tilde{Z}_i' = r_i\left(y_i - \tilde{\bar{Y}}_r'^{(2)}\right) - \frac{\tilde{R}'}{N} x_i$.

The term $\frac{1}{T_2}\sum_{i\in s}\frac{\tilde{Z}_i'}{\pi_i}$ is in the form of Horvitz and Thompson's [6] estimator, and then

$$\hat{V}_1 \cong \hat{V}_p\left(\frac{1}{T_2}\sum_{i\in s}\frac{\tilde{Z}_i'}{\pi_i}\middle|\boldsymbol{R}\right) = \frac{1}{\hat{T}_2^2}\left[\sum_{i\in s}\hat{F}_i\hat{Z}_i'^2 + \sum_{i\in s}\sum_{j\backslash\{i\}\in s}\hat{J}_{ij}\hat{Z}_i'\hat{Z}_j'\right]$$
$$= \frac{1}{\left(\sum_{i\in s}\frac{r_i}{\pi_i}\right)^2}\left[\sum_{i\in s}\hat{F}_i\hat{Z}_i'^2 + \sum_{i\in s}\sum_{j\backslash\{i\}\in s}\hat{J}_{ij}\hat{Z}_i'\hat{Z}_j'\right].$$

Therefore,

$$\hat{V}_1 \cong \frac{1}{\left(\sum_{i\in s}\frac{r_i}{\pi_i}\right)^2}\left[\sum_{i\in s}\hat{F}_i\hat{Z}_i'^2 + \sum_{i\in s}\sum_{j\backslash\{i\}\in s}\hat{J}_{ij}\hat{Z}_i'\hat{Z}_j'\right]. \tag{29}$$

Step 2: Estimate $V_2 = V_q E_p\left(\hat{\tilde{Y}}_R'^{*}\middle|\boldsymbol{R}\right)$.

From equation (25), $V_2 \cong \frac{1}{N^2 p^2}\sum_{i\in U} E_i'(y_i - \bar{Y})^2$. Considering $\sum_{i\in s}\frac{r_i}{\pi_i}$ under the reverse framework $E_q E_p\left(\sum_{i\in s}\frac{r_i}{\pi_i}\middle|\boldsymbol{R}\right) = Np$, then $\sum_{i\in s}\frac{r_i}{\pi_i}$ is the estimator of Np. Furthermore, $\sum_{i\in s}\frac{r_i}{p}\hat{E}_i'(y_i - \hat{\tilde{Y}}_R'^{(2)})^2$ is the approximately unbiased estimator of $\sum_{i\in U} E_i'(y_i - \bar{Y})^2$. Then, the estimator of V_2 is

$$\hat{V}_2 \cong \frac{1}{\left(\sum_{i\in s}\frac{r_i}{\pi_i}\right)^2}\sum_{i\in s}\frac{r_i}{p}\hat{E}_i'(y_i - \hat{\tilde{Y}}_R'^{*})^2. \tag{30}$$

Substituting equations (29) and (30) in equation (27), $V'(\hat{\tilde{Y}}_R'^{*})$ can be approximated by

$$\hat{V}\left(\hat{\bar{Y}}_R'^{*}\right)=\frac{1}{\left(\sum_{i\in s}\frac{r_i}{\pi_i}\right)^2}\left[\sum_{i\in s}\hat{F}_i\hat{Z}_i'^2+\sum_{i\in s}\sum_{j\backslash\{i\}\in s}\hat{J}_{ij}\hat{Z}_i'\hat{Z}_j'+\sum_{i\in s}\frac{r_i}{p}\hat{E}_i'(y_i-\hat{\bar{Y}}_R'^{(2)})^2\right].$$

(3) Assuming that p is unknown, then under the reverse framework it can be approximated by $\hat{p}=\frac{r}{n}$ or $\hat{p}=\sum_{i\in s}\frac{r_i}{\pi_i}\left(\sum_{i\in s}\frac{1}{\pi_i}\right)^{-1}$. Therefore, the variance of $\hat{\bar{Y}}_R'^{*}$ can be approximated by

$$\hat{V}\left(\hat{\bar{Y}}_R'^{*}\right)=\frac{1}{\left(\sum_{i\in s}\frac{r_i}{\pi_i}\right)^2}\left[\sum_{i\in s}\hat{F}_i\hat{Z}_i'^2+\sum_{i\in s}\sum_{j\backslash\{i\}\in s}\hat{J}_{ij}\hat{Z}_i'\hat{Z}_j'+\sum_{i\in s}\frac{r_i}{\hat{p}}\hat{E}_i'(y_i-\hat{\bar{Y}}_R'^{(2)})^2\right],$$

where $\hat{p}$ is the estimator of p and $\hat{E}'=\frac{E_i'}{\pi_i}=\frac{(1-p)p}{\pi_i}$.

Theorem 4. *Under a reverse framework with nonresponse mechanism MAR.*
(1) The variance of $\hat{\bar{Y}}_R''^{*}$ is

$$V\left(\hat{\bar{Y}}_R''^{*}\right)=\frac{1}{N^2}\left[\sum_{i\in U}F_iZ_i''^2+\sum_{i\in U}\sum_{j\backslash\{i\}\in U}J_{ij}Z_i''Z_j''+\sum_{i\in U}E_i''(y_i-\bar{Y})^2\right],$$

where $Z_i''=(y_i-\bar{Y})-\frac{R}{N}x_i$ $E_i''=\frac{(1-p_i)}{p_i}$, and $R=\frac{\bar{Y}}{\bar{X}}$.

(2) The estimator of $V\left(\hat{\bar{Y}}_R''^{*}\right)$ is

$$\hat{V}\left(\hat{\bar{Y}}_R''^{*}\right)=\frac{1}{\left(\sum_{i\in s}\frac{r_i}{\pi_ip_i}\right)^2}\left[\sum_{i\in s}\hat{F}_i\hat{Z}_i''^2+\sum_{i\in s}\sum_{j\backslash\{i\}\in s}\hat{J}_{ij}\hat{Z}_i''\hat{Z}_j''+\sum_{i\in s}\frac{r_i}{p_i}\hat{E}_i''(y_i-\hat{\bar{Y}}_R''^{*})^2\right],$$

where $\hat{Z}_i''=\frac{r_i}{p_i}(y_i-\hat{\bar{Y}}_R''^{*})-\frac{\hat{R}''^{(2)}}{N}x_i$, $\hat{R}''^{(2)}=\frac{\hat{\bar{Y}}_r''^{(2)}}{\hat{\bar{X}}_{HT}}$, $\hat{\bar{Y}}_r''^{(2)}=\sum_{i\in s}\frac{r_iy_i}{\pi_ip_i}\left(\sum_{i\in s}\frac{r_i}{\pi_ip_i}\right)^{-1}$,
and $\hat{E}_i''=\frac{(1-p_i)}{\pi_ip_i}$.

(3) If p_i is unknown for some $i\in s$, then the estimator of $V\left(\hat{\bar{Y}}_R''^{*}\right)$ is

$$\hat{V}\left(\hat{\bar{Y}}_R''^{*}\right)=\frac{1}{\left(\sum_{i\in s}\frac{r_i}{\pi_i\hat{p}_i}\right)^2}\left[\sum_{i\in s}\hat{F}_i\hat{Z}_i''^2+\sum_{i\in s}\sum_{j\backslash\{i\}\in s}\hat{J}_{ij}\hat{Z}_i''\hat{Z}_j''+\sum_{i\in s}\frac{r_i}{\hat{p}_i}\hat{E}_i''(y_i-\hat{\bar{Y}}_R''^{*})^2\right],$$

where $\hat{Z}_i''=\frac{r_i}{\hat{p}_i}(y_i-\hat{\bar{Y}}_R''^{*})-\frac{\hat{R}''^{(2)}}{N}x_i$, $\hat{R}''^{(2)}=\frac{\hat{\bar{Y}}_r''^{(2)}}{\hat{\bar{X}}_{HT}}$, $\hat{\bar{Y}}_r''^{(2)}=\sum_{i\in s}\frac{r_iy_i}{\pi_i\hat{p}_i}\left(\sum_{i\in s}\frac{r_i}{\pi_i\hat{p}_i}\right)^{-1}$, $\hat{E}_i''=\frac{(1-\hat{p}_i)}{\pi_i\hat{p}_i}$

and $\hat{p}_i$ is the estimator of p_i estimated by the probit or logistic regression models.

Proof. The proof is similar to Theorem 3.

The existing estimators and proposed estimators are shown in Table 1.

3.3 Simulation studies

The efficiency of the proposed estimators are compared with the existing estimators through simulation studies. The data are generated following the idea of the package 'sampling' in R program (R Core Team [21]) in a case where π_i are weakly correlated with y_i which support the results found in Särndal *et al*. [5] that the Hájek estimator [4] is better than Horvitz and Thompson [3] in some situations including situations where π_i are not strong or are negatively correlated with y_i. Since the existing estimators are developed from the Horvitz and Thompson estimator [3] and the proposed estimators are derived from the Hájek estimator [4] under the same situations when nonresponse occurs in the study variable, generating the model shown below to see the performance of the proposed estimators. The model is defined by $y_i=-20+\frac{1}{\pi_i}+8x_i^2+6w_i^2$ where $N=2,000$, $x_i\sim N(65,5)$, $w_i\sim N(30,5)$, $\pi_i=P_i\frac{(N-n)}{N-1}+\frac{(n-1)}{N-1}$, $P_i=k_i\big/\sum_{i\in U}k_i$, and $i=1,2,..,N$. The UPWOR is used to select six levels of sample sizes that are 50, 100, 200, 300, 400 and 500. In the presence of nonresponse, we consider three levels of response probabilities; 60%, 75% and 90%, and we assume the response probabilities are unknown because it is what occurs in real life. Then, we define the notation of the estimators of response probabilities for both MCAR and MAR as follows.

Let $\hat{p}_1=r/n$ and $\hat{p}_2=\sum_{i\in s}\frac{r_i}{\pi_i}\Big/\sum_{i\in s}\frac{1}{\pi_i}$ be the estimators of response probabilities under MCAR while $\hat{p}_{i1}$ and $\hat{p}_{i2}$ are the estimators of probability of responses using the logistic regression or probit models under MAR. The mean square error (MSE) is used to compare the efficiency of the proposed variance estimators and the formula of MSE

Table 1. The existing estimators and proposed estimators

Type of Estimator	Author	Nonresponse Mechanism	The estimator formula
Existing estimators	Ponkaew and Lawson [16]	MCAR	$\hat{\bar{Y}}_r'^{(1)} = \frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i p}$
		MAR	$\hat{\bar{Y}}_r''^{(1)} = \frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i p_i}$
		MCAR	$\hat{\bar{Y}}_r'^{(2)} = \frac{\sum_{i\in s}\frac{r_i y_i}{\pi_i p}}{\sum_{i\in s}\frac{r_i p}{\pi_i}} = \frac{\sum_{i\in s}\frac{r_i y_i}{\pi_i}}{\sum_{i\in s}\frac{r_i}{\pi_i}}$
	Lawson [13]	MAR	$\hat{\bar{Y}}_r''^{(2)} = \frac{\sum_{i\in s}\frac{r_i y_i}{\pi_i p_i}}{\sum_{i\in s}\frac{r_i}{\pi_i p_i}}$
	Ponkaew and Lawson [14]	MCAR	$\hat{\bar{Y}}_R'^{(1)} = \frac{\frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i p}}{\frac{1}{N}\sum_{i\in s}\frac{x_i}{\pi_i}}\bar{X}$
	Ponkaew and Lawson [20]	MAR	$\hat{\bar{Y}}_R''^{(1)} = \frac{\frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i p_i}}{\frac{1}{N}\sum_{i\in s}\frac{x_i}{\pi_i}}\bar{X}$
Proposed estimators		MCAR	$\hat{\bar{Y}}_R'^{*} = \frac{\sum_{i\in s}\frac{r_i y_i}{\pi_i}\left(\sum_{i\in s}\frac{r_i}{\pi_i}\right)^{-1}}{\frac{1}{N}\sum_{i\in s}\frac{x_i}{\pi_i}}\bar{X}$
		MAR	$\hat{\bar{Y}}_R''^{*} = \frac{\sum_{i\in s}\frac{r_i y_i}{\pi_i p_i}\left(\sum_{i\in s}\frac{r_i}{\pi_i p_i}\right)^{-1}}{\frac{1}{N}\sum_{i\in s}\frac{x_i}{\pi_i}}\bar{X}$

$$MSE\left(\hat{\bar{Y}}\right)=\frac{1}{M-1}\sum_{m=1}^{M}(\hat{\bar{Y}}_m-\bar{Y})^2 ,$$

where $\hat{\bar{Y}}_m$ is the estimator of population mean and $\bar{Y}$ is population mean. The simulation is repeated 10,000 times ($M = 10,000$). The results are shown in Tables 2-4.

The results from Table 2 show that the proposed estimators $\hat{\bar{Y}}''^{*}_{R}$ under MAR for both $\hat{p}_{i1}$ and $\hat{p}_{i2}$ gave similar results and gave smaller MSEs compared to all other estimators for all sample sizes when the response rate was equal to 60%. The proposed estimator $\hat{\bar{Y}}'^{*}_{R}$ under MCAR performed better than all existing estimators under MCAR and better than some existing estimators with MAR. The proposed estimators gave a lot smaller MSEs especially for small sample sizes.

Tables 3 and 4 show similar results for the response rates equal to 75% and 90%, respectively. The proposed estimators $\hat{\bar{Y}}''^{*}_{R}$ performed the best in terms of minimum MSEs. Both $\hat{\bar{Y}}''^{*}_{R}$ with different values of $\hat{p}_{i1}$ and $\hat{p}_{i2}$ gave similar results under MAR. As expected, the bigger response rate the smaller the MSEs seen in the study. The proposed estimators performed a lot better than the existing estimators under both MCAR and MAR, especially for small sample sizes.

3.4 An application to air pollution data

The air pollution data in Suan Luang District, Bangkok, Thailand from the Air Quality and Noise Management Division Bangkok [22] collected hourly in August 2022 are applied in this study to assess the performances of the estimators. The study variable y is PM2.5 (mg/m^3), the temperature (ºC), and particulate matter 10 µm or less in diameter (pm 10: mg/m^3), are auxiliary variables x and w, respectively. The average PM2.5 is 11.41 mg/m^3. The variable x is used to construct a proposed ratio estimator while variable w is used to estimate response probability using logistic regression under the MAR nonresponse mechanism. The size variable k is relative humidity (%). The correlation coefficient between y and x is 0.12 and between y and k is -0.067. The nonresponse rate is 1.95% in this study. The Midzuno scheme [23] is used to select a sample size $n = 150$ records from a population of size 718 records. The results in Table 5 show that the proposed estimators $\hat{\bar{Y}}''^{*}_{R}$ under MAR for both $\hat{p}_{i1}$ and $\hat{p}_{i2}$ gave similar results to the results found in the simulation studies and gave a smaller estimated variance of the mean of PM2.5 compared to all other estimators. The proposed estimators under MAR gave a closer estimated mean of PM2.5 compared to the others. The proposed estimators under MCAR gave smaller variances with respect to all other existing estimators under MCAR and some existing estimators under MAR.

4. Conclusions

New ratio estimators are developed in this study under two nonresponse mechanisms: MCAR and MAR when the population mean of the auxiliary variable is known. The sampling plan is studied under PPSWOR and the variance estimators are considered under a reverse framework, both under MCAR where nonresponse occurs in the study variable under uniform nonresponse and under MAR

Table 2. The mean square error of the existing estimators and proposed estimators with 60% response rate

Estimator	Nonresponse Mechanism	$\hat{p}$	Mean Square Error					
			n					
			50	**100**	**200**	**300**	**400**	**500**
$\hat{\bar{Y}}_r'^{(1)}$	MCAR	$\hat{p}_1$	7.35×10^5	3.63×10^5	1.72×10^5	1.16×10^5	7.65×10^4	5.87×10^4
		$\hat{p}_2$	7.31×10^5	3.62×10^5	1.72×10^5	1.15×10^5	7.65×10^4	5.87×10^4
$\hat{\bar{Y}}_r''^{(1)}$	MAR	$\hat{p}_{i1}$	7.35×10^5	3.65×10^5	1.72×10^5	1.10×10^5	7.44×10^4	5.65×10^4
		$\hat{p}_{i2}$	7.29×10^5	3.63×10^5	1.72×10^5	1.10×10^5	7.43×10^4	5.65×10^4
$\hat{\bar{Y}}_r'^{(2)}$	MCAR	-	7.22×10^5	3.61×10^5	1.72×10^5	1.15×10^5	7.65×10^4	5.87×10^4
$\hat{\bar{Y}}_r''^{(2)}$	MAR	$\hat{p}_{i1}$	7.08×10^5	3.54×10^5	1.71×10^5	1.09×10^5	7.43×10^4	5.65×10^4
		$\hat{p}_{i2}$	7.09×10^5	3.54×10^5	1.71×10^5	1.09×10^5	7.43×10^4	5.65×10^4
$\hat{\bar{Y}}_R'^{(1)}$	MCAR	$\hat{p}_1$	3.07×10^5	1.49×10^5	7.01×10^4	5.08×10^4	3×10^4	2.39×10^4
		$\hat{p}_2$	3.04×10^5	1.49×10^5	7×10^4	5.07×10^4	3×10^4	2.39×10^4
$\hat{\bar{Y}}_R''^{(1)}$	MAR	$\hat{p}_{i1}$	3.12×10^5	1.52×10^5	6.92×10^4	4.49×10^4	2.83×10^4	2.24×10^4
		$\hat{p}_{i2}$	3.05×10^5	1.49×10^5	6.92×10^4	4.49×10^4	2.83×10^4	2.25×10^4
$\hat{\bar{Y}}_R'^{*}$	MCAR	-	3.05×10^5	1.49×10^5	6.99×10^4	4.47×10^4	3×10^4	2.39×10^4
$\hat{\bar{Y}}_R''^{*}$	MAR	$\hat{p}_{i1}$	2.93×10^5	1.43×10^5	6.86×10^4	4.46×10^4	2.82×10^4	2.24×10^4
		$\hat{p}_{i2}$	2.93×10^5	1.43×10^5	6.87×10^4	4.46×10^4	2.82×10^4	2.24×10^4

Table 3. The mean square error of the existing estimators and proposed estimators with 75% response rate

Estimator	**Nonresponse Mechanism**	$\hat{p}$	**Mean Square Error**					
			n					
			50	**100**	**200**	**300**	**400**	**500**
$\hat{\bar{Y}}_r'^{(1)}$	MCAR	$\hat{p}_1$	7.15×10^5	3.46×10^5	1.64×10^5	1.08×10^5	7.86×10^4	6.37×10^4
		$\hat{p}_2$	7.17×10^5	3.46×10^5	1.64×10^5	1.08×10^5	7.86×10^4	6.37×10^4
$\hat{\bar{Y}}_r''^{(1)}$	MAR	$\hat{p}_{i1}$	7.06×10^5	3.38×10^5	1.58×10^5	1.07×10^5	7.76×10^4	6.07×10^4
		$\hat{p}_{i2}$	6.99×10^5	3.37×10^5	1.58×10^5	1.07×10^5	7.76×10^4	6.07×10^4
$\hat{\bar{Y}}_r'^{(2)}$	MCAR	-	7.10×10^5	3.45×10^5	1.64×10^5	1.08×10^5	7.86×10^4	6.37×10^4
$\hat{\bar{Y}}_r''^{(2)}$	MAR	$\hat{p}_{i1}$	6.86×10^5	3.35×10^5	1.58×10^5	1.07×10^5	7.74×10^4	6.06×10^4
		$\hat{p}_{i2}$	6.86×10^5	3.35×10^5	1.58×10^5	1.07×10^5	7.74×10^4	6.06×10^4
$\hat{\bar{Y}}_R'^{(1)}$	MCAR	$\hat{p}_1$	2.76×10^5	1.53×10^5	7.88×10^4	4.51×10^4	3.22×10^4	2.67×10^4
		$\hat{p}_2$	2.76×10^5	1.53×10^5	7.85×10^4	4.52×10^4	3.22×10^4	2.67×10^4
$\hat{\bar{Y}}_R''^{(1)}$	MAR	$\hat{p}_{i1}$	2.75×10^5	1.45×10^5	7.27×10^4	4.30×10^4	3.15×10^4	2.42×10^4
		$\hat{p}_{i2}$	2.66×10^5	1.45×10^5	7.27×10^4	4.30×10^4	3.15×10^4	2.42×10^4
$\hat{\bar{Y}}_R'^{*}$	MCAR	-	2.80×10^5	1.53×10^5	7.56×10^4	4.51×10^4	3.23×10^4	2.67×10^4
$\hat{\bar{Y}}_R''^{*}$	MAR	$\hat{p}_{i1}$	2.59×10^5	1.43×10^5	7.26×10^4	4.29×10^4	3.13×10^4	2.42×10^4
		$\hat{p}_{i2}$	2.59×10^5	1.43×10^5	7.26×10^4	4.29×10^4	3.13×10^4	2.42×10^4

Table 4. The mean square error of the existing estimators and proposed estimators with 90% response rate

Estimator	Nonresponse Mechanism	$\hat{p}$	Mean Square Error					
			n					
			50	**100**	**200**	**300**	**400**	**500**
$\hat{\bar{Y}}_r'^{(1)}$	MCAR	$\hat{p}_1$	4.02×10^5	3.73×10^5	1.63×10^5	1.01×10^5	8.71×10^4	5.66×10^4
		$\hat{p}_2$	4.02×10^5	3.72×10^5	1.63×10^5	1.01×10^5	8.71×10^4	5.66×10^4
$\hat{\bar{Y}}_r''^{(1)}$	MAR	$\hat{p}_{i1}$	3.94×10^5	3.65×10^5	1.54×10^5	9.75×10^4	7.91×10^4	5.49×10^4
		$\hat{p}_{i2}$	3.94×10^5	3.64×10^5	1.54×10^5	9.74×10^4	7.91×10^4	5.49×10^4
$\hat{\bar{Y}}_r'^{(2)}$	MCAR	-	4.02×10^5	3.71×10^5	1.62×10^5	1.01×10^5	8.71×10^4	5.65×10^4
$\hat{\bar{Y}}_r''^{(2)}$	MAR	$\hat{p}_{i1}$	3.89×10^5	3.62×10^5	1.53×10^5	9.71×10^4	7.85×10^4	5.48×10^4
		$\hat{p}_{i2}$	3.89×10^5	3.62×10^5	1.53×10^5	9.71×10^4	7.86×10^4	5.48×10^4
$\hat{\bar{Y}}_R'^{(1)}$	MCAR	$\hat{p}_1$	1.77×10^5	1.52×10^5	6.72×10^4	4.35×10^4	4.27×10^4	2.30×10^4
		$\hat{p}_2$	1.77×10^5	1.51×10^5	6.72×10^4	4.35×10^4	4.27×10^4	2.29×10^4
$\hat{\bar{Y}}_R''^{(1)}$	MAR	$\hat{p}_{i1}$	1.72×10^5	1.45×10^5	6.30×10^4	3.88×10^4	3.44×10^4	2.15×10^4
		$\hat{p}_{i2}$	1.72×10^5	1.45×10^5	6.29×10^4	3.88×10^4	3.43×10^4	2.15×10^4
$\hat{\bar{Y}}_R'^{*}$	MCAR	-	1.77×10^5	1.51×10^5	6.30×10^4	4.35×10^4	4.27×10^4	2.29×10^4
$\hat{\bar{Y}}_R''^{*}$	MAR	$\hat{p}_{i1}$	1.68×10^5	1.43×10^5	6.26×10^4	3.85×10^4	3.36×10^4	2.14×10^4
		$\hat{p}_{i2}$	1.68×10^5	1.43×10^5	6.26×10^4	3.85×10^4	3.37×10^4	2.14×10^4

Table 5. The estimated mean and variance of PM2.5 in Suan Luang District

Estimator	Nonresponse Mechanism	$\hat{p}$	Estimated Mean of PM2.5	Estimated Variance of Mean of PM2.5
$\hat{\bar{Y}}_r'^{(1)}$	MCAR	$\hat{p}_1$	12.16	0.17881
		$\hat{p}_2$	12.16	0.17882
$\hat{\bar{Y}}_r''^{(1)}$	MAR	$\hat{p}_{i1}$	12.14	0.17721
		$\hat{p}_{i2}$	12.14	0.17757
$\hat{\bar{Y}}_r'^{(2)}$	MCAR	$\hat{p}_1$		0.17438
		$\hat{p}_2$	12.16	0.17436
$\hat{\bar{Y}}_r''^{(2)}$	MAR	$\hat{p}_{i1}$	12.14	0.17384
		$\hat{p}_{i2}$	12.14	0.17379
$\hat{\bar{Y}}_R'^{(1)}$	MCAR	$\hat{p}_1$	12.00	0.17674
		$\hat{p}_2$	12.00	0.17672
$\hat{\bar{Y}}_R''^{(1)}$	MAR	$\hat{p}_{i1}$	11.98	0.17621
		$\hat{p}_{i2}$	11.98	0.17615
$\hat{\bar{Y}}_R'^{*}$	MCAR	$\hat{p}_1$		0.17437
		$\hat{p}_2$	11.58	0.17438
$\hat{\bar{Y}}_R''^{*}$	MAR	$\hat{p}_{i1}$	11.41	0.17373
		$\hat{p}_{i2}$	11.41	0.17369

which enhances versatility, and where the nonresponse is not uniform which is more likely to happen in the real world. The proposed estimators are shown in theory to be approximately unbiased estimators. The results showed in both simulation studies and in the case of an application to air pollution data that the proposed estimators under MAR performed the best with respect to all estimators for all levels of nonresponse rates and sample sizes, and the proposed estimator under MCAR also performed well when compared to other existing estimators, especially for small sampling fractions. Although the new estimators offer more flexibility when missing values occur in the study variable under MAR, the population mean of the auxiliary variable is required. In future, we can extend our work to cases where the population mean of the auxiliary variable is unknown and also to situations where nonresponse can occur in both the study and auxiliary variables under different sampling plans. Nevertheless, the new estimators work well especially for small sample sizes which can help in saving money and time for researchers in collecting data from surveys and can be applied to real world problems with more accuracy in the estimation process.

5. Acknowledgements

This research was funded by the National Science, Research and Innovation Fund (NSRF), and King Mongkut's University of Technology North Bangkok Contract no. KMUTNB-FF-66-56.

References

[1] Cochran, W.G., 1977. *Sampling Techniques*. New York: John Wiley and Sons.
[2] Bacanli, S. and Kadilar, C., 2008. Ratio estimators with unequal probability designs. *Pakistan Journal of Statistics*, 24(3), 167-172.
[3] Horvitz, D.F. and Thompson, D.J., 1952. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260), 663-685, https://doi.org/10.2307/2280784.
[4] Hájek, J., 1981. *Sampling from Finite Population*. New York, Marcel Dekker.
[5] Särndal, C.-E., Swensson, B. and Wretman, J., 1992. *Model Assisted Survey Sampling*. Berlin: Springer-Verlag Publishing.
[6] Thongsak, N. and Lawson, N., 2022. 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*, 30(3), https://doi.org/10.1080/08898480.2022.2145790.
[7] Thongsak, N. and Lawson, N., 2022. Classes of combined population mean estimators utilizing transformed variables under double sampling with application to air pollution in Chiang Rai, Thailand. *Songklanakarin Journal of Science and Technology*, 44(5), 1390-1398.
[8] Thongsak, N. and Lawson, N., 2022. A combined family of dual to ratio estimators using a transformed auxiliary variable. *Lobachevskii Journal of Mathematics*, 43(9), 2621-2633.
[9] 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.
[10] Lawson, N., 2017. Variance estimation in the presence of nonresponse under probability proportional to size sampling. *Proceedings of the 6th Annual International Conference on Computational Mathematics, Computational Geometry and Statistics (CMCGS 2017),* Singapore, March 6-7, 2017, https://doi.org/10.5176/2251-1911_CMCGS17.32.
[11] Ponkaew, C. and Lawson, N., 2018. A new ratio estimator for population total in the presence of nonresponse under unequal probability sampling without replacement. *Thai Journal of Mathematics*, Special Issue 2018, 417-429.
[12] Särndal, C.E. and Lundström, S., 2005. *Estimation in surveys with nonresponse*. New York: John Wiley and Sons.
[13] Ponkaew, C. and Lawson, N. 2019. Estimating variance in the presence of nonresponse under unequal probability sampling. *Suranaree Journal of Science and Technology*, 26(3), 293-302.
[14] Lawson, N. and Ponkaew, C., 2019. New generalized regression estimator in the presence of nonresponse under unequal probability sampling. *Communications in Statistics - Theory and Methods,* 48(10), 2483-2498, https://doi.org/10.1080/03610926.2018.1465091.
[15] Lawson, N. and Siripanich, P., 2020. A new generalized regression estimator and variance estimation for unequal probability sampling without replacement for missing data. *Communications in Statistics - Theory and Methods,* 51(18), 6296-6318, https://doi.org/10.1080/03610926.2020.1860224.

[16] Ponkaew, C. and Lawson, N., 2023. New generalized regression estimators using a ratio method and its variance estimation for unequal probability sampling without replacement in the presence of nonresponse. *Current Applied Science and Technology,* 23(2), https://doi.org/ 10.55003/cast.2022.02.23.007.

[17] Ponkaew, C. and Lawson, N., 2022. New estimators for estimating population total: an application to water demand in Thailand under unequal probability sampling without replacement for missing data. *PeerJ,* https://doi.org/10.7717/peerj.14551.

[18] Chodjuntug, K. and Lawson, N., 2022. Imputation for estimating the population mean in the presence of nonresponse, with application to fine particle density in Bangkok. *Mathematical Population Studies*, 29(4), 204-225, https://doi.org/ 10.1080/08898480.2021.1997466.

[19] Chodjuntug, K. and Lawson, N., 2022. A chain regression exponential type imputation method for mean estimation in the presence of missing data. *Songklanakarin Journal of Science and Technology,* 44(4), 1109-1118.

[20] Lawson, N., 2023. New imputation method for estimating population mean in the presence of missing data. *Lobachevskii Journal of Mathematics,* 44(9). (article in print).

[21] R Core Team., 2021. *R : A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria*. [online] Available at: https://www.R-project.org/.

[22] Air Quality and Noise Management Division Bangkok. 2023. *Report Measurement*. [online] Available at: https://bangkokairquality.com/bma/report?lang=en.

[23] Midzuno, H., 1952. On sampling system with probability proportional to sum of sizes. *Annals of the Institute of Statistical Mathematics*, 3(1), 99-107.